Intelligent Augmented Reality (iAR): Context-aware Inference and Adaptation in AR

Shakiba Davari

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Computer Science and Application

Douglas A. Bowman, Chair

Christopher L. North

Yalong Yang

Wallace S. Lages

Missie I. Smith

09/08/2024

Blacksburg, Virginia

Keywords: Augmented Reality, Context-Aware, Adaptive AR, Intelligent systems, UI/UX, XR Design

Copyright 2024, Shakiba Davari

Intelligent Augmented Reality (iAR): Context-aware Inference and Adaptation in AR

Shakiba Davari

(ABSTRACT)

Augmented Reality (AR) transforms the entire 3D space around the user into a dynamic screen, surpassing the limitations of traditional displays and enabling efficient access to multiple pieces of information simultaneously, all day, every day. Recent developments in AR eyeglasses promise that AR could become the next generation of personal computing devices. To realize this vision of pervasive AR, the AR interface must address the challenges posed by constant and omnipresent virtual content. As the user's context changes, the virtual content in AR head-worn displays can occasionally become obtrusive, hindering the user's perception and awareness of their surroundings and their interaction with both the virtual and physical worlds. An intelligent interface is needed to adapt the presentation and interaction of AR content. This dissertation outlines a roadmap towards effective, efficient, and unobtrusive AR through intelligent AR (iAR) systems that automatically learn and adapt the interface to the user's context. To achieve this goal, we: (1) Identify multiple AR design principles and guidelines that maintain efficiency while addressing challenges such as occlusion, social interaction, and content placement in AR. (2) Demonstrate the impact of context on AR effectiveness, validating the advantages of context-awareness and highlighting the complexities of implementing a context-aware approach in pervasive AR, particularly in scenarios involving context-switching. (3) Propose a design space for XR interfaces; (4) Develop a taxonomy of quantifiable contextual components and a framework for designing iAR interfaces.

Intelligent Augmented Reality (iAR): Context-aware Inference and Adaptation in AR

Shakiba Davari

(GENERAL AUDIENCE ABSTRACT)

Augmented Reality (AR) integrates digital information with the real world in real-time, transforming the surrounding physical space into a dynamic, interactive screen. This technology can simultaneously provide hands-free access to unlimited virtual applications and information, facilitating fast and easy access. With recent advancements in AR eyeglasses, AR is anticipated to become the next generation of personal computing, potentially replacing mobile phones and computers. However, to be seamlessly integrated into daily life, AR must overcome challenges such as occluding important real-world objects, distraction, visual clutter, and information overload. This dissertation presents a roadmap for developing intelligent AR (iAR) systems that automatically adapt to the user's context. To achieve this goal, we identify the design space for adaptable AR elements and design and test various context-aware AR interfaces. We identify key AR design principles that ensure efficiency while addressing challenges like occlusion, social interaction, and content placement. We also highlight the impact of context on AR effectiveness and the complexities of implementing a context-aware approach, especially in context-switching scenarios. Additionally, we informed the design process of iAR interfaces by identifying the contextual components that influence their effectiveness and providing a framework and architecture for utilizing this information for automatic adaptations. These efforts aim to enhance AR effectiveness and efficiency while ensuring it remains unobtrusive in everyday use.

Dedication

To Women, Life, Freedom

This dissertation is dedicated to the courageous freedom fighters of the Mahsa Amini Movement—those who have lost their lives, those who have been forced to leave their homes, and those who are imprisoned for standing up for truth and justice. May we one day live in a world where being a woman does not endanger your life and where freedom is not threatened simply because of your gender.

Acknowledgments

I have been fortunate to encounter and be blessed with incredible people throughout this journey. Many remarkable individuals have made this challenging path immensely fulfilling, making it difficult to leave Blacksburg.

First and foremost, I want to express my deepest gratitude to my advisor, Doug A. Bowman, for being the best advisor I could have asked for. Doug, you are an exceptional academic and leader from whom I have learned so much. My current and future accomplishments are a testament to your guidance. Beyond that, as an amazing role model, you have shown me, by example, how to be a better person. While words cannot fully capture the depth of my respect and gratitude, please know that I will forever be thankful for your unwavering trust and support and for being part of the 3DI family. *It was truly an honor!*

I also extend my heartfelt thanks to the supportive members of my 3DI family: Cory Ilo, Ibrahim Tahmid, Alexander Giovannelli, Logan Lane, Cherelle Connor, Feiyu Lu, Leonardo Pavanatto, Jerald (& Rosey) Thomas, Daniel Stover, Yuan Li, Lee Lisle, Run Yu, Lei Zhang, and Kylie Davidson, Christos Lougiakis, and Francielly Rodrigues. I truly enjoyed getting to know each of you and learned a great deal from everyone. This lab exemplified how intelligent, motivated, and creative people can support and respect one another, engage in healthy competition, and push each other to improve.

I am deeply grateful to my mentor, Missie Smith, for her generosity, continuous support, and encouragement. Thank you, Missie, for believing in me, sharing your experiences, and always providing invaluable advice. I deeply appreciate your time and immensely respect you as both a researcher and a compassionate, supportive mentor.

I also want to thank my Ph.D. committee members, Chris North, Wallace Lages, Missie

Smith, and Yalong Yang, for their invaluable guidance, thought-provoking questions, and suggestions. Your unwavering support has significantly shaped the direction of this dissertation.

My sincere thanks go to my mentors during my Microsoft Research and Adobe internships: Andrew Begel, Ed Cutrell, Rich Stoakeley, Carmen Badea, Stefano Petrangeli, and Jane Hoffswell. Thank you for the countless meetings and discussions, for sharing your expertise and resources, and for guiding me in becoming a better industrial researcher.

On a personal note, I am deeply thankful to my family for everything they have done for me, for believing in me, and for giving me the freedom to choose my own path. From both my brothers' weddings to my niece's birth, there were many wonderful moments I missed during this Ph.D. journey. Thank you for understanding and encouraging me to continue.

I would like to express my heartfelt appreciation to my amazing boyfriend, Dr. Miles Frantz, and his supportive family for being there for me and supporting me through tough times. Miles, I admire your kindness, devotion, and patience. Thank you for being an incredibly understanding, loving partner and for standing by me.

Special thanks to Alexander Giovannelli and Cory Ilo, who are not just brilliant researchers but also amazing friends who pushed me to the finish line. Alex, your creativity, diverse knowledge, and sense of humor made every conversation a joy. Thank you, Gio, for always cheering me up. Cory, your intelligence, resilience, and kindness have taught me how to navigate tough situations with strength. Thank you for being an understanding friend and helping me grow.

The list of incredible people who have supported me is extensive, and while I cannot mention everyone by name, I want to extend my sincere thanks to the CS department at Virginia Tech, along with the amazing professors and staff who welcomed me into this community and supported me throughout these past years. I am also deeply grateful to my friends from the Iranian Society and the CS department at VT for their camaraderie and encouragement.

I would like to acknowledge the Virginia Tech IRB and express my gratitude to the Virginia Tech community members who participated in my user study for their valuable insights and time.

Lastly, I want to note that the use of AI tools in this work was limited to the final stages of polishing and refining my initial draft.

Contents

List of Figures xv

List of Tables xx

1 Introduction 1

1.1 Motivation 1

1.2 Research Questions 5

1.3 Proposed Approach 8

1.3.1 Desig and Evaluation of Context-Aware AR Interfaces in Different Contexts 8

1.3.2 Design Space & A Context-aware AR Placement Strategy in Different Contexts 9

1.3.3 Taxonomy of Context & iAR Framework 10

1.3.4 An Architecture for iAR & An Empirical Study 10

1.4 Contributions 11

2 Review of Literature 13

2.1 Information Acquisition and Awareness Information 13

2.2 AR Applications and Challenges 15

2.3 AR Design Space 17

2.3.1 Presentation 18

2.3.2 Interaction 19

2.4 Context-Awareness in AR 21

3 Designing Context-Aware AR Interfaces in Different Contexts 26

3.1 Overview 26

3.2 Glanceable AR 27

3.3 Experiment 1: AR Occlusion Management in a Real-World Monitoring Context 28

3.3.1 Design of Occlusion-Aware AR Interfaces for Real-World Monitoring Tasks 29

3.3.2 Experiment Setup 32

3.3.3 Evaluation Metrics 35

3.3.4 Hypotheses 36

3.3.5 Participants 37

3.3.6 Results 38

3.3.7 Discussion 42

3.4 Experiment 2: AR Information Acquisition in Social and Solo Context . . . 46

3.4.1 Research Questions 47

3.4.2 Design of Context-Aware AR Interfaces 48

3.4.3 Experiment Setup 51

3.4.4 Evaluation Metrics 57

3.4.5 Hypotheses 58

3.4.6 Participants 60

3.4.7 Results 60

3.4.8 Discussion and Future Work 65

3.5 Conclusions 70

4 Design Space & A Context-aware AR Placement Strategy **72**

4.1 Overview 72

4.2 XR Design Dimensions 74

4.2.1 Content Design 75

4.2.2 Presentation Design 76

4.2.3 Input Design 85

4.3 Context-aware Environment-referenced Placement 86

4.4 Experiment 3: AR Content Placement Strategy in Four Contexts 89

4.4.1 Experimental Design 90

4.4.2 Hypotheses 94

4.4.3 Apparatus 97

4.4.4 Procedure 99

4.4.5 Evaluation Metrics 100

4.4.6 Participants 102

4.4.7 Results 103

4.4.8 Discussion 111

4.5 Conclusions 117

5 Taxonomy of Context & iAR Framework 118

5.1 Overview 118

5.2 Taxonomy of Context 119

5.2.1 User 122

5.2.2 Setting 126

5.2.3 Setting-User Interplay 130

5.3 Framework 132

5.4 Conclusion 133

6 An Architecture for iAR & An Empirical Study 135

6.1 Overview 135

6.2 Architecture 136

6.2.1 Context Unit 137

6.2.2 Inference Unit 138

6.2.3 Adaptation 143

6.3 Reproducing the Design of Occlusion-Aware AR Interfaces for Real-World Monitoring Task . . . 145

6.4 Experiment 4: User-Specified AR Adaptations in Eight Contexts . . . 148

6.4.1 Experimental Design . . . 148

6.4.2 Apparatus . . . 152

6.4.3 Procedure . . . 153

6.4.4 Participants . . . 154

6.4.5 Data Collection . . . 154

6.4.6 Results . . . 155

6.4.7 Discussion . . . 165

6.5 Conclusion . . . 168

7 Conclusions and Future Work **170**

7.1 Summary . . . 170

7.2 Insights and Reflections . . . 171

7.3 Future Work Directions . . . 176

7.3.1 Input and Non-visual Presentation in iAR . . . 177

7.3.2 Frequent Context Switch and Adaptation in iAR . . . 177

7.3.3 Data Collection and Implementation . . . 178

7.3.4 Security & Privacy . . . 179

Bibliography **180**

Appendices **201**

Appendix A Experiment 1: AR Occlusion Management in a Real-World Monitoring Context **202**

A.1 Questionnaires and Interview 203

A.1.1 Post-Condition: NASA TLX Questionnaire 203

A.1.2 Post-Condition Usability Questionnaire 204

A.1.3 Interview Questions 205

A.2 IRB Approved Informed Consent Form 205

Appendix B Experiment 2: AR Information Acquisition in Social and Solo Context **210**

B.1 Experimental Design 210

B.1.1 Reading Material for the Solo Context 210

B.1.2 Conversation Material for the Social Context 214

B.2 Questionnaires and Interview 216

B.2.1 Post-Condition Questionnaire 216

B.2.2 Interview Questions 223

B.3 IRB Approved Informed Consent Form 225

Appendix C Experiment 3: AR Content Placement Strategy in Four Con-

texts **229**

C.1 IRB Approved Informed Consent Form 230

Appendix D Experiment 4: User-Specified AR Adaptations in Eight Contexts **235**

D.1 IRB Approved Informed Consent Form 236

List of Figures

3.1 A screenshot of the virtual environment showcasing the Automatic-Reposition interface. This figure shows a trial in which the user is required to answer a question from the virtual content (appearing on the TV) and react to the occurrence of an RW event (the waving of the child). 33

3.2 User's choice of best interface 39

3.3 Mean of perceived ease of accessing information from the real world and virtual content based on interface dimensions 41

3.4 Time for accessing information from the real world and the virtual content based on interface 42

3.5 Average operation count based on an interface. Most pairwise comparisons are significant; they are not shown in the figure to avoid visual clutter. . . . 43

3.6 We propose a context-aware augmented reality interface that minimizes intrusiveness while providing fast and easy information access during both solo and social contexts: (a) In a solo context, the virtual content is glanceable and opaque when not occluding any object of importance in the real world; (b) In a social context, the real-world is prioritized, keeping the virtual content transparent when not needed, avoiding occlusion, distraction, and visual clutter; (c) In a social context, a piece of virtual content relevant to the ongoing conversation moves up and turns opaque to provide fast and easy access without blocking the conversation partner's face. 46

3.7 Screenshots of our prototype in the social context. (a) When in conversation, the real world is prioritized by keeping the virtual content's level of transparency high. (b) The desired virtual content is provided automatically based on the conversation content without occluding the conversation partner's faces. 49

3.8 Conditions compared to Context-Aware Glanceable AR in the experiment. (a) Mobile phone interface. (b) Non-adaptive Glanceable AR interface, in which the user can air-tap an app to turn it opaque or transparent. 55

3.9 Time to answer questions (left) and time away from the reading task (right) in the solo context. 61

3.10 Time to answer questions (left) and time away from the conversation (right) for both question types in the social context. 62

3.11 Missed social cues overall (left) and during information access (right) in the social context. 64

3.12 UEQ results in the solo context. 65

3.13 UEQ results in the social context. 66

4.1 Design Dimensions of XR interfaces 74

4.2 XR design elements for non-visual modalities 82

4.3 Illustration of our two placement strategies. **(1-a, 2-a)**: Body-fixed placement: each XR object's pose remains consistent based on the orientation of the user, i.e., frame of reference. **(1-b, 2-b)**: Environment-referenced placement: each XR object's pose remains consistent based on the orientation of its intermediary relative to the frame of reference, maintaining a consistent distance from the user. **(1-a, 1-b)**: The user is at the session start point, positioned at an equal distance from the intermediaries. **(2-a, 2-b)**: The user has moved near the Sports category's intermediary. 88

4.4 Each document displays the answer to a trivia question 91

4.5 Environment-referenced placement in dynamic contexts: a) The category panel relevant to each trivia host is displayed above their head. b) the panels move along with their associated trivia host. 93

4.6 Body-fixed placement in dynamic contexts: a) All three category panels in front of the user's body. b) No AR content is placed near the trivia hosts, not directly in front of the user. 94

4.7 a) The questions are presented word-by-word. b) The header for each category panel reappears after the question is asked. 98

4.8 Accuracy within various RW settings 104

4.9 Time efficiency within various RW Settings 105

4.10 Gaze switch efficiency within various contexts 106

4.11 User preference on placement strategy in various RW settings 110

4.12 The influence of Info-focus Relevance on gaze switch efficiency within static contexts 114

4.13 The influence of Info-focus Relevance on time efficiency within static contexts 115

5.1 To identify context and provide the proper adaptations as output, an intelligent AR interface uses quantifiable data from the user, sensors, and personal or publicly-available data storage and databases. 119

5.2 Our taxonomy identifies three principal categories of components within context: User, Setting, and Setting-User Interplay. 122

5.3 User is described by persistent context components about the user profile and dynamic context components about their current state. 123

5.4 The real-world environment, consists of persistent and dynamic information about the local environment, and dynamic context components about the structure and state of the physical items and phenomena in the user's immediate environment. 125

5.5 The principal subcategories of context components within User, Setting and SUI in our taxonomy of context. 127

5.6 Digital setting is a collection of persistent and publicly available global information, and dynamic context components representing the hardware and firmware state and capabilities of available digital devices and the AR interface, including the current state of the AR interface's design elements. 128

5.7 Social setting is presented by dynamic context components providing accessible information from the profile and current state of other people (attendees) present in the user context and their interactions with each other and with RW and digital setting. 130

5.8 SUI includes context components that represent the user's connections and interplay with other people and the RW and digital setting. 131

5.9 Our framework infers potential adjustments for interface performance using the context identified by our taxonomy. The Adaptation unit optimizes these inferences to adapt the interface. 132

6.1 Our framework infers potential adjustments for interface performance using the context identified by our taxonomy. The Adaptation unit optimizes these inferences to adapt the interface. 136

6.2 The architecture of intelligent AR systems. 137

6.3 Highest Density Region Contours (HDS) of Scale vs. Opacity in each sector within the Body-fixed FoR 157

6.4 HDS of Scale vs. Opacity in each sector within the Head-fixed FoR 157

6.5 User-specified app Visibility within various Contexts. 160

6.6 Influence of App Role on the user-specified design dimensions within the Body-fixed FoR 161

6.7 The histogram of the Opacity and Scale of apps with different App Roles within Body-fixed (Top) and Head-fixed (Bottom) FoRs 163

List of Tables

List of Abbreviations

3D Three-dimensional

AI Artificial Intelligence

AR Augmented Reality interactively registers a combination of physical and virtual objects in 3D space in real-time [5].

CV Computer Vision

FoV The Field of View in a display is the angular range of the observable world it allows at a given moment.

HCI Human Computer Interaction

HWD Head-worn Display

LoD Level of Detail

ML Machine Learning

MR Mixed Reality

PAR Pervasive AR

UI User Interface

UX User Experience

VE Virtual Environment

VR Virtual Reality

XR Extended Reality

Chapter 1

Introduction

1.1 Motivation

Digital information plays a vital role in people's personal and professional affairs, offering the potential to enhance productivity, foster social interactions, and occasionally enhance safety measures [23]. This leads to near-constant information acquisition in different situations, and contexts [126]. Physical displays such as mobile phones, laptops, tablets, monitors, and TVs have proven to be sufficient for presenting this information [31]. Yet, the quantity and dimensions of these physical 2D displays are bound by financial, physical, and computing power constraints, allowing them to present limited content within a finite space. These constraints can increase the duration and steps involved in information access and distract users from their task and the surrounding environment. While the limited size of physical displays and its trade-off with portability can restrict availability and hinder multitasking and efficiency, Augmented Reality (AR) offers the potential to overcome these constraints, enhancing both efficiency and user experience.

AR interactively integrates virtual and real objects in real-time [5]. Future AR interfaces can support continuous hands-free access to multiple pieces of information simultaneously [136]. By displaying information in the entire 3D space around the user, AR allows effortless interaction with numerous virtual objects and information with little to no limitation on their presentation and portability. For instance, while conversing about scheduling a hike

on a sunny day, using a mobile phone involves multiple steps to 1) retrieve the device, 2) unlock the screen, 3) locate the calendar and weather apps, and 4) frequently switch between them. Furthermore, this process diverts the user's attention away from their conversation partner and social cues. By allowing simultaneous access to both apps, AR reduces the steps involved in transitioning between and navigating them. Additionally, AR offers the potential to position the apps near the interlocutor's face, enabling the user to maintain social awareness and engage in the conversation seamlessly [29].

In recent years, AR attracted considerable attention across various domains. Due to its potential to continuously provide effortless and hands-free access to information anywhere, and at any time, AR has enabled new opportunities in entertainment, communication, and productivity. Recent advancements in AR head-worn displays (HWDs) increase the potential for the widespread adoption and accessibility of lightweight and powerful AR glasses in everyday life. Future AR is envisioned to become the next generation of personal computing devices, replacing our smartphones and PCs and providing the information "that we expect to see both at work and at play [43, 56, 133]." However, these AR interfaces also present challenges to the user's perception and awareness of their surroundings, as well as their interactions with the world and how they are perceived by others.

While AR's persistent display enhances efficiency and productivity, it can also lead to obtrusiveness, information overload, distraction, clutter, occlusion, and, in extreme cases, reduction of users' environmental awareness and safety. Inappropriate presentation of continuous virtual content can increase physical and cognitive effort, potentially compromising user safety, convenience, efficiency, and social interactions [11, 41, 55]. For example, AR emergency exit indicators can be lifesaving during an evacuation when visibility is obstructed by smoke. However, the continuous display of large, bright pointers can become occlusive, distracting, and stress-inducing in everyday situations. An optimal interface should dy-

namically adapt to maintain effectiveness and unobtrusiveness, enhancing the user's overall experience.

Designing effective interfaces for ubiquitous everyday AR goes beyond providing efficient information access; it involves unobtrusive, non-distracting, and safe integration of digital information into the real world [57]. Context-aware adaptations are crucial for addressing these challenges. Considering both context and the presentation of AR information enables designing interfaces that provide the necessary information, preserve the user's awareness of the real world, and maintain usability and clarity. By leveraging contextual information about the user's requirements, constraints, and preferences, AR interfaces can dynamically adjust to ensure effectiveness. AR eyeglasses enable collecting contextual information, such as the user's gaze, head and body pose, and their view, providing insights into the user's attention, activity, and location. Understanding how individuals perceive and interact with virtual and physical environments allows an interface to present the right information at the right time and in the right form while supporting real-world awareness and interaction [82]. AR's capability to provide such insights makes them particularly valuable in designing context-aware systems that personalize information presentation and optimize the interface for the current situation.

An optimal AR interface aims to sense and adapt to the user's context, balancing efficiency with unobtrusiveness. For instance, an AR interface persistently displaying multiple frequently used apps is beneficial in a static, familiar environment like one's room, prioritizing information access. However, in a public scenario, such as walking down a busy street, the same interface can cause occlusion and visual clutter, compromising safety and situational awareness. Such public contexts involve unique challenges, such as social interactions and limited time for information access [126], requiring different interface adaptations. Recently, context-aware AR interfaces have attracted significant attention in the research community

as a means to optimally provide digital information and enhance users' safety and efficiency in achieving their personal and social goals [17, 55, 89].

Context-aware AR identifies *one or multiple contextual components* within a specific scenario and adapts the AR interface *based on specific design choices and principles*. However, the complexity of designing optimal AR interfaces arises from the interdependence and interaction of various contextual components. Any small changes to the context can introduce new restrictions and opportunities, potentially rendering a previously optimal interface ineffective or harmful. As a result, no single contextual component or a limited number of them can consistently dictate design choices indefinitely. Consequently, the effectiveness of these context-aware interfaces remains limited to their specific scenarios.

An optimal AR interface must function seamlessly in diverse real-world contexts, each presenting unique challenges and success criteria. However, due to the countless subtle changes that can shift a user's context, it is impractical to manually identify the specific design choices and principles needed for each context. Integrating intelligent systems in AR interfaces offers the potential to automatically learn and adapt the interface, optimizing the user's experience and efficiency.

Therefore, we propose to study **Intelligent AR (iAR)**, which detects and integrates *all contextual components* that could potentially influence AR effectiveness and intelligently predicts the *impact of various adaptations* within the current context. Accordingly, iAR provides a solution for adapting the interface for each unique combination of components without necessitating predefined design choices and principles.

1.2 Research Questions

To enable pervasive context-aware AR and pave the way for designing iAR interfaces that automatically learn and adapt to context, this dissertation focuses on addressing the following Research Questions (RQs):

RQ1. What design principles can enhance AR effectiveness in occlusion-prone contexts?

The occlusion of real-world objects by digital elements in AR presents significant challenges, particularly in maintaining user situational awareness. When critical real-world features, such as moving vehicles, steps, or obstacles, are obscured, safety risks such as accidents or injuries can arise. Even in less critical scenarios, such occlusions may lead to confusion, discomfort, and reduced trust in the AR system.

This issue is especially crucial in occlusion-prone contexts where the user tasks require high environmental awareness or where user safety is paramount, such as industrial environments and navigation systems. Leveraging contextual information in AR interfaces is key to enabling the seamless and safe integration of digital and real-world elements.

Identifying design principles that enhance AR effectiveness in such occlusion-prone contexts is vital for enabling safe and effective context-aware AR.

RQ2. How does utilizing social context in the design of an AR interface influence AR effectiveness?

AR can profoundly influence users' interactions and awareness of others, reshaping how they perceive and engage with their environment. It can either enhance or obscure social cues, communication, and both individual and collaborative tasks. Integrating social context into AR interfaces can improve user awareness, foster better interactions, and create more

seamless, intuitive experiences. Therefore, exploring design principles for incorporating social context into AR interface design is essential.

Addressing this RQ provides a practical approach to identifying design principles, contextual components, and AR design elements influencing AR effectiveness in social contexts. Additionally, understanding the benefits of context awareness is critical to validating the motivation behind designing iAR systems, thereby building a solid case for their development and adoption.

RQ3. How does utilizing environmental context in AR content placement influence AR effectiveness?

The alignment of digital elements with the real-world environment plays a crucial role in determining the usability and effectiveness of AR. Strategic placement of these elements can enhance situational awareness, reduce cognitive load, and facilitate AR access. Exploring design principles for utilizing environmental context in AR content placement is essential for creating interfaces that seamlessly integrate with users' surroundings. Addressing this research question offers a practical approach to identifying fundamental design principles and contextual factors for context-aware AR placement, and validates the rationale behind the design of intelligent AR systems.

RQ4. How do contextual changes influence AR effectiveness?

Understanding how contextual changes influence the effectiveness of a particular AR interface design is essential for designing iAR systems that can adapt seamlessly across diverse contexts. By examining AR performance in different contexts, we can identify the key contextual components that impact interface effectiveness and determine the scenarios in which a design is optimal. Addressing this research question is crucial for establishing design principles that guide adapting or applying specific AR interface designs in response to contextual shifts. Moreover, it underscores the importance of context awareness in the design and adop-

tion of iAR systems, ensuring their practical functionality across varied contexts.

RQ5. What are the elements of AR interfaces that can be adapted based on user context?

iAR continuously selects the set of adaptations that best improve the effectiveness of its input and output [107]. Identifying the design space and design elements of AR interfaces poses a unique challenge due to its extensive possibilities for information presentation and interaction. Comprehensive knowledge of the specific elements of AR interfaces that can adapt is crucial for developing iAR systems that dynamically adjust to a wide range of user contexts, ensuring an effective and intuitive interface.

RQ6. How can we describe context for an iAR interface?

To identify and adapt to the current context, an iAR system must detect the current state of the comprising contextual components. Previous research has explored various taxonomies to describe context, but these taxonomies often lack components crucial to the effectiveness of an AR interface or are difficult to apply due to the inclusion of unquantifiable elements and the absence of a clear framework for identifying implicit information. The design of iAR systems begins by establishing a taxonomy of quantifiable contextual components that directly or indirectly influence AR effectiveness.

RQ7. How can an iAR system utilize context to infer information important to the optimal design of an AR interface?

To optimize effectiveness in a given context, iAR systems must identify the impact of each adaptation within that context. The complexity of designing iAR systems lies not only in detecting the context but also in intelligently inferring the impact of adaptations to specific design dimensions on effectiveness. Outlining how to detect the user context, learn, and infer the direct and indirect influences of context in determining the optimal adaptations

can provide a roadmap for future iXR design and open the door to further research.

1.3 Proposed Approach

1.3.1 Desig and Evaluation of Context-Aware AR Interfaces in Different Contexts

Our investigation into iAR commenced with exploring the design of multiple context-aware AR interfaces in a specific occlusion-prone context and assessing their effectiveness. To address *RQ1.*, in Section 3.2, we introduce criteria for classifying AR interfaces and define Glanceable interfaces according to these criteria. Experiment 1, presented in Section 3.3, explores the design of context-aware interfaces for a scenario involving occlusion in real-world monitoring tasks.

To address *RQ2.*, in Experiment 2 (Section 3.4), we design and prototype a context-aware AR interface that can detect and adapt to the social context. Towards addressing *RQ4.*, this study explores the influence of contextual changes on the social setting. It presents our evaluation of specific Glanceable AR interface designs in two different contexts: one involving a reading task in a single-user scenario and another involving a social conversation.

The results of these experiments help validate the benefits of context-awareness in iAR interfaces. Additionally, the findings provide insight into identifying contextual components that influence AR effectiveness within these contexts and propose design principles for developing effective context-aware AR interfaces in these contexts.

1.3.2 Design Space & A Context-aware AR Placement Strategy in Different Contexts

Designing intelligent iAR systems necessitates identifying design dimensions that can adapt to the context and enhance AR effectiveness. We regard an XR object as the fundamental component of any AR interface. To address *RQ5.* and identify the adaptations that influence AR effectiveness, in Chapter 4, we explore the design dimensions that influence these XR objects' content, presentation, and interaction. In Section 4.2, we identify a robust and comprehensive list of the design elements for the visual presentation of these XR objects through a systematic methodology involving numerous iterations and refinements. We provide a roadmap for future XR design by identifying adaptations in iAR systems and present a comprehensive visual design space for XR interfaces.

Within these visual design elements, we focused on the spatial layout of XR content and investigated a specific use case of context-aware adaptations. In Section 4.4, we:

1. Present our solution to *RQ3.* and propose an *Environment-referenced* adaptive placement strategy that uses a relevant intermediary from the environment within a Hybrid Frame of Reference (FoR) for each AR object.

2. Towards *RQ4.*, evaluate the effectiveness of this context-aware placement strategy and a non-adaptive *Body-Fixed* placement strategy in four contextual scenarios varying in terms of social setting and user mobility in the environment. Our results emphasize the importance of the relevance of intermediaries to the user's focus, providing insights into designing context-aware AR for these contexts and identifying contextual components (*RQ6.*, *RQ7.*). These findings underscore the significance of context-aware interfaces, indicating that the appropriate use of an adaptive content placement strategy enabled by iAR can significantly improve task efficiency, accuracy, and usability.

1.3.3 Taxonomy of Context & iAR Framework

To inform the design and development of iAR, it is essential to describe the context comprehensively through data that can serve as input for these systems. Prior research has explored various taxonomies to describe context, but these taxonomies often miss components crucial to the effectiveness of an AR interface or are difficult to apply due to the inclusion of unquantifiable elements and a lack of a clear framework for identifying implicit information. Chapter 5 investigates *RQ6.* and *RQ7.* and explores quantifiable contextual factors that influence AR effectiveness and an approach to design iAR systems that utilize these contextual factors.

We investigate prior research on the conceptualization and characterization of context, including an examination of previous taxonomies of context proposed in the literature. Taking into account the insights gained from previous work, through an iterative refinement process, we reevaluate various quantifiable contextual components and their influence on AR effectiveness. Section 5.2, introduces a comprehensive taxonomy that describes the context of iAR systems through explicit and quantifiable input data. Section 5.3 proposes a framework for iAR to receive context as their input, infer the impact of various adaptations within the current context, and make the optimal adaptations accordingly.

1.3.4 An Architecture for iAR & An Empirical Study

iAR offers a solution to intelligently learn and identify the optimal adaptations for each specific context. Designing iAR systems requires investigating the architecture that implements our framework and utilizes our taxonomy for context-aware inference and adaptation (*RQ7.*). Additionally, developing iAR systems relies heavily on sufficient training data derived from users' authentic behaviors and adaptations to the AR interface across diverse

contexts. Chapter 6 provides a roadmap for solutions to these challenges.

In alignment with our framework, Section 6.2 proposes a potential architecture for designing iAR systems that optimize adaptations based on available quantifiable contextual components. This architecture elucidates a potential approach for implementing our proposed design space (*RQ5.*) and context taxonomy (*RQ6.*) within iAR design.

Finally, in Section 6.4, we further explore *RQ4.* and present an empirical study that records contextual factors and user-specified adaptations to the interface in a context-switching scenario. We explore the design of context-aware AR for these scenarios and investigate patterns within various user-specified AR interface designs across different contexts. These results highlight the importance of implicit inferences in future iAR systems, demonstrating how detected and inferred contexts can optimize the effectiveness of the AR interface.

1.4 Contributions

This research contributes to developing iAR interfaces that seamlessly integrate with users' daily activities and appropriately provide relevant information without distraction or interference. We contribute toward the development of these interfaces by:

(1) Identifying design principles that enhance AR effectiveness in such occlusion-prone contexts.

(2) Designing and evaluating a context-aware AR interface for social contexts.

(3) Designing and evaluating a context-aware AR placements strategy.

(4) Exploring the influence of contextual changes in the user's state and their environment and social setting on the effectiveness of various AR interface designs.

(5) Providing a Design Space for XR interfaces.

(6) Providing a taxonomy for describing context in iAR through explicit, quantifiable information.

(7) Proposing a framework and an architecture to design iAR systems that infer the potential impact of adaptations to optimize their effectiveness within a context.

Chapter 2

Review of Literature

2.1 Information Acquisition and Awareness Information

Today, people's decision-making and performance of daily and professional tasks in various contexts rely on digital information, mainly accessed through mobile phones, as the most pervasive personal computing devices [23, 31, 126]. The rise of the internet and the prevalence of mobile phones and personal computing devices provide users with any form of information at any time and place. These information acquisitions influence different aspects of people's lives. For instance, frequent use of mobile phones can be obtrusive to social interactions and perceived as impolite [1, 60]. Ubiquitous computing aims to integrate computers and digital interfaces into the user's physical world and minimize the perception of their presence [136].

"Awareness information" refers to peripheral yet important information that people remain attuned to and aware of without explicitly focusing on it [104, 137, 138]. Ideally, accessing this information requires a near-constant, yet unobtrusive presentation [56]. In recent years, numerous research projects, such as ShutEye, infoCanvas, sideshow, and scope, explored the provision of awareness information for various purposes, such as promoting healthy behaviors and monitoring the state of specific information from daily applications and the real world with minimal cognitive effort. These works utilize secondary screens and "ambient displays",

i.e., physical forms and material from the user's environment, to display information such as notifications, emails, and images to the user [7, 19, 63, 67, 104, 131, 141]. However, these methods often present limited information for special-purpose applications, require extra hardware and manual setup, and are stationary.

Unlike Virtual Reality (VR), which immerses users in a computer-generated environment, AR interactively integrates computer-generated content into the real world [5]. Recent advancements in hardware promise a future where powerful, lightweight AR glasses attain the form factor of standard eyeglasses.This introduces the potential to facilitate rapid and effortless access to extensive awareness information and ubiquitous digital content through AR [100, 136].

These AR glasses are anticipated to be worn throughout the day, providing continuous access to daily information across various contexts [43, 56, 133]. Prior research comparing AR to mobile phones, interactive videos, and printed instructions in various contexts has demonstrated AR's capability to reduce physical and cognitive load while enhancing efficiency, user experience, and long-term memory [28, 92, 130]. However, the primary challenge in achieving this effectiveness lies in AR's ability to provide access to the "right" information at the "right" time and in the "right" form.

This work explores the design of intelligent Augmented Reality (iAR) systems that enable effective access to awareness information and ubiquitous digital content. By focusing on context-aware adaptations, we lay the groundwork for developing iAR systems that dynamically adjust to the user's needs and environmental conditions, ensuring the timely and appropriate presentation and interaction of relevant information. This approach seeks to maximize the potential benefits of information acquisition through AR while minimizing its challenges and potential intrusions to the user.

2.2 AR Applications and Challenges

Various research explored the applications of AR for ubiquitous information access. Within domestic settings, such as homes, a survey explored AR's potential to enhance user experience and provide users with on-demand information, assistance, enhanced sensory perception, and entertainment in various locations [76]. LED lights on uncomplicated and lightweight HWDs were explored as a cost-efficient approach to provide omnipresent access to limited information such as notifications [24]. DiVerdi proposed ARWin to explore access to everyday information such as calendars, clocks, and web browsers in a physical workplace [33, 34]. ARBrowser was designed to provide the user with relevant information at any time [82]. Biener et al. demonstrated XR's potential to facilitate new forms of collaboration and learning, enhancing productivity and creativity for knowledge work in everyday environments [11]. While AR's potential for continuous and ubiquitous information display can facilitate information acquisition, it also presents certain challenges.

AR's effectiveness is significantly influenced by its interface design and its content, presentation, and interaction suitability within the context. Inappropriate or untimely presentation of AR content can result in information overload and occlusion of important cues within the environment [9, 77], increase cognitive load and discomfort, and reduce the user's situational awareness and performance [38, 55, 127]. Previous research has explored these challenges and the design of AR interfaces to address them in specific contexts and tasks.

The design of AR interfaces for public and social contexts plays a significant role in the adoption of this technology for ubiquitous personal computing. HWDs are inherently designed for single-user immersion, which limits their social acceptance. Given the dynamic nature of these contexts, inappropriate AR interfaces can lead to social isolation and exclusion. For instance, facial expressions are crucial in social interactions as they indicate the

intention to communicate [3, 47]. Inappropriate AR interfaces may obstruct the interlocutor's face, hindering the user's ability to interpret facial expressions [58]. This can lead to a lack of attention to the speaker, as seen with the frequent use of mobile phones during social situations [1, 60]. Furthermore, the invisibility of personal and private virtual experiences to others in close proximity, can influence public trust and raise privacy concerns [59]. Prior research has explored non-invasive AR interfaces and view management systems that maintain visual attention to interlocutors in social scenarios [66, 108]. In Section 3.4 and Section 4.4, we further investigate addressing visual clutter and facilitating access to socially relevant information while maintaining the user's visual attention to interlocutors in social AR interfaces.

The spatial setting of the real world, along with spatial relationships, depth cues, and occlusions between virtual and real-world (RW) objects, significantly impact the effectiveness of the AR interface [25, 35, 87, 91]. Occlusion can lead to reduced situational awareness, performance degradation, cognitive and mental load, and user discomfort if not properly handled [38, 55, 127]. Prior work has investigated solutions to handle occlusion in AR systems, focusing on two main types of occlusion management: (1) handling occlusion between RW objects and virtual content, and (2) handling occlusion between virtual content pieces. Numerous studies proposed techniques to preserve the credibility of registering virtual content caused by their occlusion with RW objects [10, 86, 123, 142]. Ens et al., [39] suggested using surfaces to tackle such occlusion issues. This approach relies on the availability of empty surfaces and undermines AR's advantages of placing information in any location in the 3D world. Meaningful approaches for view management and placing the virtual information have been investigated to avoid the occlusion of virtual content with each other and ensure a structured view and the visibility of virtual content [4, 8, 52, 97]. Many such approaches utilize user-triggered adaptation to the transparency, Level of Detail (LoD), and spatial layout of

XR content [40, 79, 118].

However, handling the occlusion of RW objects by virtual content and the impact of occlusion and RW spatial setting on AR effectiveness has been previously under-explored. In Section 3.3, we examine various approaches for handling the RW occlusion by virtual content and assess their effect on user's situational awareness. In Section 6.4, we examine the influence of the RW spatial setting on the user-preferred AR interface. On the other hand, in some contexts, manually resolving occlusions can be inefficient or impractical, e.g., if the user is unaware of the occurrence of occlusions. This work aims to enable iAR systems that can automatically detect and adapt to such occlusions.

Placement of virtual information in the periphery has also been explored as a general approach to providing users with easily accessible and non-intrusive information [19, 92, 93]. Maglio et al. explore the trade-offs of this approach and that this approach enhances the user's performance and awareness and improves accurate perception and comprehension [96]. Another approach to reducing visual clutter and distraction in AR is to prioritize the real world and only summon virtual content through explicit interactions. This approach empowers the user to control the presentation of their virtual content. Previous research has extensively applied this approach, with virtual content initially kept invisible, in the periphery, or at a lower level of detail [34, 89, 99, 102, 113, 114, 118, 137].

2.3 AR Design Space

The challenges of AR interface design stem from its potential to intrude on the user's attention, perception, and interaction with the physical environment through the inappropriate or untimely presentation and interaction of virtual content. This highlights the importance of having a deep understanding of both the AR design elements that influence AR perfor-

mance and the internal and external factors that shape their design and effectiveness. The various modalities and techniques for the virtual presentation (Section 2.3.1) and interaction (Section 2.3.2) of XR objects have been thoroughly studied in the literature, underscoring the broad design space of AR interfaces [15, 54]. This work investigates previous studies and identifies the AR design elements that influence AR effectiveness, consolidating and organizing this extensive information we identify the XR design dimensions and provide a comprehensive visual design space for XR (Section 4.2).

2.3.1 Presentation

Creating effective and engaging visuals relies upon various decisions about concepts, visual/sound design, and interaction design [16]. Previous work explored various innovative applications and design dimensions of XR objects to spatially place 2D & 3D objects and transition between them, visualize hierarchies, and provide persistent and portable presentation of the personal information [33, 37, 42, 82, 92]. To prioritize the real-world [28], previous work explored initiating the AR content hidden, in the peripheral, or at a lower LoD, and granting access to them through explicit gaze, hand, and head-based interactions [99, 112, 114].

To make information understandable, adaptations to visual presentation design elements such as scale, position, texture, and shape were explored[99]. Morrison et al. investigated AR's unique design elements for enhancing accessibility for visually impaired children [105]. Schmalstieg and Hollerer present design guidelines for creating effective AR visualizations using visual cues, color, and animation [122]. Laviola found that in a minimal authoring approach, using auxiliary models and color coding is preferred, and conveying the orientation of an object requires animated product models [84].

Various approaches for intuitive and seamless integration of XR content into the environment were extensively explored. Adaptations to the XR content's availability, transparency, placement, and LoD were investigated for enhancing efficiency [8, 34, 112]. Spatial layout and size were explored for minimizing intrusiveness [22, 40].

For adaptations to the XR content placement, the concept of the frame of reference, also referred to as fixation, was introduced [42]. Different Frame of References were explored for registering virtual information as world-fixed, object-fixed, head-fixed, body-fixed, or device-fixed [16]. In AR, the most commonly used settings are world-fixed, display-fixed, and body-fixed [42].

2.3.2 Interaction

Existing research extensively explores various input modalities and interaction methods and their advantages and disadvantages in different contexts and situations. Specifically, controller, head, hand, and gaze-based interactions are identified as the most commonly used and researched modalities for explicit interactions with virtual content in AR [21, 78, 109, 145]. The availability of controllers can be challenging in some situations and hand interactions can interrupt the user by occupying their hands. Another design strategy for 3D and AR interfaces is multi-modal interactions [16]. WatchThru explored pairing AR with a wrist-worn display to enable better interactions [139]. Mobile devices and smartphones have also been used to enable interaction with real-world objects [64, 120]. Hand-free head and gaze interaction have been explored as potential interactions for adaptations to the user context [34, 146]. Research has been conducted on exploring and using gaze-direction to display relevant information and adapt the virtual content's transparency and information [50, 75, 112, 113, 114, 115].

Hand input has been explored mostly for specific-purpose interactions. Xu. et al. explored the applications of hand gestures for pointing and selecting words in a text entry task [144]. Mutasim et al. investigated target selection techniques, including hand-based and gaze-dwell[106]. Due to the limitations of tracking technologies, ensuring the visibility of hands to the camera can create challenges in the use and reliability of hand gestures. Additionally, public use of hand gestures has been proven to pose issues in social acceptance [117].

Natural User Interfaces (NUI) provide interactions that are intuitive and close to everyday human behavior [90]. Gaze interaction as a potential modality to provide NUI and enhance the speed and efficiency of interaction has been gaining a lot of attention in the past few years [128].

User's gaze direction and gaze depth can indicate the user's visual attention and focus [46] and allows fast object selections, visualizations, and information acquired through a glance [14, 65]. Recently, ML models have been developed to improve the accuracy of gaze-depth, which can be influenced by vergence-accommodation conflict [85, 122, 134].

The use of blinking as a selection approach has been explored in the previous work [21, 53]. Since people blink constantly, it can create a Midas touch effect, i.e., unintended selection of content. Dwell has been used as the main approach to prevent Midas touch [71]. In Gameplay, gaze has been used for interactive tasks [36, 101]. Tung et al. explored providing input to a game in public using hand gestures defined by the user [129]. Gaze and hand coordination in multi-modal interaction methods have also proven to improve the effectiveness of object manipulation and selection. [69, 73] The distance of head movement in the depth dimension has been used to receive explicit user input in adaptive interfaces and mode-switching [34, 61] and to perform mode-switching [124]. In a study evaluating head-depth for interaction in VR, Yu et al. found head-depth to be faster than speed dwell [145].

2.4 Context-Awareness in AR

To provide an effective interface, AR interfaces must present the right information at the right time and in the appropriate form for the current context. This requires a thorough understanding of the real world and determining the appropriate content design, positioning, transparency, level of detail, etc., for information presentation. The concept of "Pervasive AR" was introduced by Grubert et al. in 2016, which refers to the integration of AR in everyday scenarios and tasks to provide "continuous, universal, and omnipresent" information [56]. This vision of pervasive AR relies on designing an automatic AR system that can adapt to the changing requirements and constraints of the user's context, enabling continuous use. Moreover, AR devices equipped with sensors offer unique opportunities for context detection. Morrison et al. used these capabilities to design an AI system that helps people with visual impairments better understand their immediate social environment. This highlights AR's potential to identify and support the user's everyday information needs [105].

Static AR interfaces provide information without considering the context in which they are used. For instance, the interfaces available on current AR devices are mostly static, meaning they provide the same information in the same form regardless of specific contextual information about the user. This lack of consideration for context can lead to AR challenges. Context-aware AR interfaces are specific-purpose interfaces designed to use components of one specific context and provide an effective interface within that scenario. However, these interfaces cannot be adapted to other contexts for which they were not designed. For example, in ARWin, the frame of reference of 2D virtual applications is world-fixed to physical surfaces around a desktop in a workspace [33]. This context-aware interface uses contextual information to identify the physical surfaces to which it attaches the digital content. HoloDoc utilizes context to create an interactive experience for users when

working on document-intensive tasks [88]; Context-awareness allows HoloDoc to identify the tangible qualities of physical paper documents and combine them with digital features like "searchability", multimedia support, and indexing.

However, a desirable interface in one context can be completely inappropriate in another. The design of an AR interface depends on the specific context in which it will be used. For mobile contexts, AR interfaces must consider social interactions and limited time for information access [126]. On the other hand, AR interfaces designed for industrial settings, such as a manufacturing or maintenance facility, must consider the presence of heavy machinery and other safety concerns and prioritize ease of use and efficiency, as they are likely to be used for extended periods of time.

Manually adaptive context-aware AR interfaces can be considered as interfaces that provide multiple choices of non-adaptive context-aware interfaces between which the user can manually switch using explicit interactions [34, 61, 113, 118]. This requires more effort on the user's part but could provide a reliable and more predictable approach, which can be beneficial in dynamic contexts and scenarios where the automatic adaptations are prone to creating distractions by making too many adaptations. Lages and Bowman proposed an interface to seamlessly move around and adapt the placement and position of the presented digital content through manual interactions on a handheld controller [79, 80]. Nirit explored AR designed to create an engaging experience for training industrial maintenance and assembly tasks. These interfaces can be tailored according to users' individual needs by manually adjusting the difficulty level or task complexity [48].

User-triggered adaptations have been shown to offer better control and predictability [119]. However, requiring the user's explicit input for every adaptation can become cumbersome and increases the user's workload. AR interfaces that can be adapted automatically detect and respond to changes in contextual information at the system level, which can increase

the level of efficiency in information access as compared to manual adaptations [28].

Recent research has shifted towards exploring "context-aware AR interfaces" that automatically adapt based on changes in context.

Context-aware interfaces can provide timely access to the "right" information in AR. Langlotz et al. utilized context in their ARBrowser to provide relevant information [82]. To provide information at an appropriate LoD, Lindlbauer explored online adaptations tailored to the user's current cognitive load and knowledge about their tasks and environment [89]. To prompt the appropriate information when needed, the use of user's conversation content with others has been suggested as a viable approach [126]. Adaptation to the amount of information presented to users based on their cognitive load was examined and corroborated that users prefer low levels of information when engaged in high-level tasks [12, 116]. Additionally, to reduce unwanted information, researchers have explored ways to minimize this obtrusive information in various contexts [102, 113, 114, 118].

Knowledge of the user's context and environment can also enable the effective presentation of the AR content. Langlotz et al. created a context-aware AR browser that can adjust itself according to different conditions and environments, e.g., automatically adjust to the environmental lighting [81]. Attentive user interfaces were proposed as systems to use the user's point of attention and provide the related services for the user [132]. Lindlbauer's online context-aware interface adapted the information display, specifically their frame of reference and the spatial location of the virtual content, to the user's current cognitive load and knowledge about the tasks and environment [89]. Lages & Bowman highlighted the significance of adapting the AR content placement strategy to avoid occlusions and accommodate activities like walking [80]. Feiner and White designed novel techniques to present and interact with spatially-situated and semantically-situated visualization in three cases, mobile visualization, objects as context, and scenes as context [140]. To maintain efficiency

and unobtrusiveness in context-switching scenarios, in Chapter 4, we explore leveraging environmental information for the **spatial placement** of XR objects. We investigate the effectiveness of our proposed context-aware placement strategy across four different contexts varying in user mobility and social setting.

Cheng et al. proposed the detection of the user's fatigue and drowsiness to automatically provide alerts to the following vehicles[20]. More recently, they also proposed a system that uses semantic changes in a mobile user's context to adapt the layout of AR windows [22]. Non-invasive AR systems were proposed for social scenarios to use context in a social scenario and change the interface layouts automatically to avoid occlusion of people's heads and interference with interpersonal interactions [66, 108].

The concept of "embodied virtuality" or "invisible computing" refers to the integration of computers into our daily lives in a seamless and unobtrusive manner, as described by the well-known quote from Weiser, "The most profound technologies are those that disappear." [135, 136, 137, 138]. AR interfaces that automatically detect and respond to contextual changes at the system level have the potential to enable invisible computing. However, this phenomenon has yet to be widely observed in the literature. This is partly because the current context-aware AR interfaces are custom-built to address specific problems within a particular scenario. This work aims to provide a framework for designing intelligent interfaces that can be adapted to various contexts. such framework must support the detection and representation of the context, its delivery to the decision-making process, and the system's reaction [32, 111]. The minimum necessary information included in context is the answer to the 5 "Ws" (Who, What, Where, When, Why) [2, 111]. Previous work has explored the context and its composing components [2, 56, 110, 111, 121]. Abowd et al. establish location, identity, time, and activity as the main context categories [111]. Other researchers focused on specific components of context and investigated objects-aware [70, 140], background-

aware [95], depth perception [44], and labeling and exploring the user's physical environment [22, 52]. Chapter 5 and Chapter 6 present our proposed taxonomy to detect and context and outline our framework to deliver and react to the context.

Chapter 3

Designing Context-Aware AR Interfaces in Different Contexts

3.1 Overview

Context-aware AR interfaces are believed to be crucial for achieving pervasive AR in the future [56]. This chapter provides our work to address *RQ1.*, *RQ2.*, and *RQ4.* on designing and evaluating context-aware interfaces in various contexts. This includes designing and evaluating context-aware interfaces for three specific contexts. The selection of contexts in this chapter aims to investigate the impact of specific contextual components on the effectiveness of different AR interface designs in resolving common AR challenges. The occlusion study discussed in Section 3.3 delves into utilizing context to address the challenge of *maintaining user's awareness of their surroundings*(*RQ1.*). The challenges of *distraction and social interactions* in AR are explored in the social and solo contexts in Section 3.4 (*RQ2.* and *RQ4.*).

Through two studies, we explore the design and effectiveness of various context-aware interfaces in each context by evaluating their usability and efficiency. These studies mainly focus on Glanceable AR (Section 3.2), a type of all-day AR interface that enables continuous virtual information display and interaction through a glance. The virtual information provided by these interfaces can be irrelevant or assistive to the user's primary task. The

first study (Section 3.3) focused on utilizing context to address Real-World (RW) occlusion in AR interfaces and evaluate its impact. In Section 3.4, distraction and social challenges of AR were studied through an ecologically valid user study in both solo and social contexts.

3.2 Glanceable AR

We suggest that AR interfaces can be classified using the following criteria.

Accessibility

Temporary interfaces are accessible upon activating a specific trigger. This trigger can be in the form of an action (glancing, summoning, etc.) performed by the user or the occurrence of a change in the state of the system (notification). *Persistent interfaces* are always accessible to the user, needless of any actions or triggers. Heads-up display interfaces in which the digital content resides at the edge of the user's field of view and are always visible as an example of such interfaces.

Number of Tasks

Single-task interfaces are those in which the user is immersed. Such interfaces are not necessarily virtual environments, but the user solely interacts with them and not any other virtual or RW content. *Multi-tasking interfaces* are used and interacted with, simultaneous with other cooperative XR interfaces or RW tasks.

Level of Focus

Primary interfaces present the user's primary task and central focus. Such interfaces interpret all user interactions as inputs. *Secondary interfaces* are available for access while the user is performing another primary task.

Level of Information Detail

Concise interfaces eliminate details and provide a to-the-point presentation of data to decrease the required time and focus for grasping information. *Verbose interfaces* present thorough and detailed information while requiring more of the user's time and focus.

Placement

World-fixed interfaces are placed in a fixed position relative to the real world [42]. They can be *fixed to an object* and move with that object, or *fixed to a global location*. *User-fixed interfaces* are placed relative to the user and follow them [42]. Such interfaces can be *body-fixed* and follow the user while maintaining a fixed orientation, or they can be *head-fixed* and follow the user's position and head orientation.

Based on the classification above, we define **Glanceable AR** interfaces as *secondary*, *concise*, and *multi-tasking* AR interfaces that are *user-fixed* and *temporary*. In other words, Glanceable AR interfaces present information/content designed to be accessed and understood with a glance while performing another primary task either in the real world or with virtual content. Many such information display interfaces can be available at once (different "apps"), and they follow the user to be available at any time and place. Using a Glanceable AR interface is analogous to using a smartwatch to quickly check information such as time, date, weather, or upcoming calendar events while doing some other primary task.

3.3 Experiment 1: AR Occlusion Management in a Real-World Monitoring Context

Context-aware interfaces must simultaneously allow effortless and active access to virtual content and maintain users' awareness of their surroundings. This work explores the effect of context awareness on both information access and user awareness of the real world. To

achieve this goal, we design multiple context-aware AR interfaces for information acquisition in RW monitoring tasks.

3.3.1 Design of Occlusion-Aware AR Interfaces for Real-World Monitoring Tasks

Interface Design Choices

The occlusion of important RW elements is a critical issue for Glanceable AR interfaces. In this section, we delineate our Design Choices for utilizing context in occlusion management techniques.

Content Prioritization

Based on the user's priority, a context-aware AR interface can prioritize RW content over the virtual or vice versa. When *prioritizing the real world*, no virtual content will occlude it by default. When *prioritizing the virtual content*, all the glanceable content is initially accessible by default.

Adaptation Mechanism for Resolving Occlusion

There are many possible mechanisms that could be used to resolve occlusion. We focus on two mechanisms: adapting the transparency level or the position of the occluding glanceable content [8]. From now on, these two intuitive adaptation mechanisms will be referred to as the transparency mechanism and the repositioning mechanism. In our interface *transparency mechanism* adapts the transparency level of the occluding glanceable content to enable the user to see the real world behind it but also retains enough visibility to enable the user to recognize its content and boundaries. The *reposition mechanism* raises the occluding glanceable content above the user's eye level to reveal the RW content behind the glanceable

content.

Level of Automation

When the glanceable content occludes an RW object important to the user, occlusion management can be performed using one of the following automation levels:

Full Automation is used when a context-aware AR interface detects the occlusion and utilizes an adaptation mechanism, i.e., repositioning or translucency mechanisms, to resolve it.

Automatic Detection is used when context-aware AR interfaces detect the occlusion and notify the user about the occlusion through a visual cue, i.e., a blinking red outline, on the occluding virtual content. These interfaces give the user control to choose whether they want to resolve the occlusion or continue to access the virtual content.

Fully Manual is used in non-adaptive glanceable AR interfaces that neither detect nor react to the occurrence of occlusion. This baseline level of automation does not introduce any distraction and gives the user absolute control at the expense of reducing their awareness of the surrounding real world.

Proposed AR interfaces

In this section, we delineate our occlusion management techniques for utilizing context to address RW occlusion by virtual content. Considering these interface Design Choices, our design space includes 12 (2*2*3) unique interfaces. However, four of these interfaces are not sensible, and they were excluded from the experiment. These four are the context-aware interfaces that prioritize the real world and were expected to automatically present the user with the required virtual content without any contextual information indicating such needs. The practicality of such fully automatic, context-aware interfaces requires further investigation.

To address RW occlusion by virtual content, we introduce eight occlusion management techniques. Six techniques are provided as context-aware interfaces, while the other two are non-adaptive glanceable AR interfaces used as our baseline.

Automatic-Reposition Interface

This context-aware AR interface initially prioritizes the virtual content. When occlusion is detected, this interface automatically moves any occluding glanceable content above eye level to allow the user to see the RW object of interest.

Automatic-transparency Interface

This context-aware AR interface initially prioritizes the virtual content. When occlusion is detected, this interface automatically increases the transparency level of occluding glanceable content to allow the user to see the RW object of interest.

Detection-Reposition Interface

This context-aware AR interface initially prioritizes the virtual content. When occlusion is detected, this interface outlines any occluding glanceable content with blinking red lines. In this interface, the user can manually raise the occluding content and see the real world behind it.

Detection-transparency Interface

This context-aware AR interface initially prioritizes the virtual content. When occlusion is detected, this interface outlines any occluding glanceable content with blinking red lines. In this interface, the user can manually increase the transparency level of glanceable content and see the real world behind it.

Reposition-VirtualPrioritized Interface

This non-adaptive, fully manual interface initially prioritizes the virtual content and allows the user to manually raise the occluding content and see the real world behind it.

Transparency-VirtualPrioritized Interface

This non-adaptive, fully manual interface initially prioritizes the virtual content and allows the user to manually make the glanceable content transparent and see the real world behind it.

Reposition-RW Prioritized Interface

This context-aware AR interface initially prioritizes the real world and positions all the glanceable content above eye level. This interface allows the user to access the glanceable content by manually lowering it when access to the virtual content is desired.

Transparency-RWPrioritized Interface

This context-aware AR interface initially prioritizes the real world and makes all the glanceable content transparent. When access to the virtual content is desired, the user can access the glanceable content by manually making it opaque.

Regardless of the interface, the user is allowed to interact manually with virtual content and adjust its visual features based on the interface's adaptation mechanism.

3.3.2 Experiment Setup

Scenario and Context

We designed a within-subjects experiment to evaluate our eight interfaces. To evaluate these interfaces in a controlled manner, we ran an AR simulation experiment in which users were asked to pay attention to the real world and virtual content simultaneously in a dual-task scenario. We simulate a future world in which smartphones are replaced by always-on wearable AR eyeglasses **(Figure 3.1)**. The participant, a babysitter, is placed in a virtual room, watching TV while six pieces of glanceable content, namely: Instagram, Fitbit,

Weather, Email, Snapchat, and News apps, are arranged around her. They are asked to perform the following two tasks:

Maintain awareness of the real world: The participant is required to watch Kevin, an autistic child, who occasionally seeks his babysitter's attention by waving at her **(Figure 3.1)**. Her primary task is to notice Kevin's wave and wave back at him immediately. To ensure that participants interact with the glanceable content using our interfaces, Kevin moves to different locations, and her view of him can get blocked by the glanceable content.

Access virtual information: Occasionally, a question (e.g., the temperature at a specific time of the day, number of unread emails, etc.) and four answer options appear on the TV **(Figure 3.1)**. The participant must access glanceable content to answer the question. It is

Figure 3.1: A screenshot of the virtual environment showcasing the Automatic-Reposition interface. This figure shows a trial in which the user is required to answer a question from the virtual content (appearing on the TV) and react to the occurrence of an RW event (the waving of the child).

important to note that in our implementation of adaptation mechanisms, we ensured that users would not be able to read the required virtual information unless the glanceable content

was opaque and in its initial position. This was to keep conditions similar between the two adaptation mechanisms, different glanceable content, and users. We did not want users' vision to determine when glanceable content was readable. Also, the background, content, lighting, and colors of different glanceable windows are all different. Therefore, a single user might be able to read from specific windows even when they are transparent/repositioned and not from others.

Procedure

This study was approved by the Virginia Tech Institutional Review Board (IRB). Upon arrival, our participants were asked to carefully read and sign the consent form. Before starting the experiment we collected our participant's demographic information and prior experience with AR. After a brief introduction, participants completed a training session in which they familiarized themselves with the environment, glanceable content, arrangement, and tasks. Each participant then experienced eight sets of trials, one for each interface. Each trial set included ten trials. A trial required either a wave at Kevin, a question to be answered, or both, and participants did not know in advance which type of trial would occur. When questions were asked, some were from virtual content that would occlude Kevin, and others were from different pieces of virtual content.

All participants used the same eight trial sets in the same order, while the order of presentation of interfaces was counterbalanced to ensure that a particular trial set or the order of experiencing interfaces did not bias the results. Before starting each trial set, the participant was trained to interact with that interface. After each trial set, the participant answered a questionnaire about that interface. Finally, we interviewed each participant on their overall experience and their preferences.

Apparatus

We used the Unity 2019.1.3f1 game engine to design a virtual environment that simulates our scenario. Users viewed the virtual environment with a Lenovo Mirage Solo standalone VR headset. We chose to simulate the AR scenario in VR in order to maintain experimental control so that all participants experienced the experiment in exactly the same way. This approach has been used effectively in many prior studies of AR. Participants used the Mirage Solo's three-degree-of-freedom handheld controller to interact with the virtual content. We rendered a ray emerging from the front of the virtual controller, and users could "click on" a piece of virtual content by intersecting it with the ray and pressing the controller's touchpad button. For detecting an occlusion in context-aware interfaces, we monitor the collisions between the glanceable content and a frustum drawn from the user's eye center to the corners of the important RW object (Kevin). Once a piece of glanceable content enters this frustum, occlusion is detected, and its state changes according to the interface.

3.3.3 Evaluation Metrics

We evaluate our proposed interfaces and their Design Choices based on user performance and self-reported user experience. *Perceived Ease of information access* is evaluated through the participant responses to two post-interface survey questions that used five-point scales. Using that specific interface, we asked 1) how easy they found it to access information from the real world and 2) how easy they found it to access information from the glanceable content. *Preference* is evaluated through the participant responses to the final interview questions about their preferred interface dimensions and preferred interface overall.

We logged the user's interactions and used them to evaluate the user's performance. The wrong answers to the questions and the excess wave-backs to Kevin in each trial were logged

to measure *Accuracy*. The participant's clicks on the glanceable content were logged and used to calculate the *Average operation count* for each interface. The start and end times of each task were logged and used to calculate *Task performance time*.

3.3.4 Hypotheses

Based on the five parameters we are exploring in this experiment, we developed and tested the following hypotheses:

Content Prioritization

Due to the dependency of the primary task on the initial visibility of the RW object of interest in this context, considering the content prioritization, we hypothesize that:

$\boldsymbol{H1_{study1}}$**.** Prioritizing the real world will be preferred.

$\boldsymbol{H2_{study1}}$ In terms of perceived easiness: **a)** Prioritizing the real world will be better than prioritizing the virtual when accessing information from the real world. **b)** Prioritizing virtual content will be better than prioritizing the real world when accessing virtual information.

$\boldsymbol{H3_{study1}}$ In terms of task performance time: **a)** Prioritizing the real world will be faster than prioritizing virtual content when accessing information from the real world. **b)** Prioritizing virtual content will be faster than prioritizing the real world when accessing virtual information.

Adaptation Mechanism

$\boldsymbol{H4_{study1}}$ Reposition mechanism will be preferred since this mechanism allows the user to have both virtual content and the RW object of interest in their view.

Automation Level

Our fully automatic context-aware interfaces always keep the RW object of interest visible. On the other hand, fully automatic context-aware interfaces require the least amount of interaction and manual operation, while non-adaptive interfaces require the most. For this reason, we hypothesize that:

$\boldsymbol{H5_{study1}}$ In terms of preference: **a)** Full automation will be preferred to the other two levels of automation. **b)** Automatic detect will be preferred to the full manual.

$\boldsymbol{H6_{study1}}$ In terms of perceived easiness: **a)** Full automation will be better than the other two automation levels for accessing information from the real world. **b)** Automatic detection will be better than fully manual for accessing information from the real world.

$\boldsymbol{H7_{study1}}$ In terms of task performance time: **a)** Full automation will be faster than the other two automation levels when accessing information from the real world. **b)** Automatic detection will be faster than fully manual when accessing information from the real world.

$\boldsymbol{H8_{study1}}$ Full automation will result in a lower average operation count than the other two automation levels.

Interface

$\boldsymbol{H9_{study1}}$ context-aware Automatic-Reposition will be the best interface in terms of **a)** user preference, **b)** perceived easiness, **c)** task performance time and **d)** average operation count. This is because this interface is fully automatic and keeps both virtual content and the RW object of interest visible at all times. Thus, user interaction is minimized for achieving goals.

3.3.5 Participants

We gathered data from 36 participants. Data from four participants had to be discarded due to missing data. The participants were recruited from the student body of Virginia Tech. Of the 32 remaining participants, 20 had imperfect eye vision and used eyeglasses. 8 of the

participants were female, and 8 of them were graduate students. One of the participants had never used any type of AR/VR before, 12 had limited experience, and 6 of them expressed an expert level of experience with AR/VR. Since we used eight interface orderings based on a Latin Square, we had four participants use each interface order.

3.3.6 Results

We conducted a series of analyses to test our hypotheses. For our ordinal data (i.e., perceived ease of information access) gathered from the post-interface surveys and interviews, we performed Chi-square approximations to the one-way Wilcoxon signed-rank tests.

For the quantitative data, we performed multiple two-way analyses of variance (ANOVAs) on task performance time, with *trial-type* as the first independent variable and *prioritization*, *adaptation mechanism*, *automation level* or *interface* as the second one, and report the Least Squares Means(LSM) and Standard Error(SE). We also performed multiple one-way ANOVAs on operation count with *prioritization*, *adaptation mechanism*, *automation level* or *interface* as the independent variable and reported the Mean(M), and Standard Deviation(SD). We used an $\boldsymbol{\alpha}$ level of 0.05 in all significance tests.

Preference

The majority (84.4%) of our participants prefer prioritizing the RW content in this scenario, and transparency was also the preferred adaptation mechanism for 84.4% of them. In terms of automation level, 93.8% of our participants consider full automation the best, and 56.3% of them preferred automatic detection over fully manual.

When asked which interface they preferred overall (**Figure.3.2**), all participants chose one of the four interfaces in which the child is initially visible. Of these, 75% preferred one of the interfaces that use the transparency mechanism, with 50% choosing the Transparency-

RWPrioritized interface as the best one.

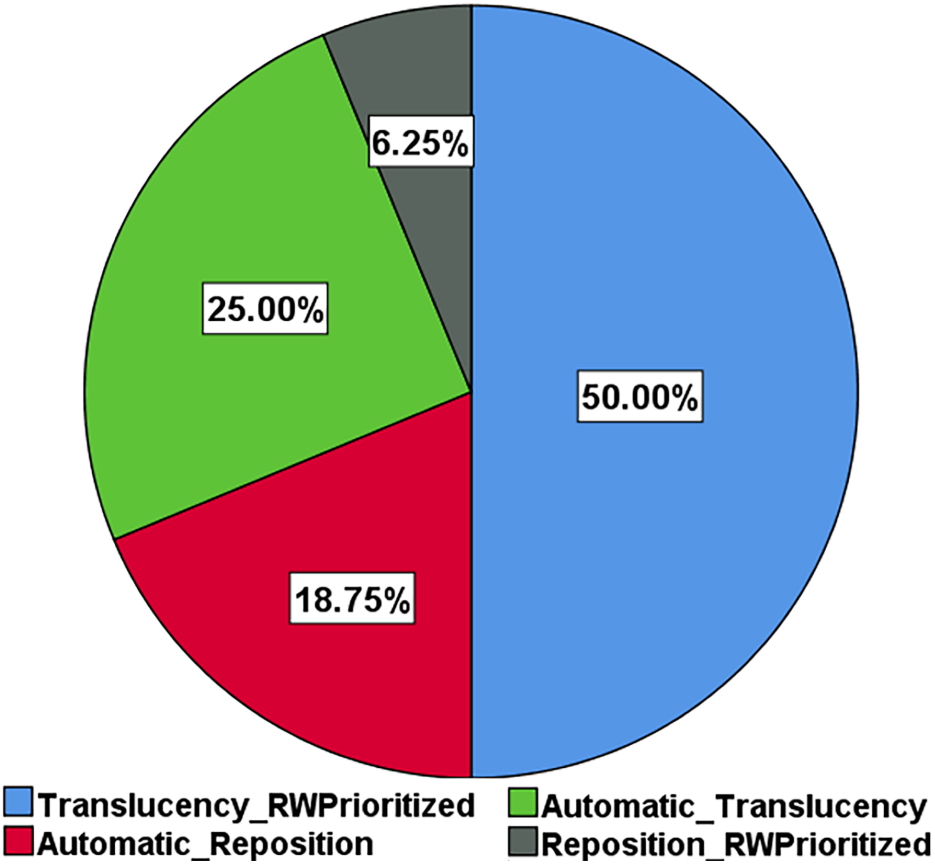


Figure 3.2: User's choice of best interface

Perceived Ease of Information Access

Figure 3.3 illustrates the effect of each interface dimension on the mean of perceived easiness when accessing information. More than 85% of participants ranked the level of ease for RW information access (i.e., monitoring Kevin) when prioritizing the real world as 5(easiest). A Wilcoxon signed-rank test yields a significant effect (p=0.0001) of prioritization on ease of RW information access. Prioritizing the real world increases the ease of RW information access (M: 4.81, SD: 0.47) rather than prioritizing virtual content (M: 3.78, SD: 1.16). However, we did not find any significant difference in the ease of virtual information access caused by content prioritization.

We found a significant effect (p=0.0069) of the adaptation mechanism on the ease of virtual information access. The transparency mechanism results in a higher ease of virtual information access (M: 4.36, SD: 0.73) than the reposition mechanism (M: 4.07, SD: 0.86).

We also found a significant effect of automation on the ease of RW information access, with significant differences between full automation and fully manual($p < 0.0001$), and between full automation and automatic detection ($p < 0.0001$). Full automation results in higher ease of RW information access (M: 4.66, SD: 0.57) compared to automatic detection (M:3.9, SD: 0.9) and full manual (M: 3.8, SD: 1.3). However, no significant difference was found for this measure when comparing automatic detection with fully manual.

No significant difference was found for ease of virtual information access when comparing different interfaces. However, we found a significant effect ($p < 0.0001$) of the interface on the ease of RW information access. interfaces that made the child initially visible were ranked consistently higher than the other interfaces for this measure.

Task Performance Time

We found a significant effect ($p < 0.0001$) of trial type on task performance time, with trials that only required the user to wave back at Kevin significantly faster than all other trials. However, we did not find any significant interactions of trial type with any of the interface dimensions.

We did not find any significant effect of prioritization or automation on task performance time. In terms of adaptation mechanism, however, transparency (LSM: 6.28, SE: 0.1) was significantly faster ($p = 0.0432$) than repositioning (LSM:6.56) for tasks that involved accessing information from virtual content.

We observed a significant effect (p= 0.0089) of the interface on task performance time when accessing information from virtual content **(Figure 3.4)**. Pairwise comparisons showed that

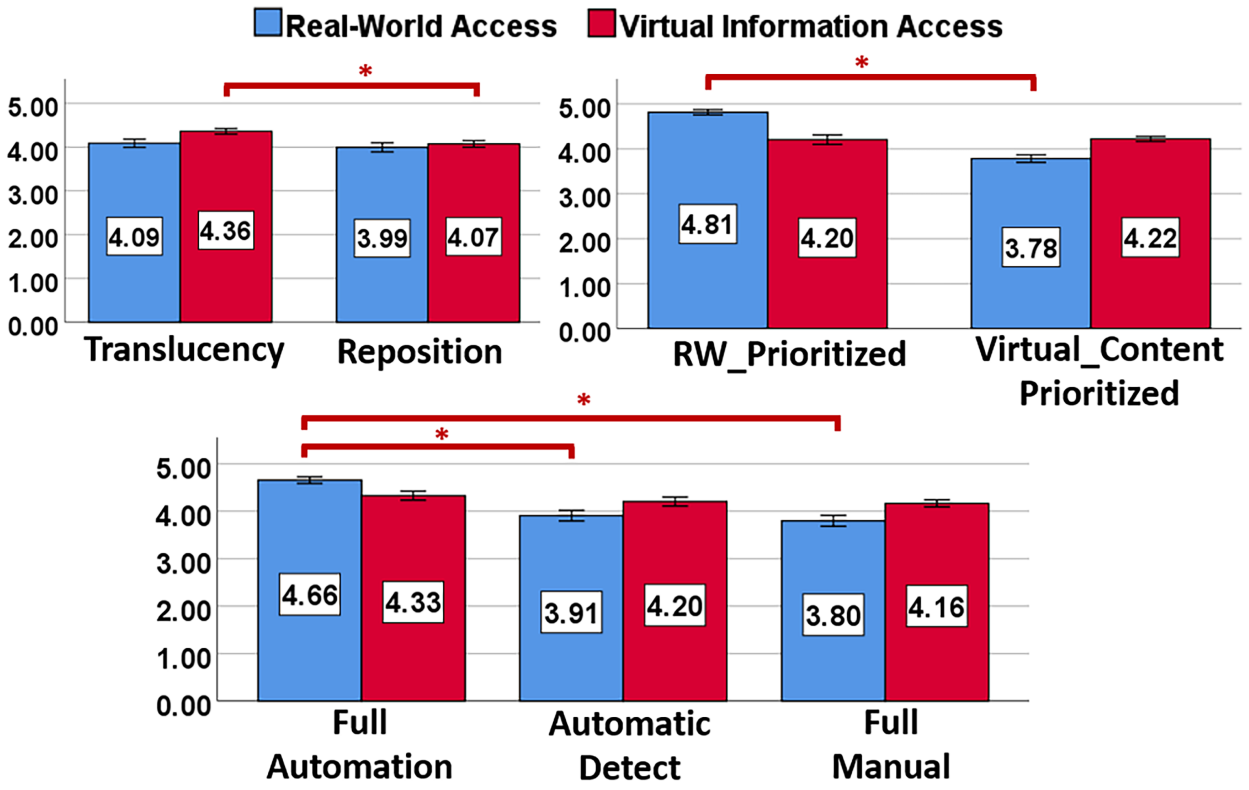


Figure 3.3: Mean of perceived ease of accessing information from the real world and virtual content based on interface dimensions

Reposition-VirtualPrioritized (M: 7.12, SE: 0.198) was significantly slower than six of the other interfaces, and it was the least efficient in terms of task performance time on virtual information access.

Operation Count

We found a significant effect ($p < 0.0001$) of prioritization on the average operation count. Prioritizing the real world results in a significantly lower average operation count (M: 1.67, SD: 1.05) than prioritizing virtual content (M: 2.87, SD: 2.29). When prioritizing the real world, the average operation count for accessing information from the real world is approximately zero. No statistically significant difference between the average operation count using transparency (M: 2.62, SD: 2.09) or repositioning (M: 0.53, SD: 2.15) was found. As expected, the average operation count of full automation (M: 1.46, SD: 1.84) is significantly

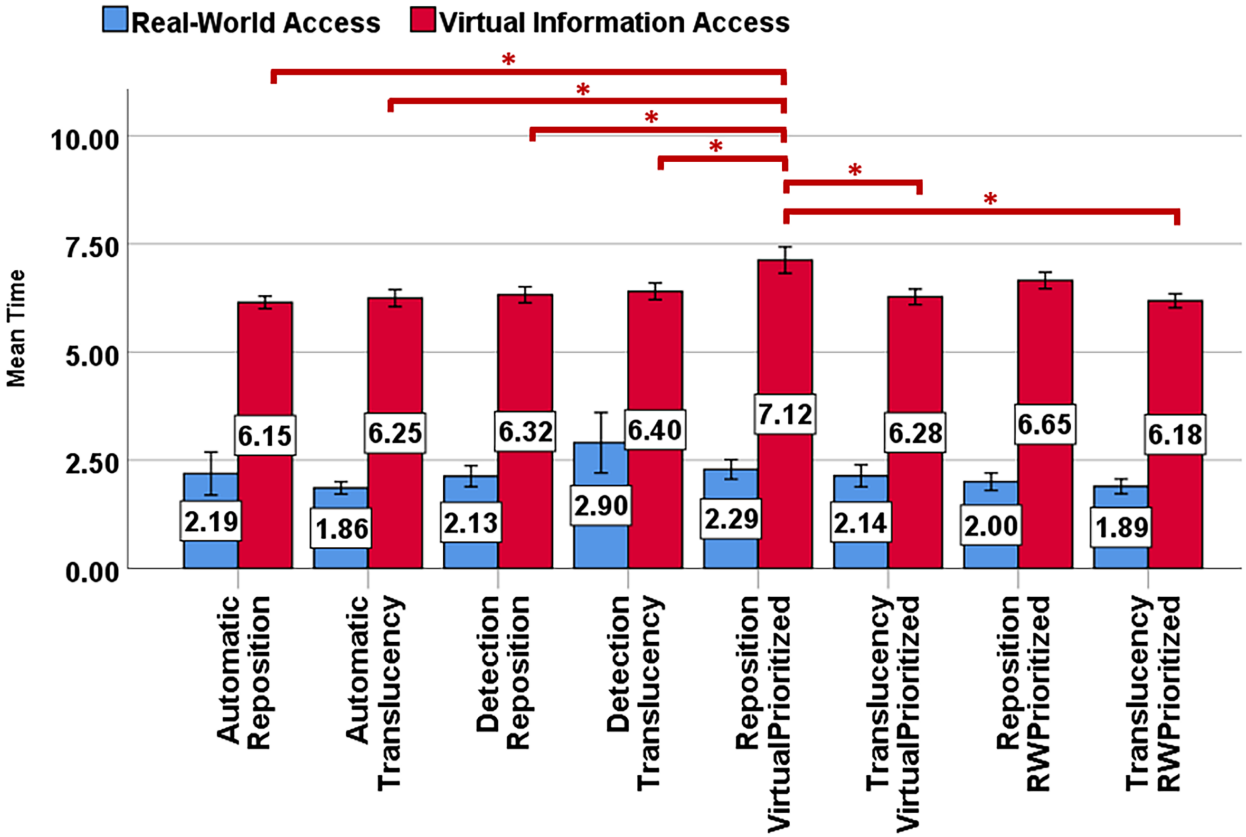


Figure 3.4: Time for accessing information from the real world and the virtual content based on interface

different than that of automatic detection (M: 2.81, SD: 1.56, $p < 0.0001$) and full manual(M: 3.01, SD: 2.28, $p < 0.0001$). The average operation count also varies significantly ($p < 0.001$) among all interfaces **(Figure 3.5)**.

3.3.7 Discussion

We hypothesized that since, in the context of this scenario, the user is primarily required to monitor the real world, prioritizing the real world would be preferred ($\boldsymbol{H1_{study1}}$), and would also facilitate ($\boldsymbol{H2.a_{study1}}$) and accelerate ($\boldsymbol{H3.a_{study1}}$) the performance of this task. Both hypotheses $\boldsymbol{H1_{study1}}$ and $\boldsymbol{H2.a_{study1}}$ were supported. However, it is noteworthy that

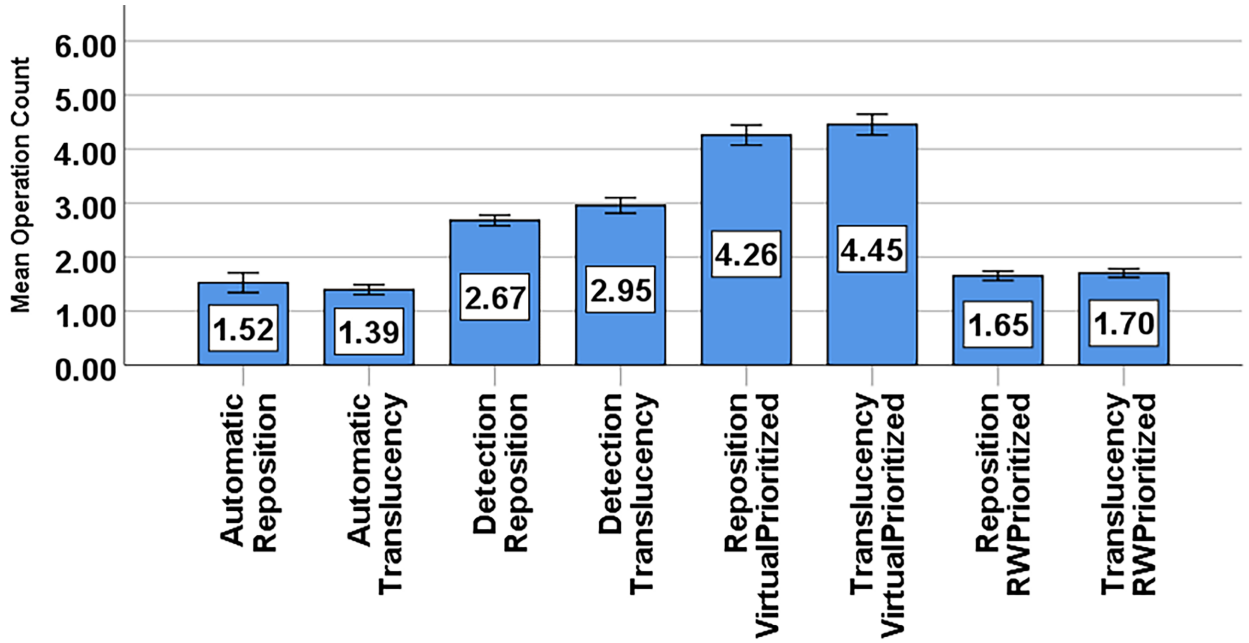


Figure 3.5: Average operation count based on an interface. Most pairwise comparisons are significant; they are not shown in the figure to avoid visual clutter.

based on the interview, this result is not merely due to the primary task and context of this experiment. Many participants expressed that prioritizing virtual content introduces unwanted visual clutter in the periphery, which feels cumbersome and distracting, and reduces the feeling of awareness, security, and control. Consequently, many participants indicated that considering the specific AR content included in this study (e.g., social media) and the irrelevance to the RW task and lack of urgency that they associate with them, no matter the context, they would not prefer prioritizing virtual content. This also may explain why $\boldsymbol{H2.b_{study1}}$ was not supported: the participants did not perceive accessing virtual information when prioritizing virtual content as easier.

None of the parameters caused any significant difference in the task performance time when accessing information from the real world; Thus $\boldsymbol{H3.a_{study1}}$, $\boldsymbol{H7_{study1}}$ and $\boldsymbol{H9_{study1}.c}$ were rejected. A likely reason for this result is that the primary task in this scenario is focused on the real world, and most participants preferred prioritizing the RW object of interest.

Our observations and the interview showed that most participants prioritized the real world manually, if necessary, once the trial started, even before they were required to access any information from the real world. For example, with the transparency-VirtualPrioritized interface, participants would typically click on every window right at the beginning of the trial to make them transparent so that they could see Kevin no matter where he was. Thus, all the interfaces had similar performance results when responding to Kevin's waves. We speculate that it was for the same reason that the interfaces with virtual content prioritized had no performance advantage over the other interfaces when accessing virtual content, and $\boldsymbol{H3.b_{study1}}$ was not supported.

Based on our results, the majority of participants prefer the transparency mechanism, so we reject $\boldsymbol{H4_{study1}}$ We suggest that this was partly because we designed the repositioning mechanism in such a way that the answers to questions were not readable when the content was in the raised position. It may also be related to the fact that transparency was immediate, while repositioning took some time. The interview revealed two additional reasons for this preference. First, when using the repositioning mechanism, the virtual content is not all at eye level. This required the user to tilt their head and point up when interacting and accessing information from the virtual content. Second, transparency resolves visual clutter and distractions, which decreases users' cognitive task load and information overload.

As we anticipated in $\boldsymbol{H5_{study1}.a}$, $\boldsymbol{H6_{study1}.a}$, and $\boldsymbol{H8_{study1}}$, full automation was superior in several ways to the other two automation levels. 93.8% of participants preferred full automation, accessing information from the real world was perceived as easier with full automation, and the average operation count was smaller when using full automation compared to the other two. Unless manually specified otherwise, full automation always makes the important RW object visible, which lowers the physical task load. However, the interview results revealed another explanation for some of these results. Automatically changing

the visual appearance (transparency/position) of the occluding virtual content draws the user's attention toward the visually different glanceable content and, thus, the RW object of interest. This increases the user's awareness of important RW events and facilitates their task.

A majority (56.3%) of participants preferred automatic detection over fully manual. This by itself is not strong support for $\boldsymbol{H5_{study1}.b}$. However, it is notable that during the interview, many participants who chose full manual over automatic detection acknowledged that automatic detection increased their awareness of the real world while giving them control over the system. Many expressed that a more appealing choice of interface design components for automatic detect interfaces (i.e., less stress-inducing colors, more noticeable blinking speed and highlight line thickness, or the ability to manually deactivate the highlights) would swing their vote towards it. This, along with the fact that automatic detection still requires manual interaction when accessing information from the real world and is not easier to perform compared to full manual, may explain why $\boldsymbol{H6_{study1}.b}$ was not supported.

As opposed to our expectation, only 18.8% chose the Automatic-Reposition interface as the best interface. In terms of perceived easiness, this interface was the second best when accessing information from the real world and was not any different than the other interfaces when accessing virtual information. It was also not significantly faster than other interfaces, nor was it the best interface in terms of average operation count. These results lead us to reject $\boldsymbol{H9_{study1}.a}$, $\boldsymbol{H9_{study1}.b}$, $\boldsymbol{H9_{study1}.c}$, and $\boldsymbol{H9_{study1}.c}$.

Overall, the majority (75%) of participants were concerned about minimizing visual clutter and distraction, which is possible by transparency more than any other dimension of the occlusion management interfaces. The Automatic-transparency interface can lower the task load, decrease unwanted visual clutter and distraction, and even help increase user awareness of important RW events. Yet, the fact that 50% of the participants chose the transparency-

RWPrioritized interface along with their interview comments may indicate that the participants primarily care about the visibility of the real world and minimizing visual clutter and information overload which feels more natural and gives the user a feeling of control.

During the interview, the participants frequently suggested an interface that prioritizes the real world and automatically makes the required glanceable content available. The practicality of such interfaces relies on the existence of iAR systems that automatically detect the virtual content that the user needs. Our research investigates the use of context (Chapter 5) to infer (Chapter 6) user's information needs and enable such systems.

3.4 Experiment 2: AR Information Acquisition in Social and Solo Context

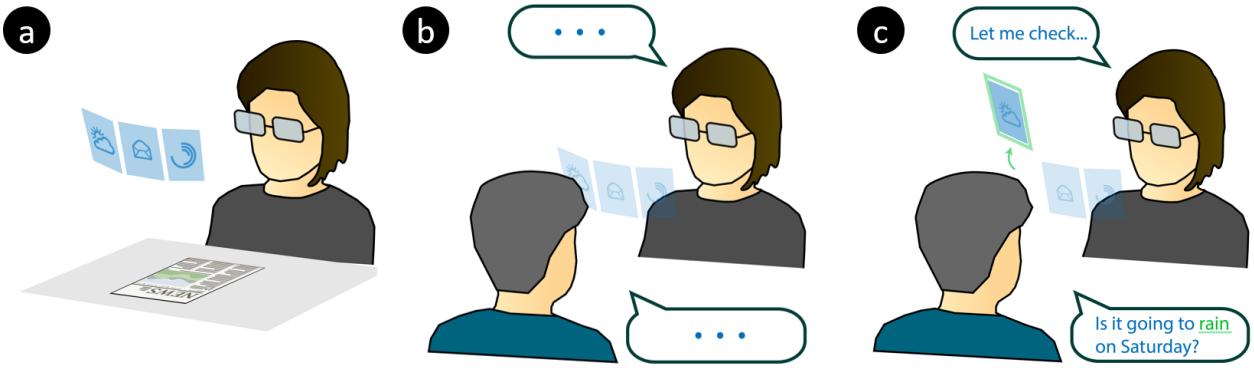


Figure 3.6: We propose a context-aware augmented reality interface that minimizes intrusiveness while providing fast and easy information access during both solo and social contexts: (a) In a solo context, the virtual content is glanceable and opaque when not occluding any object of importance in the real world; (b) In a social context, the real-world is prioritized, keeping the virtual content transparent when not needed, avoiding occlusion, distraction, and visual clutter; (c) In a social context, a piece of virtual content relevant to the ongoing conversation moves up and turns opaque to provide fast and easy access without blocking the conversation partner's face.

This experiment focuses on the evaluation of a Context-Aware Glanceable AR interface that

adapts between two common usage contexts: a static, solo context and a dynamic social context. As a proof-of-concept, we prototyped an interface that can adapt to provide a balance of efficiency and unobtrusiveness in these contexts. To our knowledge, our work was the first to prototype a socially aware and supportive Context-Aware AR interface that facilitates conversation between two users and evaluates it in a highly ecologically valid conversation.

3.4.1 Research Questions

To explore the benefits and limitations of Context-Aware Glanceable AR, we evaluated our prototype interface in both solo and social contexts. We also evaluated a naive Glanceable AR interface (without context awareness). A mobile phone interface was included as a baseline since it is the most pervasive tool for information access. Our study aimed to address three research questions. The first two questions address the benefits of Glanceable AR over mobile phones for information access, while the third question addresses the benefits of context awareness.

$\boldsymbol{RQ1_{Exp2}}$. *Do Glanceable AR interfaces improve the speed and convenience of information access in different contexts, compared to mobile phones?* The transition to Glanceable AR interfaces is justifiable only if these interfaces offer more efficient and appealing information access compared to the current gold standard of mobile phones. Therefore, we designed the study to compare the speed and convenience of information access using Glanceable AR and mobile phones and to explore the effect of user context on these measures. Is one always interface more efficient and convenient, regardless of the user's context, or does the context influence the benefits of one interface over the other?

$\boldsymbol{RQ2_{Exp2}}$. *Compared to mobile phones, does information access using Glanceable AR improve*

the user experience and focus on the primary task in a solo context? The physical interaction requirements and time-consuming process of information access using mobile phones can distract users' focus from their primary tasks. We compared Glanceable AR and mobile phones in terms of user experience and distraction from the primary task when accessing information.

RQ3$_{Exp2}$. *Do the automatic adaptations in Context-Aware Glanceable AR improve the user experience and reduce social intrusiveness during information access in a social context?* In a dynamic social context, naive Glanceable AR may result in occlusion, distraction, and loss of social awareness, and its advantages compared to mobile phones may be lost. We designed our study to compare user experience and social awareness with three interfaces (Context-Aware Glanceable AR, naive Glanceable AR, and mobile phone) for information access during a social conversation in order to evaluate whether Context-Aware AR can balance efficiency and awareness, retaining the advantages of Glanceable AR while mitigating its weaknesses.

3.4.2 Design of Context-Aware AR Interfaces

Interface Design for a Solo context

In the solo context, the user is seated, stationary, alone, and focused only on a single object (e.g., a book or laptop) in the real world. The RW object of interest is somewhere below the user's eye level. Therefore, virtual content at the user's eye level will be unobtrusive and not occlude the object of interest. The interface can also prioritize virtual content (by making it opaque) without distracting the user. This interface for the solo context (Figure 3.6a) allows efficient information access (via a quick upward glance) without causing any obstruction or visual clutter.

Interface Design for a Social context

Figure 3.7: Screenshots of our prototype in the social context. (a) When in conversation, the real world is prioritized by keeping the virtual content's level of transparency high. (b) The desired virtual content is provided automatically based on the conversation content without occluding the conversation partner's faces.

The interface for the solo context would be completely inappropriate for a social context, however. In the social context, the user is standing and conversing with one or more others. The conversation partners may move, leave, or join the conversation in progress. Furthermore, some conversation topics may necessitate information access. Thus, the context-aware interface for this context should also be dynamic and intelligent, adapting to both the presence of conversation partners and the topics of conversation. To design a Context-Aware Glanceable AR interface for the social context, we propose three design considerations (DCs). Using the user context, DC1 and DC2 aim to address the existing challenges of AR (i.e., visual clutter and obtrusiveness), while DC3 aims at improving information access efficiency.

DC1. **RW prioritization.** Constantly visible virtual content produces visual clutter and distraction when engaged in dynamic RW tasks [28]. Our design prioritizes the real world by keeping the virtual content's level of transparency high (Figure 3.7a). This allows the user to see the outline of the content and know where to find the information when needed but ignore it otherwise.

DC2. **Support for viewing social cues.** Facial expressions play an important role in social interactions [3]. For example, ostensive gestures such as the eyebrow flash can indicate the intention to communicate and are of great importance to the receiver [47]. If virtual content blocks the interlocutor's face, the user will miss their facial expressions, just as they would if they were looking down at their mobile phone. Similar to Höllerer et al., [66], we propose that in conversation, the AR interface should not occlude other people's faces, leaving space for sufficient attention and social awareness (Figure 3.7b). However, our interface moves the virtual content away only when it occludes the faces of people actively engaged in a conversation with the user. This reduces over-generalization of the no-occluded-faces rule and enables users to place virtual content anywhere they want without undesired adaptations.

DC3. **Support for socially relevant information access.** Since a significant number of information inquiries arise from something mentioned by others, the conversation is a context that prompts many information acquisitions [126]. Based on the content of the conversation, an intelligent interface can support the conversation by automatically providing the desired information, leading to a smooth and less interrupted conversation experience (Figure 3.7b). Our interface supports this by automatically making relevant apps opaque based on speech recognition during the conversation and placing those apps near the interlocutor's face without occluding it.

Figure 3.6 illustrates our proof-of-concept context-aware Glanceable AR interface in a social and a solo context. Our system can adapt virtual content placement and transparency automatically to reflect both the static and dynamic aspects of the user's current context.

3.4.3 Experiment Setup

Solo Context

The solo session was designed to simulate situations in which a user is doing some RW activity alone in a static environment and is motivated internally to access digital information. For example, the user may be reading a book when she thinks to check the current weather. We refer to such questions as Self-Triggered Questions (**STQs**). In our experiment, the user was asked to read an article. While reading, at random times, they were asked to use the available interface to answer an STQ (played from a computer). The participants were asked to say the answer and immediately resume reading. Each user performed the session twice: once using the mobile phone and once using Context-Aware Glanceable AR. The order of interfaces was counterbalanced. With both interfaces, the user had access to three different applications: Email, Exercise, and Weather. The answers to the STQs were different each time (i.e., the content in the apps changed throughout the session). Still, the questions from each application were always the same ("How many VIP emails do you have?", "How many zone-minutes of exercise have you had today?" and "What is the high temperature on Sunday?"). Each question was asked twice, with six questions per session.

Social Context

The social context was designed to represent an authentic social conversation. In each session, the experimenter would start a conversation with the participant about a specific topic ("Hiking places in the area," "Local food and restaurants," or "Weather"). Throughout, the experimenter asked them questions to keep them engaged in a two-way conversation. Each participant performed the session three times: one time with each of three different interfaces: mobile phone, non-adaptive Glanceable AR, and Context-Aware AR. The order of

interfaces was counterbalanced with a Latin Square design. In this context, the participants were asked to perform three tasks:

Task 1: Answer self-triggered questions. During the conversation at random times, the participant answered an STQ, played from a computer, just as they did in the solo context. The participants were asked to say only the answer to the question and immediately resume the conversation without interruption. Each STQ was asked twice, with six STQs per session.

Task 2: Answer conversation-triggered questions. During the conversation, the experimenter would ask the participant to answer specific questions, which we call Conversation-Triggered Questions (**CTQs**), using the interface. The answer to these CTQs was different each time, but the question from each application was always the same: ("How many unread messages do you have on your email?", "How many steps have you walked so far today?", and "Do you know by what percentage it is going to rain on Saturday?"). Each CTQ was asked once, for a total of three CTQs per session.

Task 3: Maintain awareness of the conversation partner's facial expression. At certain times during the conversation, the experimenter would wink at the participant. The participants were asked to raise their hand to indicate that they saw the wink. During each session, the experimenter winked at the participant seven times total, three of which were right after a question was asked. This task allowed us to evaluate how the interface affected social awareness.

Interfaces

To answer STQs and CTQs during the experiment, participants used three interfaces for information access.

Mobile Phone Interface: The participant was given a mobile phone placed face down

on a table in front of them. To answer a question, participants had to pick up the phone and turn the screen on by swiping up the display (Figure 3.8a). Once turned on, the phone displayed the icons of three apps. Users then tapped on the appropriate app to display a screen from which they could read the answer to the question. To simulate today's most commonly used information access method for answering an unexpected question and to keep conditions similar for answering all questions, we asked participants to go back to the home screen, turn the mobile phone's screen off, and place it face down on the table after answering each question. In our experiment, the user only had access to three applications and was not required to use any password, fingerprint, or face recognition to unlock the mobile phone, making it faster than similar RW situations.

Glanceable AR Interface: Glanceable AR interfaces are "secondary, concise and Multi-tasking AR interfaces that are user fixed and temporary" [28]. The Glanceable AR interface continuously displayed the three applications at the eye-level of the user. The apps had an identical appearance to those displayed on the mobile phone, but an AR HWD rendered the apps, and all three apps were visible at once. In the social context, since we were interested in investigating the worst-case scenario of using Glanceable AR in a social context, the scene was set so that the virtual content would block the conversation partner's face. Previous work has explored different input modalities, for example, gaze, hand, or head-based interactions, to toggle content visibility in the interfaces for a glance [93, 94]. To allow users to decide whether to view the face or the content, users could use a hand-based input (i.e., "air-tap") on individual apps to toggle them between transparent and opaque modes (Figure 3.8b).

Context-Aware Glanceable AR Interface: Our context-aware interface actively detects whether the user is in a social context using face and speech detection. Specifically, if our interface detects speech and one or more faces in front of the AR user, this is interpreted as a social context. In the solo context, when no social conversation is detected, the Context-

Aware AR interface prioritizes the virtual content (i.e., displays it as opaque, at the user's eye-level), so it is identical to the non-adaptive Glanceable AR interface (Figure 3.6 (a)). Thus, there were only two interface conditions in the solo context sessions.

When the social context is detected, the Context-Aware AR interface automatically switches to RW prioritized mode (i.e., the apps become transparent when not needed (Figure 3.6b). Based on the content of the conversation detected by the speech recognition system, the system identifies when the user needs to access information from one of the apps (i.e., when a CTQ is asked) and automatically makes the app opaque. Since the STQs simulated questions in the user's head, the interface has no knowledge of the user's need for that information. Thus, the user must manually click on the intended app to answer STQs. To simulate these questions in our experiment, the wording of our STQs were designed not to be recognized as associated with any of the available apps. If a piece of virtual content is opaque and blocks the conversation partner's face (the face detection algorithm gives the position of the face), the interface automatically moves it above the face (Figure 3.6c). Finally, all opaque apps automatically turn transparent seven seconds after the last time they were activated.

Apparatus

For the mobile phone interface, we used a Samsung Galaxy S7 device. We designed a web app for the mobile phone interface in Python 3 using the Flask package. The web app was running on a Virtual Machine (VM) Server at our university (12 Core, ARCH: x86_64, RAM 32 GB, OS: Ubuntu 20.04).

For the AR interfaces, the experiment used a Microsoft HoloLens (1st gen) AR HWD. This device is wireless and has a resolution of 1268 x 720 per eye and 30 deg x 17 deg FoV. The software was developed via Unity 2018.4.22f1 using the MRTK toolkit provided by Microsoft.

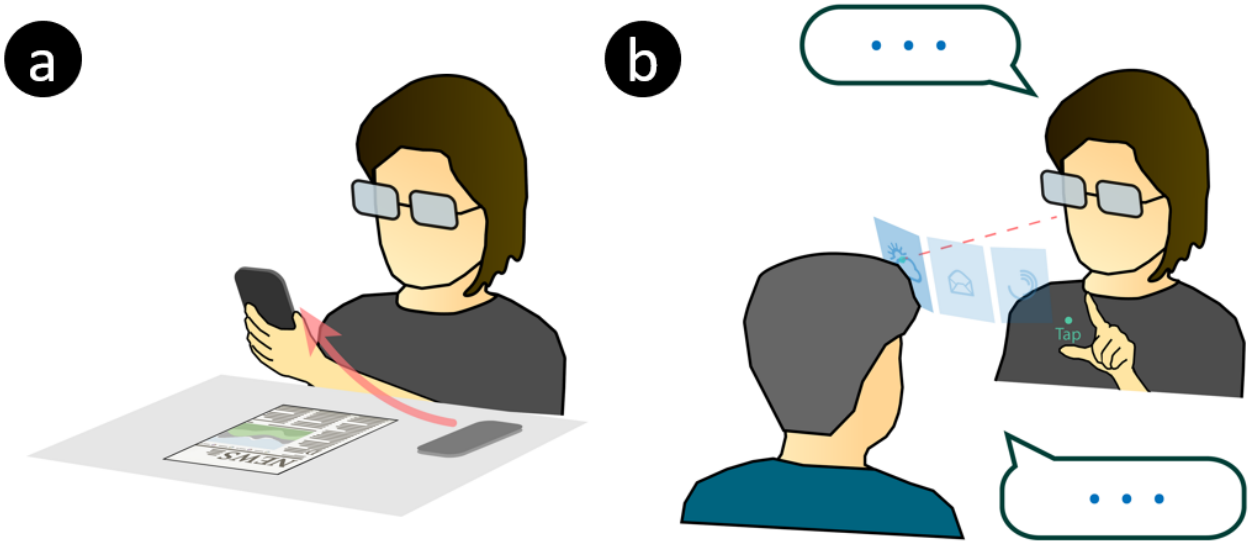


Figure 3.8: Conditions compared to Context-Aware Glanceable AR in the experiment. (a) Mobile phone interface. (b) Non-adaptive Glanceable AR interface, in which the user can air-tap an app to turn it opaque or transparent.

For the Context-Aware AR interface, we developed a face detection system using Python 3. We used the OpenCV Deep Neural Network (DNN) Module, with caffemodel as our pre-trained neural network. We ran the face detection on our VM Server. Using Python 3 sockets, we made a low-level network connection to the HoloLens device to receive the image from the HoloLens camera and send back the bounding box around each detected face. On the HoloLens, we used UnityEngine.XR.WSA.WebCam to capture a photo from the HoloLens camera. The UnityEngine.The networking module was used to send the image to the server and receive the face detection results. By transforming the bounding box of the detected face from the camera space to the screen space, we detected whether any of the virtual apps was blocking a detected face. Using MRTK speech recognition and Dictation Recognizer, we detected whether someone was speaking and, if so, whether the content of the conversation contained keywords (e.g., “rain” or “exercise”) related to one of the available apps. This fully implemented context-aware interface was used during the user study without any manual intervention by the experimenter.

All the sessions were video-recorded by a Logitech C615 HD Webcam with 1080p resolution for comprehensive observation of user behaviors.

Procedure

In this within-subjects study, the experiment was divided into two parts. Each participant conducted five different sessions. In the first part, participants completed two sessions, one with each interface, in the solo context. In the second part, participants completed three sessions, one with each interface, in the social context.

The IRB of our university approved this study. Upon arrival, we welcomed participants and asked them to read the consent form, if they had not already, and sign it. We then asked them to fill out a pre-study questionnaire to collect demographic information and prior experience with AR and to answer one question regarding their tendency to adopt AR for daily information access. Next, the participants were introduced to the experiment background, hardware, and the sessions and contexts involved in the study. When participants had no further questions, we introduced the first part of the study, the solo context, and the tasks involved in this part to the participant. Afterward, participants experienced each interface (Context-Aware AR and mobile phone) in separate sessions. After the first part, the second part of the study, the social context, and the tasks involved in this part were introduced to the participants. Afterward, the participant experienced each interface (Context-Aware AR, non-adaptive Glanceable AR, and mobile phone) in separate sessions.

Before each of the five sessions, the participants completed training to familiarize themselves with the tasks and interface. After each session, they were asked to fill out a custom-designed questionnaire and the UEQ on a computer. In our post-context interview, after each part of the study, the participants were asked about the interfaces they used in that context, their

preferences, and the advantages and disadvantages of each interface.

After finishing both parts (all five sessions), participants were interviewed about their overall experience in both contexts, the benefits and disadvantages of each interface in each context, and their tendency to adopt AR for daily information access. The entire experiment took about 75 minutes. Participants were allowed to take a break anytime in between sessions.

3.4.4 Evaluation Metrics

For each context, we assessed the interfaces based on Information Access Efficiency, Intrusiveness to the Primary Task, and Overall User Experience & Preference.

Information Access Efficiency: We measured *information access time* and *perceived ease of information access*. We used the video recording of the sessions to measure *information access time*, (T-Answer), based on how long it took participants to answer a question (from the time the question was asked verbally). *Perceived ease of information access* was evaluated through participant responses to the post-session survey question (seven-point rating scale): "When using the current interface (while reading an article / during the conversation), how easy it was to access information from the apps?"

Intrusiveness to the Primary Task: We measured *time away from the primary task* (T-Away) and *ease of resuming the primary task* after information access in both contexts. We used the video recording of the sessions to measure the time the user was away from the primary task (reading or conversation) and interacting with the interface (i.e., holding the mobile device, clicking on the interface, or looking at it). After finishing all sessions for each context, the participants were interviewed and asked to "rank the interfaces from the easiest to the most difficult for resuming reading/conversation) after information access."

For the solo scenario, we measured participants' *self-reported focus on reading*, by asking

them in the post-context interview if they felt that they read more of the article using a specific interface.

In the social context, we evaluated *social intrusiveness* by measuring participants' *awareness of others' facial expressions* and their *perception of the interface's social intrusiveness*. We used the video recording of the sessions to measure the number of times the participant missed the investigator's wink during the conversation. Users' *perception of interface's social intrusiveness* was evaluated through participant responses to a post-session survey question (seven-point rating scale): "when using the current interface during the conversation, how intrusive was the interface to your awareness of the social cues happening around you?"

User Experience & Preference: To assess the user experience (UX), we used the standard User Experience Questionnaire (UEQ) after each session [83]. After finishing all sessions for each context, the participants were interviewed and asked to (1) rank the interfaces based on their *preference* for that context, and (2) describe the advantages and disadvantages of each interface in that context.

Participants' *tendency to adopt AR* for their daily information access was evaluated through their responses to a five-point rating-scale question asked before and after the study: "If you had access to lightweight AR eyeglasses, how likely do you find it to use them instead of your mobile phone for your daily information access? Please explain why."

3.4.5 Hypotheses

We developed and tested the following hypotheses.

Information Access Efficiency:

$\boldsymbol{H1_{study2}}$ Since using the mobile phone involves physical interaction, more steps, and in-

creased workload, the mobile phone will be the least efficient interface for information access in both contexts.

$\boldsymbol{H2_{study2}}$ In the social context, the Context-Aware AR interface will take significantly less time to answer CTQs than the other two interfaces because it automatically makes the required content available to the user.

Intrusiveness to the Primary Task:

$\boldsymbol{H3_{study2}}$ In the solo context, Context-Aware AR will be less intrusive to the user's primary task of reading than the mobile phone, since information acquisition will be faster and less cognitively demanding.

$\boldsymbol{H4_{study2}}$ In the social context, Context-Aware AR will be the least intrusive interface to the conversation, since the virtual content automatically defaults to transparent when not needed, and the conversation partner's face is never occluded.

$\boldsymbol{H5_{study2}}$ In the social context, the non-adaptive Glanceable AR interface will be more *socially intrusive* than the mobile phone, since the virtual content in Glanceable AR can occlude the conversation partner's face.

User Experience & Preference:

$\boldsymbol{H6_{study2}}$ In both contexts, Context-Aware AR will be the best interface in terms of user experience and overall ranking.

$\boldsymbol{H7_{study2}}$ In the social context, the mobile phone will be the worst interface in terms of user experience and overall ranking.

$\boldsymbol{H8_{study2}}$ Users' reported tendency to adopt AR as their primary information access tool will increase from the beginning of the study to the end of the study, due to their exposure to Context-Aware Glanceable AR.

3.4.6 Participants

We gathered data from 36 participants (14 female), between 18 and 45 years of age (M = 26.53, SD = 6.06), recruited from our local university community. Nearly half of the participants (17) had little to no experience with AR/VR, and only two had used AR/VR more than 10 times.

3.4.7 Results

We conducted a series of analyses to test our hypotheses and explore the trade-offs between the interfaces for each context individually. For all the analyses, therefore, we separated the data based on the context. Shapiro-Wilk tests found that our data was not normally distributed. As such, we ran non-parametric Friedman tests to test the significant effect of independent variables. Wilcoxon signed-rank tests were conducted for pairwise comparisons with Bonferroni corrections applied. We also used Wilcoxon signed-rank tests for our qualitative data collected from the questionnaire and interview. We used an $\boldsymbol{\alpha}$ level of $\mathbf{0.05}$ in all significance tests.

Information Access Efficiency

We evaluated the effect of interface on *information access time* (T-Answer) in both contexts. In the solo context, a Wilcoxon signed-rank test ($\boldsymbol{Z = 17.62, p < 0.001}$) revealed that Context-Aware AR ($\boldsymbol{M = 2.83}$) was significantly faster than the mobile phone ($\boldsymbol{M = 7.31}$) (Figure 3.9, left).

In the social context, a Friedman test on T-Answer for CTQs ($\boldsymbol{\chi^2(2) = 162.66, p < 0.001}$) and STQs ($\boldsymbol{\chi^2(2) = 226.8, p < 0.001}$) revealed that the type of interface had a significant effect on *information access time* for both question types. Pairwise Wilcoxon signed-rank tests showed that answering both question types took significantly longer using

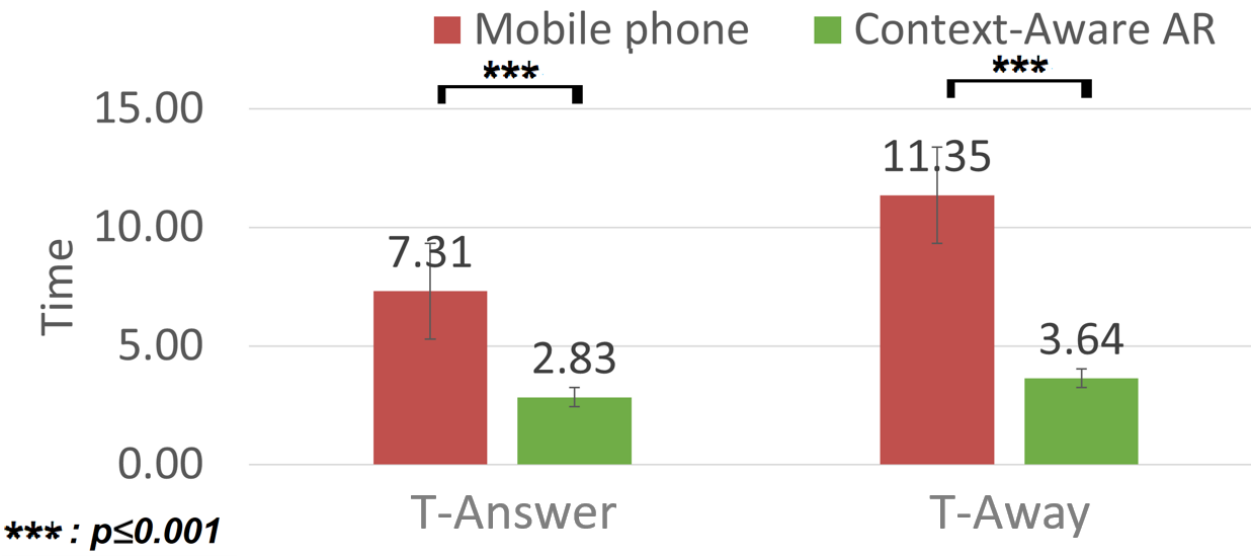


Figure 3.9: Time to answer questions (left) and time away from the reading task (right) in the solo context.

the mobile phone. For CTQs, mobile phone ($M = 7.44$) took significantly longer than both Glanceable AR ($M = 3.19$) ($Z = 12.16, p < 0.001$) and Context-Aware AR ($M = 2.78$) ($Z = 12.09, p < 0.001$). For STQs, mobile phone ($M = 8.46$) took significantly longer than both Glanceable AR ($M = 3.83$) ($Z = 15.34, p < 0.001$) and Context-Aware AR ($M = 4.78$) ($Z = 13.07, p < 0.001$). These pair-wise tests also revealed that compared to Glanceable AR, Context-Aware AR significantly reduced T-Answer for CTQs ($Z = 2.98, p = 0.009$), but it significantly increased T-Answer for STQs ($Z = -5.33, p < 0.001$) (Figure 3.10, left).

We also evaluated the effect of interface on *perceived ease of information access*. In the solo context, a Wilcoxon signed-rank test revealed that Context-Aware AR ($M = 6.11$) was perceived as significantly easier compared to mobile phones ($M = 4.69$) ($Z = -3.85, p < 0.001$).

In the social context, a Friedman test showed a significant effect of the interface ($\chi^2(2) = 27.469, p < 0.001$) on *perceived ease of information access*. Pairwise Wilcoxon signed-rank

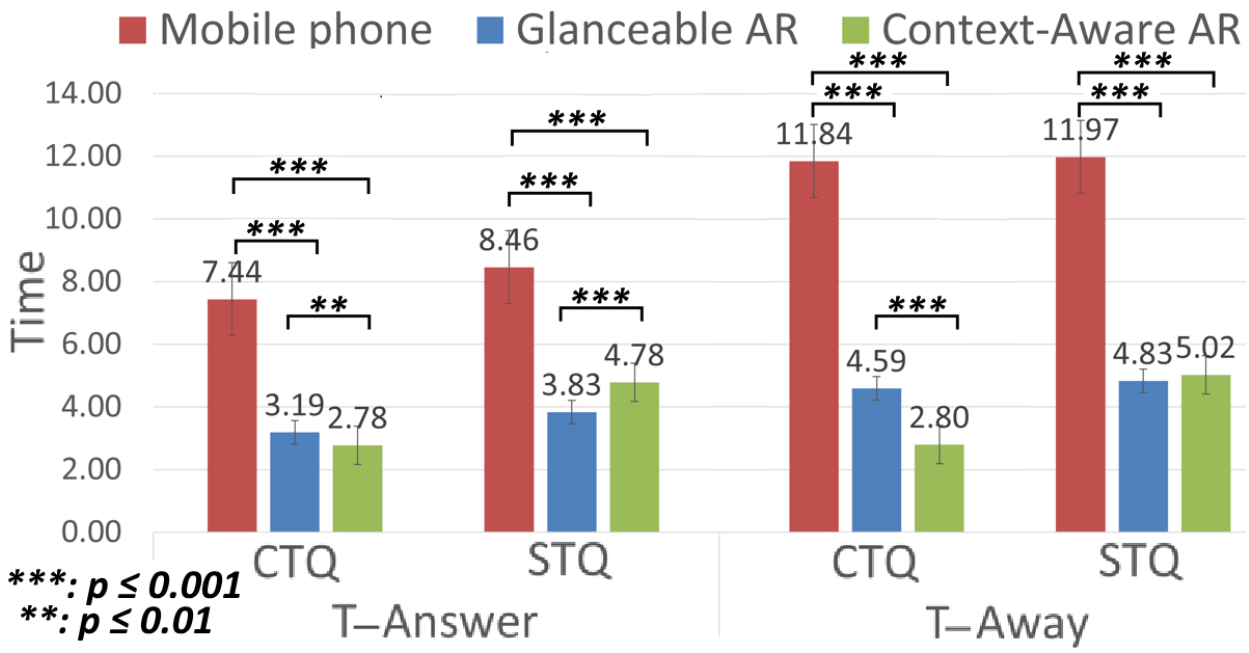


Figure 3.10: Time to answer questions (left) and time away from the conversation (right) for both question types in the social context.

tests revealed that mobile phone ($M = 3.83$) was perceived as significantly more difficult than both Glanceable AR ($M = 5.42$) ($Z = -4.14, p < 0.001$) and Context-Aware AR ($M = 5.64$) ($Z = -4.78, p < 0.001$).

Intrusiveness to the Primary Task

We evaluated the effect of interface on the user's *time away* from their primary task when accessing information. In the solo context, a Wilcoxon signed-rank test on T-Away revealed that Context-Aware AR ($M = 3.64$) significantly reduced *time away* from reading, compared to mobile phones ($M = 11.35$) ($Z = 17.98, p < 0.001$) (Figure 3.9, right).

In the social context, a Friedman test for CTQs ($\chi^2(2) = 181.45, p < .001$) and STQs ($\chi^2(2) = 268.6, p < .001$) revealed a significant effect of interface on the user's *time away* from the conversation. We ran Wilcoxon signed-rank tests for each question type on each interface pair. The test results showed that *time away* from the conversation was significantly

longer with the mobile phone. For CTQs, T-Away for the mobile phone ($\boldsymbol{M = 11.84}$) was significantly higher than for both Glanceable AR ($\boldsymbol{M = 4.59}$) ($\boldsymbol{Z = 11.92, p < 0.001}$) and Context-Aware AR ($\boldsymbol{M = 2.8}$) ($\boldsymbol{Z = 12.39, p < 0.001}$). For STQs, T-Away when using mobile phone ($\boldsymbol{M = 11.97}$) was significantly higher than both Glanceable AR ($\boldsymbol{M = 4.83}$) ($\boldsymbol{Z = 16.02, p < 0.001}$) and Context-Aware AR ($\boldsymbol{M = 5.02}$) ($\boldsymbol{Z = 16.01, p < 0.001}$). We also found that compared to Glanceable AR, Context-Aware AR significantly reduces T-Away for CTQs ($\boldsymbol{Z = 6.99, p < 0.001}$) (Figure 3.10, right).

We asked participants to rank the interfaces based on *ease of resuming the primary task* after information access. In both contexts, most participants (97% in the solo context and 67% in the social context) considered Context-Aware AR as the easiest to resume the primary task. In the social context, 72% ranked the mobile phone as the most difficult.

In the solo context, the participants were given the same amount of time to read the article in both sessions. However, their *self-reported focus on reading* indicates that 75% read more and understood the article better when using Context-Aware AR.

We also evaluated the *social intrusiveness* of the interfaces. A Friedman test on the total number of missed winks during the conversation ($\boldsymbol{\chi^2(2) = 36.24, p < .001}$) and the number of missed winks during information access ($\boldsymbol{\chi^2(2) = 25.35, p < .001}$) revealed significant effects of the interface on *awareness of others' facial expressions* (Figure 3.11). Wilcoxon signed-rank tests for each interface pair showed that users missed a significantly smaller number of winks overall, and even during information access when using Context-Aware AR. Using Context-Aware AR ($\boldsymbol{M = 1.42}$) significantly reduced the total number of missed winks compared to both Glanceable AR ($\boldsymbol{M = 2.89}$) ($\boldsymbol{Z = 4.36, p < 0.001}$) and mobile phone ($\boldsymbol{M = 2.97}$) ($\boldsymbol{Z = 4.66, p < 0.001}$). During information access, Context-Aware AR ($\boldsymbol{M = 1.28}$) significantly reduced the number of missed winks compared to both Glanceable AR ($\boldsymbol{M = 2.28}$) ($\boldsymbol{Z = 4.4, p < 0.001}$) and mobile phone ($\boldsymbol{M = 2.28}$)

$(\boldsymbol{Z = 4.3, p < 0.001})$.

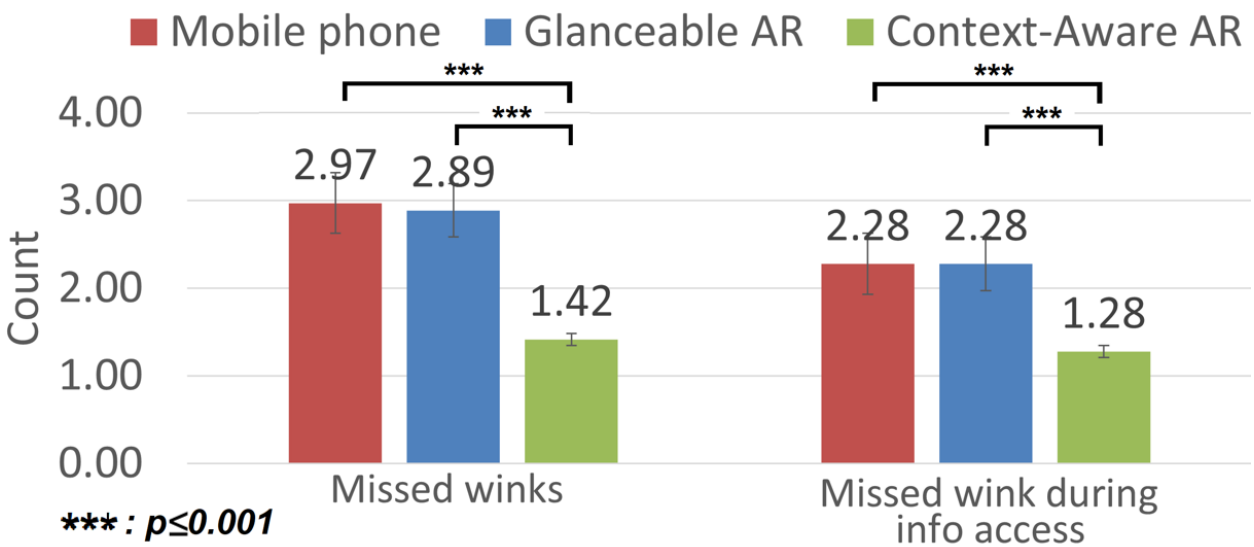


Figure 3.11: Missed social cues overall (left) and during information access (right) in the social context.

A Friedman test showed a significant effect of interface on participants' self-reported *perception of interface's social intrusiveness* $(\boldsymbol{\chi^2(2) = 19.451, p < 0.001})$. Wilcoxon signed-rank tests for each pair showed that mobile phone $(\boldsymbol{M = 5.44})$ was perceived as significantly more intrusive than both Glanceable AR $(\boldsymbol{M = 3.64})$ $(\boldsymbol{Z = 3.64, p < 0.001})$ and Context-Aware AR $(\boldsymbol{M = 4.14})$ $(\boldsymbol{Z = 3.88, p < 0.001})$.

User Experience & Preference

Context-Aware AR was ranked first for *preference* by the majority of participants (89% in the solo context and 64% in the social context). In the social context, 64% of our participants ranked mobile phones as their least preferred.

We performed Wilcoxon signed-rank tests for each interface pair on our UEQ results from both contexts to compare *UX aspects*. In the solo context, Context-Aware AR was significantly higher in attractiveness, efficiency, stimulation, and novelty (Figure 3.12). For the social context, compared to the mobile phone, both Context-Aware AR and Glanceable AR

had significantly higher attractiveness, stimulation, and novelty and lower dependability and perspicuity (Figure 3.13).

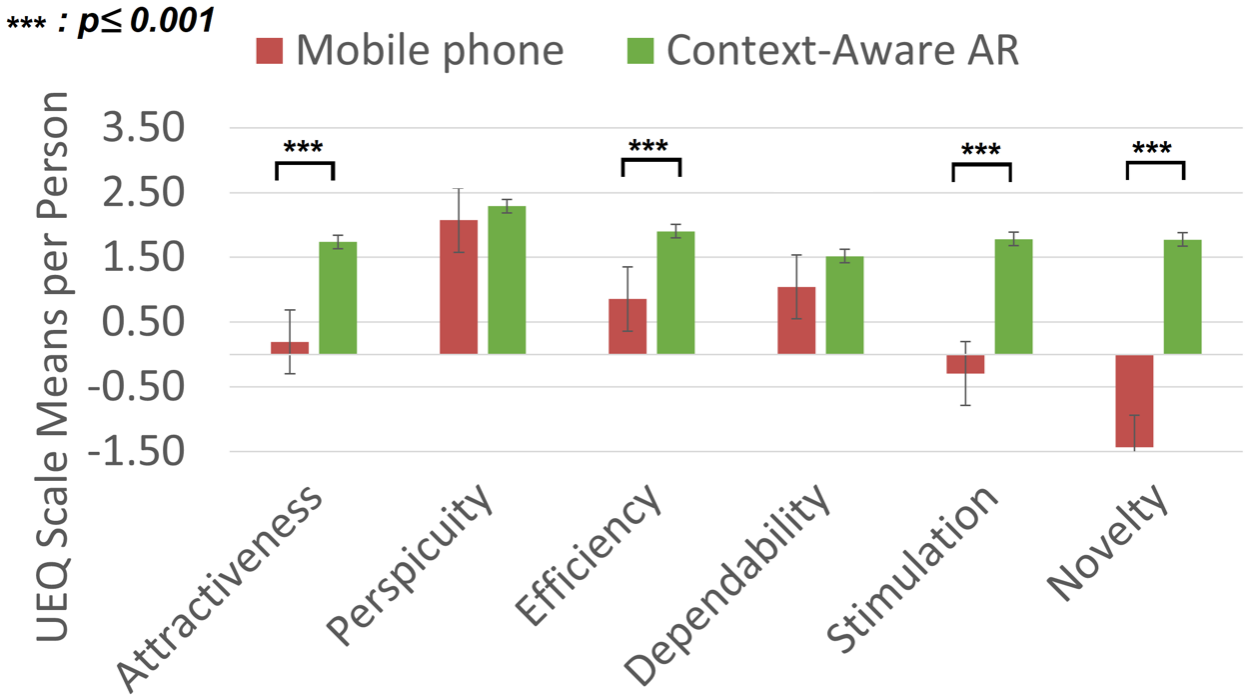


Figure 3.12: UEQ results in the solo context.

Finally, a Wilcoxon signed-rank test ($\boldsymbol{Z = 2.44, p < 0.01}$) on the participants' *tendency to adopt AR* comparing their responses before the study with those afterward showed that participating in the study significantly increased their tendency to adopt AR for their daily information access ($\boldsymbol{M_{pre-Study} = 3.28}$, $\boldsymbol{M_{post-Study} = 3.89}$).

3.4.8 Discussion and Future Work

Information Access Efficiency

In $\boldsymbol{H1_{study2}}$, we hypothesized that the mobile phone would be the least efficient interface for information access in both contexts. Our results support this hypothesis by showing that, compared to the other two interfaces, the mobile phone increased *information access time*, regardless of the context or question type, and was *perceived as more difficult* for information

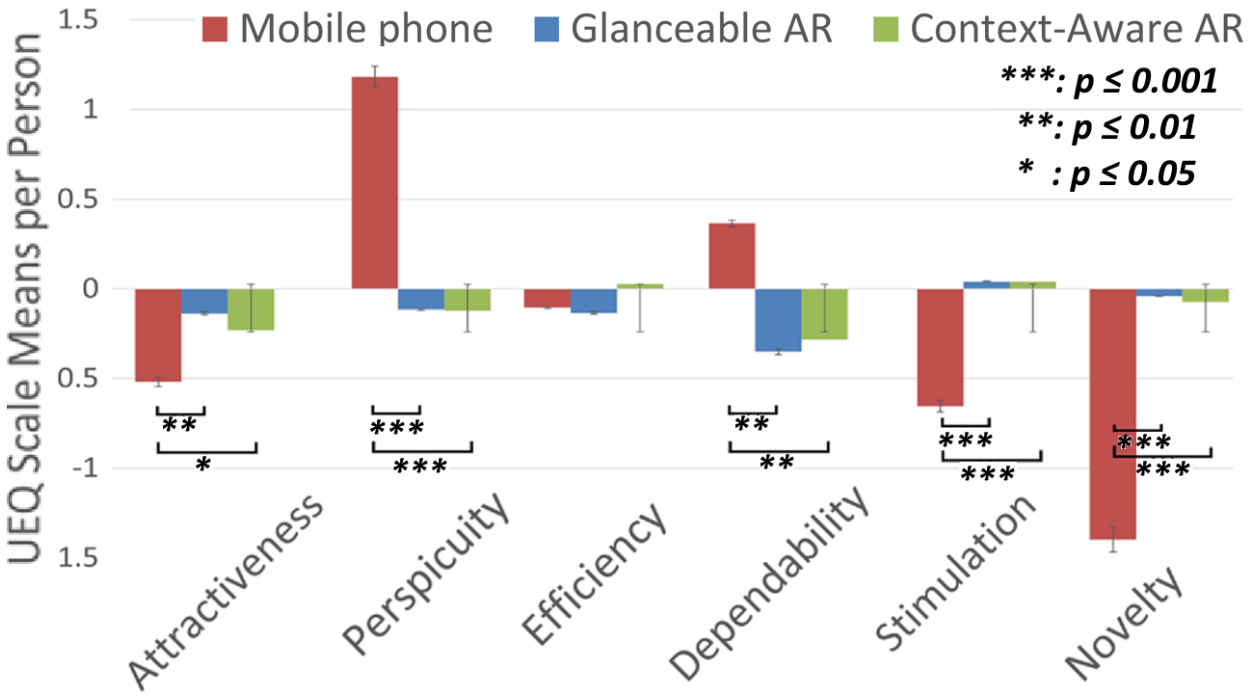


Figure 3.13: UEQ results in the social context.

access in both contexts. The study also showed that Context-Aware AR took less time to answer CTQs than the other two interfaces, supporting $\boldsymbol{H2_{study2}}$.

Together, this constitutes strong evidence that Glanceable AR can provide rapid information access with a quick glance and that the expansive and persistent display provided by Glanceable AR has important benefits over mobile phones. Moreover, in the social context, we showed that the automatic activation of apps based on conversational content can be a powerful way to support information access in a social situation, without requiring any manual interaction.

Intrusiveness to the Primary Task

When using the Context-Aware AR interface in the solo context, our participants spent less *time away* from reading, felt it was *easier to resume reading* after information access and reported *higher focus on reading* compared to the mobile phone. This supports $\boldsymbol{H3_{study2}}$, that Context-Aware AR is less intrusive to the user's primary task.

Meanwhile, $\boldsymbol{H4_{study2}}$ was partially supported by our data. When using Context-Aware AR, participants took the least *time away* for information access, but only when answering CTQs. They also perceived the interface as the *easiest for resuming the conversation* after information access and less *socially intrusive* compared to mobile phones. Finally, they had the most *awareness of the interlocutor's facial expressions* in the Context-Aware condition.

We also noticed that participants occasionally missed the STQ, either completely ignoring the question or forgetting it. Only 20% of these missed questions occurred with Context-Aware AR, compared to 54% with the mobile phone, which indicates a greater ability to effectively multi-task with Context-Aware AR.

Together, the results from $\boldsymbol{H1_{study2}} - \boldsymbol{H4_{study2}}$ demonstrate that our Context-Aware Glanceable AR prototype achieved its goals of balancing rapid information access with minimal intrusion on the primary task. In the solo context and in the case of CTQs in the social context, the Context-Aware condition automatically both ensures the visibility of needed content at a glance and avoids occlusion of the primary task. In the worst case (STQs during a conversation), the user still has to interact to activate an app manually, but this is still significantly faster than the mobile phone and results in less time away from the conversation. We speculate that the non-adaptive Glanceable AR had less time away than the Context-Aware condition for STQs because participants waited for the app to automatically become opaque before deciding to click on it themselves.

$\boldsymbol{H5_{study2}}$ hypothesized higher *social intrusiveness* of the non-adaptive Glanceable AR compared to the mobile phone, but our results do not support this hypothesis. We found no significant difference in participants' *awareness of other person's social expressions*, or in their self-reported *perception of the interface's social intrusiveness*. We can explain this finding by observing that all participants manually made the apps transparent during the conversation, only activating them during information access. When using the mobile phone,

time away from the conversation was longer due to the multiple steps of interacting with the device, and participants found it difficult to focus on the conversation or detect the social cues during this time of interaction with the device (T-Away).

User Experience & Preference

Our results partially support $\boldsymbol{H6_{study2}}$ and $\boldsymbol{H7_{study2}}$. In both contexts, Context-Aware AR was ranked as the *most preferred*, and mobile phone as the *least preferred* interface. Context-Aware was also rated significantly higher than the mobile phone on the UEQ factors of attractiveness, stimulation, and novelty in both contexts. On the other hand, the mobile phone had significantly higher ratings for perspicuity (i.e., ease of understanding) and dependability in the social context, which is reasonable since users have high familiarity and experience with mobile phone interfaces. In addition, we did not find any significant advantages of Context-Aware over non-adaptive Glanceable AR on any of the UEQ factors. We suggest that these results may be explained by general uneasiness about automatic adaptations in the Context-Aware condition. Many users expressed concerns of this sort, such as: "What if I don't need something and it keeps opening the content?" and "What if I need other applications?"

Our results on users' reported tendency to adopt AR as their primary information access tool indicated an increase from the beginning of the study to the end of the study, supporting $\boldsymbol{H8_{study2}}$. This suggests that experience with our Context-Aware Glanceable AR prototype helped participants envision the potential benefits of someday replacing their smartphones with iAR glasses.

Limitations and Future work

This study aimed to compare the most commonly used interface for daily information access against a glanceable AR interface accessible through all-day wearable eyeglasses. We assume that future lightweight AR glasses will stay unlocked as long as they are worn after

being "logged in" once the user puts them on. We also assume such devices will become commonplace and socially accepted. However, these assumptions might not prove to be fully accurate, potentially limiting the validity of our results. Our prototype only supported two contexts. However, a Context-Aware Glanceable AR interface must differentiate and adapt to various daily contexts. This raises research questions regarding the definition of AR user contexts and proper adaptations to them.

Design Limitations: We limited our prototype to three Glanceable apps and used air-tap interactions for both contexts. Our interface focused on occlusion and intrusiveness challenges in AR and improving information access (**DC1, DC2, DC3**), adapting the content prioritization, virtual content's transparency, and placement. Future studies should identify and address other challenges of AR and investigate Context-Aware interfaces that intelligently adapt the interaction technique and the group of available apps.

User Study Limitations: Our participants' familiarity and experience with the three interfaces were not equal, which could have biased against the AR conditions and the Context-Aware AR condition. This, in fact, makes our positive results about AR even stronger, but it also helps to explain why Context-Aware AR did not completely live up to our expectations. Furthermore, the Context-Aware AR condition had a significant computational load, even though the computer vision and speech recognition were done on a remote server. This made the performance of the Context-Aware condition less than ideal, especially when users needed to air tap on apps manually. This also helps explain some of the missed expectations. Future studies should also examine the effects of Context-Aware AR use on both sides of a social situation and on larger social situations with more than two participants since we did not evaluate how Context-Aware AR or the other conditions affect the perceptions or social acceptance of the interlocutor.

3.5 Conclusions

To address *RQ1.*, *RQ2.*, and *RQ4.*, this chapter explored the design of various AR interfaces for three different contexts and evaluated their usability and efficiency in these contexts. In Section 3.2, we introduce criteria for classifying AR interfaces and define Glanceable AR interfaces based on these criteria. These Glanceable AR interfaces allow users to efficiently access relevant information with a quick glance while performing other tasks in the real world. However, their constant presence and potential intrusiveness necessitate the use of iAR systems, making them an ideal subject for exploring context awareness and iAR functionalities.

Section 3.3 presented the design and evaluation of multiple context-aware adaptations in a simulated AR environment, imitating occlusion in a RW monitoring task (*RQ1.*). The results indicate the advantages of context awareness on the usability and efficiency of the AR adaptations and the user's awareness of the RW in an occlusion context.

This work was published in the IEEE VR 2020 Workshop on Everyday Virtual Reality:

- [28] Shakiba Davari, Feiyu Lu, and Doug A. Bowman. "Occlusion management techniques for everyday glanceable AR interfaces." 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pages 324–330. IEEE, 2020.
  DOI:10.1109/VRW50115.2020.00072

Section 3.4 presents our design and prototype for a context-aware AR interface that can detect and adapt to the social context (*RQ2.*). This interface was evaluated across two distinct contexts: a reading task in a single-user context and a social conversation context (*RQ4.*). The results of evaluating these interfaces in their corresponding contexts are presented. This

prototype demonstrates the potential of context-aware AR interfaces to improve users' focus and social awareness and the usability and efficiency of AR interfaces in different contexts.

This work was published in IEEE VR 2022:

- [29] Shakiba Davari, Feiyu Lu, and Doug A. Bowman. "Validating the benefits of glanceable and context-aware augmented reality for everyday information access tasks" In 2022 IEEE Virtual Reality and 3D User Interfaces (VR), pages 336–444. IEEE, 2022 DOI:10.1109/VR51125.2022.00063

The findings from these studies offer valuable insights for identifying and categorizing AR design elements and contextual components, as well as for developing effective context-aware AR interfaces within iXR systems (*RQ5.*, *RQ6.*, and *RQ7.*).

Chapter 4

Design Space & A Context-aware AR Placement Strategy

4.1 Overview

By converting the entire 3D space around the user into a screen, AR can ameliorate traditional displays' space limitations and facilitate the consumption of multiple pieces of information at a time. However, the extensive possibilities for providing information inherent in AR pose a unique challenge for designing interfaces that seamlessly integrate into users' environments while maintaining usability and clarity. For example, an AR navigation app overlaying directional cues onto a busy street must ensure these cues remain visible yet unobtrusive amid obstacles and pedestrians.

An AR interface is composed of a set of interactable XR (Extended Reality) objects that present information in various modalities. We regard an XR object as the fundamental component of any AR interface, irrespective of its modality for presenting digital information or its potential for interactivity. Adapting the design elements of these XR objects enables adaptations within the iAR. To answer *RQ5.*, in this chapter, we explore XR objects' design

dimensions that enable suitable presentation and interaction in AR.

In Section 4.2, we present the design dimensions of AR adaptations and outline a comprehensive visual design space for XR. Utilizing these design elements towards further addressing *RQ3.*, we investigate a specific use case of context-aware adaptations concentrated specifically on the spatial layout of the AR content. The inappropriate spatial layout of XR objects can lead to distractions and occlusions, overwhelm the user, and hinder their interactions with the environment. Section 4.3 explores utilizing context to adapt the **spatial placement** of XR objects and proposes a context-aware *Environment-referenced* placement strategy. This strategy utilizes a Hybrid Frame of Reference (FoR), introduced in our design space, to place each XR object using an intermediary object from the Real-World (RW) environment. This placement strategy leverages contextual information to maintain efficiency and unobtrusiveness within context-switching scenarios.

In Section 4.4, we present an experiment designed to explore *RQ4.* and the nuanced interplay between contextual components and the effectiveness of placement strategy. This experimental design evaluates the accuracy, efficiency, and usability of our proposed context-aware placement strategy and a traditional body-fixed placement across four distinct contexts. The performance of these placement strategies from our within-subjects user study emphasized the importance of intermediaries' relevance to the user's focus. These findings underscore the importance of iAR interfaces, indicating that the appropriate use of an adaptive content placement strategy in a context can significantly improve task efficiency, accuracy, and usability.

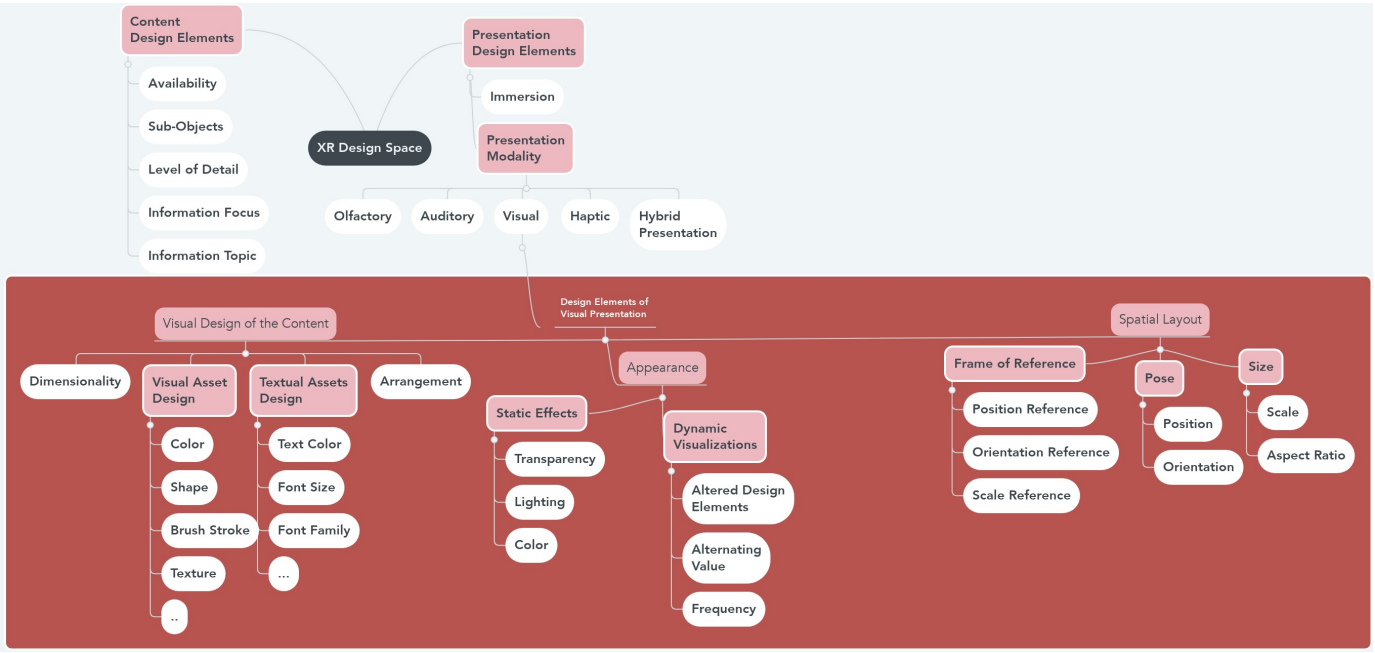


Figure 4.1: Design Dimensions of XR interfaces

4.2 XR Design Dimensions

Prior research extensively investigates design elements and adaptation techniques for Extended Reality (XR) interfaces, providing a foundation for this study [5, 16, 54, 122]. To identify adaptable design elements of XR objects we adopted a systematic methodology. We began by compiling a preliminary list of XR design elements from the existing literature.

We then explored the challenges and capabilities of XR through brainstorming sessions and reviews of existing XR interfaces, focusing on features and design elements associated with these challenges and capabilities. Throughout this process, we iteratively referred to our preliminary list, continuously adding, grouping, and removing design elements. For example, the issue of occlusion of real-world objects highlights the importance of position, size, and opacity of XR objects, while the link between eye fatigue, insomnia, and blue light usage emphasizes color theme as an adaptable design elements (see Figure 4.1).

Each XR object may comprise multiple other sub-objects and can provide applications, auditory notifications, 3D models, 2D images, system recommendations, menus, and the like. We divide the design elements of XR objects into two high-level categories: content design and presentation design (Figure 4.1).

4.2.1 Content Design

Previous research explored adapting the **availability** of XR content to provide the right information at the right time, enhancing performance and preventing potential obstructions and distractions in various contexts [9, 29, 89]. Effectively adapting in-use XR objects' **availability** allows the creation of interfaces that "close", "open", or "minimize" apps tailored to the user. Within the design of an XR interface, the availability of each XR object can be modified either manually by the user or through context-aware adaptations, or be immutable (e.g., a constantly "opened" system-level XR object). For instance, in educational XR aimed at facilitating student learning experiences, the availability of specific XR objects may be contingent upon the student's location, fulfillment of predetermined criteria, or their class and exam schedule.

Each XR object is dedicated to a specific subject inherent to its essence and remains immutable. For instance, one XR object functions as a weather app while another provides recommendations or notifications to the user. However, within an interface, adaptations can also be made to the content of an XR object, varying its **sub-objects**, **Level of Detail (LoD)**, **information focus**, and **topic** [33, 34, 89]. For instance, the content within the weather app could display data focused on either the hourly forecast or the 10-day forecast, exclusively feature precipitation forecasts, or offer minimal LoD, showcasing solely visualizations of the current temperature and conditions. Adapting the content within an XR object

allows for the creation of a personalized interface, offering only the intended content. Imagine an XR object that provides recommendations to the user. Context-aware adaptations to this XR object's content facilitate the recommendation of various apps and information tailored to the user's needs, thereby enhancing their performance. The objective of an effective XR interface is to present pertinent information in a timely and seamlessly integrated manner, ensuring it remains intuitive, easily accessible, understandable, and unobtrusive to the user. This underscores the pivotal role of content presentation in the design process of XR objects.

4.2.2 Presentation Design

The presentation design of an XR object involves its immersion and the modality in which it is presented (refer to Figure 4.1). The **immersion** design element of an XR object determines whether it is immersive or non-immersive. A non-immersive XR object allows simultaneous interaction with other XR objects. However, an immersive XR object restricts user engagement solely to itself and its sub-objects, limiting interaction with external XR objects.

Presentation Modality

Skarbez et al. propose that all technology-mediated experiences fall under the term mixed reality, suggesting that an XR object can present information through various modalities [125]. The efficacy of an XR interface and the information it provides also relies on its presentation modality. In modern XR interfaces, XR objects are predominantly showcased as 2D and 3D overlays within the XR environment, utilizing the *visual modality*. However, XR objects can possess the capacity to convey content through *audio, haptic, and olfactory modalities*, as well as through *hybrid presentations*.

Visual Presentation Design Space

Extensive research in the literature has investigated the visual design of XR interfaces and their design elements. We propose a taxonomy for the visual design elements of XR objects, categorizing them into the visual design of the content of an XR object, and the XR object's appearance, and spatial layout within the environment (Figure 4.1).

Visual Design of the Content

An XR object's visual content involves a structured arrangement of various 2D or 3D visual and textual assets, impacting the clarity, readability, and comprehensibility of the XR object. The **dimensionality** design element indicates whether the XR object employs 2D or 3D visualizations to present its content, while the **arrangement** design element determines how these assets are organized and integrated to create a coherent visualization.

Visual Asset Design: Design elements such as **texture**, **brushstroke**, **size**, and **color** are instrumental in designing 2D or 3D images, models, lines, and shapes [84]. These visual assets represent either the entirety or specific aspects of an XR object's content.

Textual Asset Design: Design elements pertinent to the typography of the text influence the legibility, readability, and aesthetic appeal of textual assets, such as text boxes [6]. These typography features encompass various design elements such as **typeface selection**, **text color**, **font and point size**, **line length**, **line and letter spacing**. Typography design elements can serve readability purposes and creative endeavors, such as conveying specific messages and establishing hierarchy, tone, and visual rhythm. For instance, context-aware adaptations to typography enable XR to accommodate low-vision or dyslexic users.

Appearance

A multitude of static and dynamic design elements collectively influence the realism, perceptibility, and noticeability of an XR object's appearance. Previous work proposed guidelines for limiting unnecessary or non-conforming dynamic visuals and effectively using visual cues, color, and animation to enhance performance and prevent potential obstructions and distractions [9, 122]. These elements contribute to the natural integration and alignment of visual XR objects within the XR environment.

Static Effects: Various visual effects can be applied to manipulate the **transparency**, **lighting**, and **colors** of the 2D or 3D visualization of an XR object. Some common static visual effects include night/day mode color themes and adaptations to color saturation, warmth, tint, brightness, and contrast. These adaptations significantly impact the opacity, legibility, and intuitiveness of the XR interface within its context. For example, adjusting an XR object's transparency and colors enables its seamless integration into the XR environment [8]. Similarly, aligning the shadow and brightness of an XR object with the environmental lighting conditions ensures its noticeability without overwhelming brightness, enhancing the natural appearance of the scene.

Dynamic Visualizations: Adapting at least one design element of an XR object at a specified frequency results in dynamic 2D or 3D animated visualizations within the XR object. For example, an animated XR object that dynamically moves at speed $\boldsymbol{v}$ visualizes pose adaptations to its position design element at the frequency of $\boldsymbol{v}$. While dynamic visualizations do not necessarily follow a periodic pattern, their design elements can consist of the **altered design elements** and their **alternating values** and **alteration frequencies**. Such dynamic visualizations can enhance the overall realism, noticability, and interactivity of the XR interface, and the user's awareness by highlighting relevant information or events within specific XR objects. Consider a dynamic visualization that alters the content compositing of an XR object by altering its line colors between green and transparent at

frequency $\boldsymbol{f}$. Such blinking outlines can provide notifications about new mission assignments in an immersive VR game or indicate the occlusion of an important object in a context-aware AR [28].

Spatial Layout

The spatial integration of a visual XR object plays a crucial role in the interface's efficiency and the user's engagement and safety [40, 79]. The proper placement and alignment of XR objects can facilitate and speed up access to XR content, prevent occlusions, and influence the user's perception and awareness of the virtual and real-world environments. To ensure that an XR object is displayed in a non-intrusive, realistic, and easily accessible manner within the user's field of view, it is essential to give careful consideration to the design of the frame of reference, the object's pose, and size.

Frame of Reference (FoR): A coordinate system is used as the FoR to determine the XR object's position, orientation, and scale in relation to the environment [12, 13, 16, 40, 42, 74]. The literature has identified various FoRs for registering XR objects within the environment as world-fixed, object-fixed, head-fixed, body-fixed, or device-fixed [16]. *World-fixed FoR* registers the XR object in a fixed position relative to stationary real-world locations and objects. *Object-fixed FoR* is defined relative to another XR object or real-world object (including input devices), so that the content moves along with that object. *Head-fixed FoR* registers XR objects relative to the virtual camera. This FoR is also referred to as display-fixed, Heads-Up-Display (HUD), or fixed display. *Body-fixed FoR* registers the XR objects relative to the user's body. World-fixed, display-fixed, and body-fixed settings have been commonly used as the FoR for XR objects in AR interfaces [42]. However, in all these scenarios, the coordinate system of the world or another object is uniformly employed as the FoR for position, orientation, and scale. For example, by utilizing $\boldsymbol{ref}$ as the reference

object for $\boldsymbol{O1}$, any alterations to its position, orientation, and scale would apply to $\boldsymbol{O1}$.

While previous research studies commonly apply a **Unified FoR**, we propose that an XR object's FoR can be **Hybrid** and adhere to a distinct **position reference**, **orientation reference**, and **scale reference**. Recent research by Manakhov et al. touches on the concept of hybrid FoRs, specifically discussing the *HeadDelay* and *Path* FoRs in the context of walking scenarios [98]. For instance, consider $\boldsymbol{N}$ as an XR object positioned on the north side of the user and displaying a pointer towards the north direction in an AR navigation interface. As its FoR, this XR object can utilize a custom coordinate system that employs the user as its $\boldsymbol{reference_{position}}$ and scale reference and the world as its orientation reference.

$$\boldsymbol{N_{FoR} = (ref_{position} : user,\ \ ref_{scale} : user,\ \ ref_{orientation} : world)}$$

This FoR enables the XR object to be consistently positioned on the north side of the user and point towards the north regardless of the user's orientation while simultaneously following the user's position and maintaining a fixed distance from the user in mobile contexts.

Pose: The pose of an XR object in 3D space is specified through its **position** and its **orientation** relative to its FoR. Cartesian $\boldsymbol{(x,\ y,\ z)}$ or spherical $\boldsymbol{(r,\ \theta,\ \phi)}$ coordinates can denote the **position**, while the facing or pointing **orientation** can be indicated through Euler angles, quaternions, or rotation matrices.

Size: The **scale** of an XR object's size relative to its FoR, along with its **aspect ratio**, represents its size within the environment. Scale can be depicted by a singular value, such as a uniform scale factor, or by distinct scale factors for each dimension. The size of the XR object, including its length, width, and height within the environment, indicates the space it occupies, contributes to a sense of depth, and affects the realism and occlusion of the XR environment. Maintaining the correct aspect ratio can be essential in specific XR objects,

as deviations may lead to distortion or misalignment, detrimentally affecting its integration within the environment and the user experience.

Non-Visual Presentation

XR objects can be presented through *audio, haptic, and olfactory modalities*, as well as through *hybrid presentations*. As XR technology advances, utilizing the full potential of non-visual outputs is crucial for enhancing realism and interaction in virtual and augmented environments. For instance, delivering content through olfactory sensations enhances the presence and plausibility of immersive XR experiences more effectively than instructing users to "imagine smelling grass" through visual text. However, employing many of these modalities necessitates specialized hardware and software, and their implementation relies on the specifications of the XR system.

This research primarily concentrates on visual presentation in AR. However, comprehending the design elements of non-visual modalities allows AR designers to develop more immersive, intuitive, interactive, and engaging experiences for users across various applications and contexts. In this section, we briefly examine the common design elements for delivering information through non-visual modalities and encourage researchers to further investigate and expand upon this line of inquiry (See Figure 4.2).

Audio Presentation

Delivering XR object content through audio outputs, such as speech synthesis or sound effects, via the *audio modality* is relatively commonplace. Utilizing speech specifically enables intuitive communication of complex information, reduces ambiguity, and enhances accessibility for visually impaired users. We consider the design elements of audio output in XR, focusing on its duration, volume, spatialization, and reverberation. The **duration** of audio

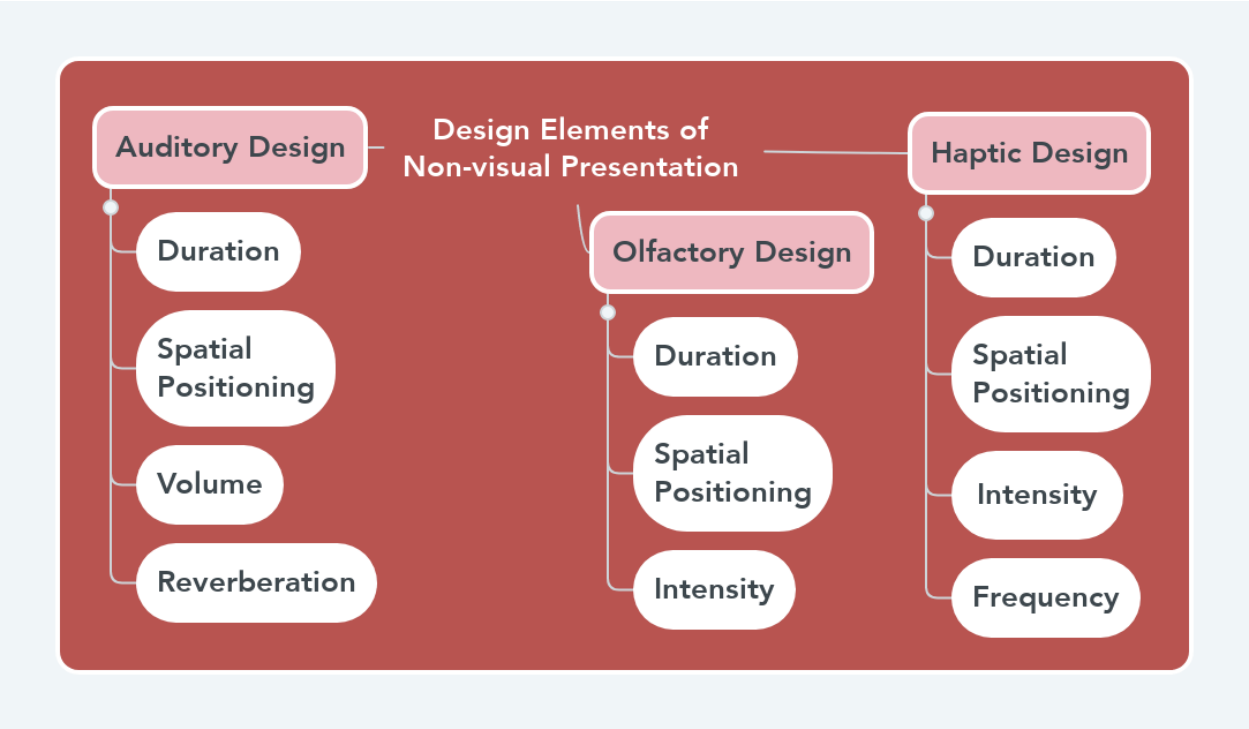


Figure 4.2: XR design elements for non-visual modalities

output in an XR interface can significantly influence distractions, disturbances, and conflicts. Prolonged audio presentations may divert the user's attention away from their task and influence their awareness of the XR environment. Additionally, repetitive or non-verbal audio played for extended periods can become annoying and disruptive. Furthermore, aligning the **spatial positioning** of the audio output in 3D space with its associated XR object and surroundings enhances realism.

Matching the audio output's **volume** and intensity to the XR environment's noise level ensures audibility without disruption. Moreover, the audio output can employ reverb effects to authentically mimic the acoustic properties of the surrounding XR environment. **Reverberation** artificially creates a spatial perception by imitating how sound bounces off surfaces in different environments, thereby shaping the acoustic ambiance.

Haptic Presentation

The *haptic modality* employs tactile sensations such as vibration, force, and thermal feedback to convey diverse information. For instance, force feedback can signify pressure and weight, while subtle variations can convey texture, shape, and deformation. We explore the design elements of haptic output in XR, focusing on its duration, spatial positioning, intensity, and frequency. The **duration** of haptic sensations during XR interaction varies based on context and purpose. Brief vibrations may signal events or interactions, while prolonged feedback can simulate ongoing physical interactions or environmental cues. Moreover, the **spatial positioning** of haptic sensations in the XR environment, relative to the user's body or XR objects, affects the realism and efficacy of haptic output. Aligning haptic output with visual or auditory cues enhances immersion and intuitiveness, enabling users to perceive virtual interactions more accurately.

User engagement and immersion in XR can be influenced by the **intensity** or amplitude of haptic sensations experienced by the user. These sensations simulate a broad spectrum of tactile experiences, ranging from intense vibrations, temperatures, or weights conveying a heightened sense of impact or force to gentle taps. Similarly, the **frequency** of haptic vibrations and output can shape user perception of the content and evoke various sensory experiences. High frequencies may signal urgency, while low frequencies can mimic gentle movements or subtle textures.

Olfactory Presentation

The *olfactory modality* utilizes scent to convey information or evoke a sense of smell. It reproduces the natural scents of the virtual environment through diffusers, fans, scented materials, electronic devices, and computational methods. Due to technical complexities, such olfactory outputs remain uncommon in XR and are primarily reserved for short, stationary immersive experiences. We discuss the main design elements of olfactory output in XR, focusing on duration, spatial positioning, and intensity. Considering the **duration** of a scent

within the design of an XR environment is crucial to ensuring its alignment with user interactions and the experience's narrative. On the other hand, strategic **spatial positioning** of scent sources relative to the user's position within the XR environment can heighten the sense of presence and realism. For instance, positioning floral scent emitters near XR objects depicting flowers enriches the immersive experience by reinforcing visual and olfactory cues. Additionally, a scent's **intensity** / potency, from subtle hint to powerful odor, can evoke various emotional responses or amplify specific narrative elements.

Hybrid Presentation

The hybrid presentation involves multimodal XR objects that integrate sub-objects of various modalities to present the content. Visual and auditory modalities are frequently integrated to enhance the efficacy of an XR object. For example, combining sound effects with visual notifications or background music with visual animations enhances their capacity to convey intended information to the user. Similarly, haptic and olfactory modalities are not commonly utilized individually, as the information they convey tends to be ambiguous and limited in range. However, they can be effectively employed in conjunction with visual or auditory modalities to enhance presence and plausibility in tailored immersive XR experiences. For example, an XR object in an immersive experience can integrate haptic and visual modalities to simulate various temperatures that match the XR object's 3D visual presentation. In more general-purpose scenarios, these modalities can also provide cues or signals to the user within a multimodal XR object and increase the efficacy of information presentation.

4.2.3 Input Design

The efficiency and functionality of an XR interface depend on the input design and user interactions. Key input design elements encompass the interactivity of each XR object and the selected input modality and interaction design for individual XR objects (Figure 4.1).

The **interactivity** of an XR object represents users' active engagement to provide input and adapt its design elements, including actions like clicking buttons, entering text, repositioning, or adjusting audio levels. An interactive XR interface enhances efficiency by enabling users to customize it, control information flow, and access relevant data according to their needs. An XR object can offer an adaptive level of interactivity for specific design elements. For instance, typically, the availability of XR objects is interactive, allowing the user to open and close them. However, some XR objects, for instance, static visual assets like walls in a VR game, may be designed with no interactivity even in terms of their availability, limiting the user to "close" and remove them from the environment. Additionally, intelligent XR interfaces can adapt the level of interactivity for individual XR objects and particular design elements depending on the context. For instance, they can limit the availability of specific XR objects to restrict students' content access during an exam.

Input modality and interaction design encompass the methods used to capture and interpret user input. These are crucial elements in XR interface design influencing ease of use, intuitiveness, accessibility, and efficiency. Inappropriate selection of input modality and poorly crafted interactions can result in user frustration and challenges.

Input modalities range from natural inputs like eye-based, hand-based, and voice-based methods to specialized equipment like controllers and touchpads. The interplay between input modality and user context shapes the XR experience, affecting users' perception of their virtual and physical surroundings and their ability to interact with the interface. Even

within a chosen input modality, selecting the right **interaction technique** significantly impacts users' ability to navigate and utilize the XR interface, influencing its learnability, usability, and overall effectiveness.

4.3 Context-aware Environment-referenced Placement

A placement strategy determines the $\boldsymbol{reference_{position}}$ and $\boldsymbol{reference_{orientation}}$ of an XR object's FoR, as well as the pose of the object within its FoR. The literature has investigated different world-fixed or user-fixed placement strategies for XR objects, including surround-fixed, object-fixed, body-fixed, head-fixed, and device-fixed [16, 28, 42]. These placement strategies utilize empirical data from previous use cases to either statically pose the XR content within a uniform FoR or propose a context-specific adaptive placement. In this work, we propose an adaptive strategy for utilizing each XR object's association with the environment to dynamically place it without prior knowledge of the specific context or environment.

We propose *Environment-referenced* placement, an adaptive placement strategy that employs a hybrid FoR tailored to each XR object utilizing an $\boldsymbol{reference_{orientation}}$ distinct from the XR object's $\boldsymbol{reference_{position}}$. Bell et al. investigated placing relevant objects in close proximity as a strategy for view management [8]. This approach intelligently identifies a relevant object or area within the environment as an intermediary for each specific XR object by leveraging machine learning (ML) and large language models (LLMs). Using each intermediary's relative position to the XR object's $\boldsymbol{reference_{position}}$, our placement strategy determines a distinct $\boldsymbol{reference_{orientation}}$ for each XR object. In essence, while the XR content moves along with and remains anchored at a consistent distance from its $\boldsymbol{reference_{position}}$, it continuously rotates to maintain a persistent pose between its

$\boldsymbol{reference_{position}}$ and intermediary (See Figure 4.3: **1-b and 2-b**).

Implementing this adaptive placement strategy within a context-aware interface that can identify relevant intermediaries for the user's task and specific context allows for placing the XR objects precisely where needed, thereby enhancing the interface's efficiency. By leveraging the user's spatial awareness and focusing their attention on nearby RW and virtual objects within their visual range, this placement strategy seamlessly integrates XR objects into the environment, simplifying the interface complexity [77]. Furthermore, the intelligent detection of intermediaries within new environments and tracking their relative position to the XR object's $\boldsymbol{reference_{position}}$ offers an adaptive placement strategy for portable and dynamic use cases. This is particularly useful when users move to new or unfamiliar environments, or when the position of the user, $\boldsymbol{reference_{position}}$, or intermediaries change within the same environment. However, in such contexts, especially when the selected intermediaries are irrelevant to the setting and the user's task, this placement strategy can be unintuitive and introduce additional mental and physical load to locate the intermediaries. Envision utilizing various placement strategies for name tags at a conference. A non-adaptive object-fixed placement strategy might use each attendee as the uniform FoR for their name tag. In this interface, the pose of each name tag depends on the specific attendee's position and orientation. This can result in the inconsistent appearance of name tags, creating a confusing interface. For instance, depending on people's relative orientation, one person's name tag might appear to their right while another's name tag appears to their left in the user's view. Additionally, an attendee's name tag might appear too small or at an angle, depending on their distance from the user or whether they are facing the user, affecting its visibility and readability. On the other hand, using the user as the uniform FoR ensures consistency in the distance and angle at which name tags appear in the user's view. However, without additional contextual information and adaptations to clearly associate each

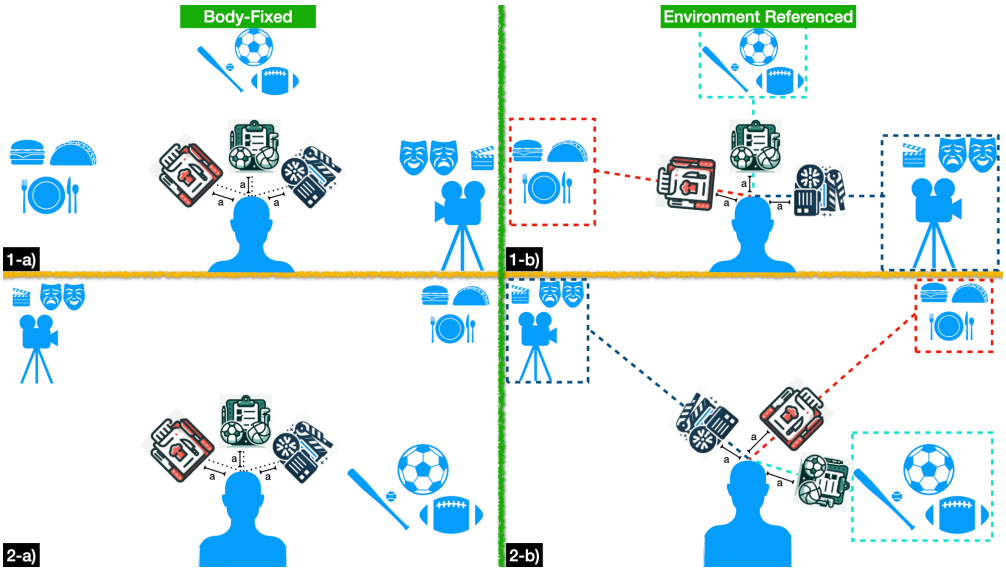


Figure 4.3: Illustration of our two placement strategies. **(1-a, 2-a)**: Body-fixed placement: each XR object's pose remains consistent based on the orientation of the user, i.e., frame of reference. **(1-b, 2-b)**: Environment-referenced placement: each XR object's pose remains consistent based on the orientation of its intermediary relative to the frame of reference, maintaining a consistent distance from the user. **(1-a, 1-b)**: The user is at the session start point, positioned at an equal distance from the intermediaries. **(2-a, 2-b)**: The user has moved near the Sports category's intermediary.

name tag with the corresponding attendee, a non-adaptive body-fixed placement strategy can result in visual clutter and overwhelm the user with excessive information.

Our placement strategy enables a solution to consider the user as the $\boldsymbol{reference_{position}}$ and each attendee as their intermediary to their name tag. This enables each person's name tag to appear in a uniform pose around them in the user's view. Using the user as the $\boldsymbol{reference_{position}}$ ensures consistency in the distance and angle at which the name tags appear in the user's view. While these XR name tags consistently follow the user, their distinct $\boldsymbol{reference_{orientation}}$ ensures they are visible solely when the specific attendee, i.e., intermediary, is within the user's field of view, minimizing visual clutter and information overload.

4.4 Experiment 3: AR Content Placement Strategy in Four Contexts

While an adaptive placement strategy adjusts to specific contextual information, relying solely on one or a few context components may be insufficient for optimal performance. The people, objects, events, and elements within the environment can affect the user's behavior, information access needs, and interests. Similarly, one's physical, cognitive, and emotional states influence perception and interactions within the environment, changing perspective, spatial awareness, directional perception, and distance from different objects. Varying the setting and the user state can influence the effectiveness and performance of an adaptive placement strategy within a scenario. This experiment explores one research question:

$\boldsymbol{RQ_{Experiment3}}$ Does adaptive XR placement *always* enhance performance?

To address this $\boldsymbol{RQ_{Experiment3}}$, we designed an experiment to assess information acquisition using our proposed adaptive placement strategy compared to a non-adaptive placement strategy across four distinct contexts: *Dynamic Mobile*, *Dynamic Stationary*, *Static Stationary*, and *Static Mobile*.

This experimental design aims to investigate the nuanced interplay between placement strategy and contextual components.Section 4.4.7 presents the results from our user study within *Dynamic Mobile*, *Static Mobile*, *Dynamic Stationary*, and *Static Stationary* contexts. It is important to note that, while our proposed adaptive strategy intelligently identifies intermediaries, this work does not aim to propose, implement, or evaluate the performance of ML models in identifying optimal intermediaries. Consequently, to ensure that the reliability and performance of the intermediary detection do not influence our results, within our experimental design we manually identify relevant intermediaries for the XR content within each context.

4.4.1 Experimental Design

Our experiment incorporated three independent variables, each with two levels: *Placement Strategy* (adaptive *Environment-referenced* or non-adaptive *Body-fixed*), *Real-World Setting* (either *Dynamic* or *Static*), and *User State* (*Mobile* or *Stationary*). We conducted a within-subjects user-study to evaluate the performance of each *Placement Strategy* within each combination of *RW Setting* and *User State*. Thus, each participant's performance was evaluated across eight sessions.

Study Task

The challenges associated with information acquisition and improper placement of XR content are exacerbated when dealing with a large amount of content. We devised a trivia game about 28 different countries to assess information acquisition within a large collection of AR documents. The trivia questions covered three categories: international food, movies, and sports. Our AR interface displayed a total of 84 documents, one for each category and country. The documents were displayed across three panels in the interface, one for each category. The user's objective was to identify the category and country mentioned in the question and open the correct document. All questions explicitly stated both the category name and the country name. The question was repeated every 30 seconds until the correct document was opened.

RW Setting

The user was positioned in an empty room featuring three designated intermediary objects, each representing one of our categories.

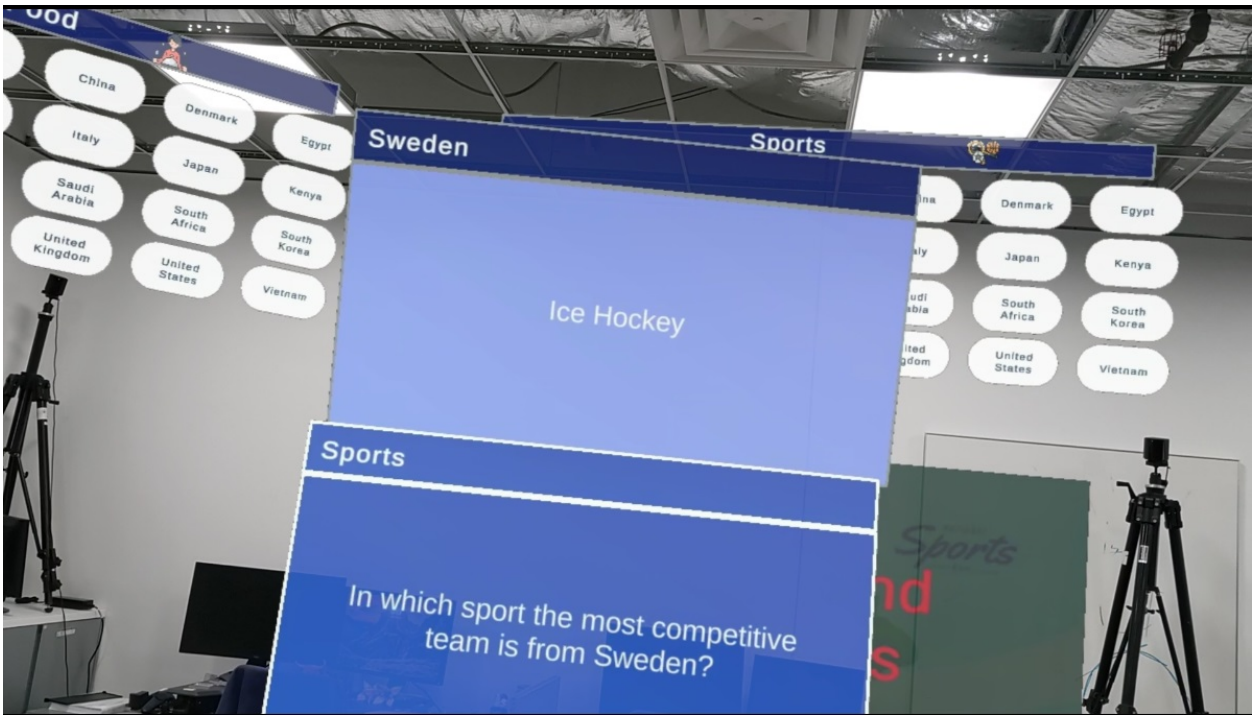


Figure 4.4: Each document displays the answer to a trivia question

Static Individual Setting: In the *Static* contexts, the user participated in a trivia game alone in the room, where questions were displayed on an AR screen in front of them. In these *Static* contexts, the RW environment involved three virtual posters, each displaying images relevant to a specific category, and remained static.

Dynamic Social Setting: In the *Dynamic* contexts, three virtual humans acted as trivia hosts. These trivia hosts engaged in dialogue with the user and each other, moved around the room, and switched positions throughout the session. The virtual avatars were each attired in clothing pertinent to one of our categories: a chef, a soccer player, and a director. Each host presented questions aligned with their respective category.

User State

At the beginning of each session, the user was standing at the *Start Point*, facing the sports intermediary.

Stationary User: In the *Stationary* contexts, the participant remained consistently positioned at the *Start Point*.

Mobile User: In contrast, in the *Dynamic* contexts, participants were prompted to move near a different intermediary at specific times during each session.

Placement Strategy

We designed a non-adaptive and adaptive placement strategy to display these panels (Refer to Figure 4.3). Variations in the spatial layout design elements of the panels influence not only their placement but also their appearance, occlusion, and ease of interaction. This emphasizes the importance of consistently presenting the panels at a uniform size, height, and distance within the user's view to ensure that our results are solely due to placement. Additionally, to access the panels in *Mobile* contexts, they must follow the user. To address these concerns, in both placement strategies, we used the user as the $\boldsymbol{reference_{position}}$ and designed all category panels in a uniform $\boldsymbol{pose}$ and $\boldsymbol{size}$.

Environment-referenced Placement: As described in Section 4.3, we implement an adaptive placement strategy to determine an individual $\boldsymbol{reference_{orientation}}$ for each category panel utilizing its intermediary's relative position to the user. In our design, we placed each panel directly between the user and its intermediary (See Figure 4.5). To ensure that intermediary detection's performance does not influence our results, within each setting, we manually assigned each panel's relevant intermediary. In *Static* contexts, the virtual poster, displaying

images relevant to each category, served as that category's intermediary.

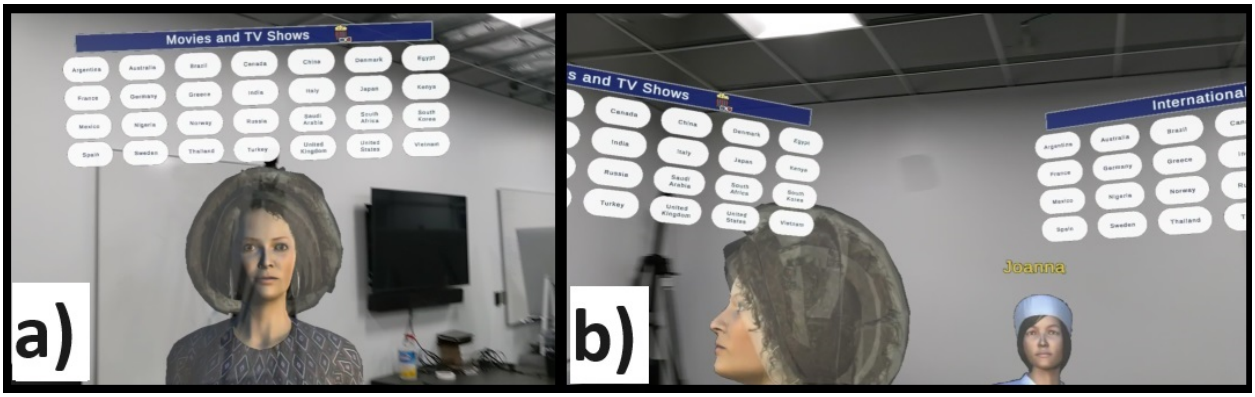


Figure 4.5: Environment-referenced placement in dynamic contexts: a) The category panel relevant to each trivia host is displayed above their head. b) the panels move along with their associated trivia host.

In social interactions, visual engagement with conversation partners is crucial. Research by Hood et al. demonstrates that individuals tend to gaze at their conversation partners for a significant portion of the interaction [68]. Miller et al. revealed that the mere presence of virtual agents influences the user's non-verbal and social behavior, with participants avoiding avertion of their gaze from avatars and treating them as real individuals [103]. Consequently, in *Dynamic* contexts, we considered each trivia host as the intermediary for their category panel.

Body-fixed Placement: As a baseline non-adaptive placement strategy, we implemented a commonly-used *Body-fixed Placement*, in which the user is considered as the uniform FoR. By using the orientation of the user's body as the $\boldsymbol{reference_{orientation}}$ for all panels, within this non-adaptive placement, a particular panel is consistently placed on the same side of the user's body. In our implementation, we situated the sports panel directly in front of the user, the food panel to their right side, and the Movies panel to their left (See Figure 4.6).

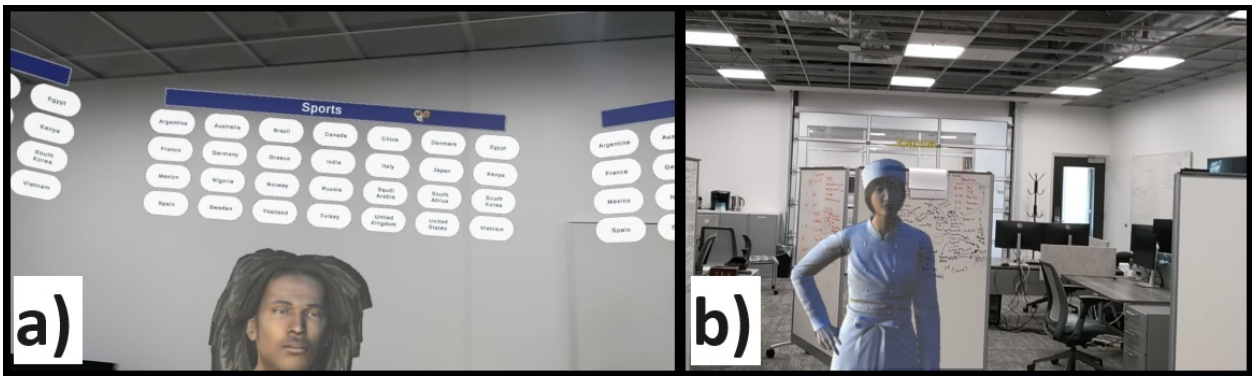


Figure 4.6: Body-fixed placement in dynamic contexts: a) All three category panels in front of the user's body. b) No AR content is placed near the trivia hosts, not directly in front of the user.

4.4.2 Hypotheses

Both adaptive and non-adaptive placement strategies have unique advantages that influence their effectiveness for the same task. Regarding $\boldsymbol{RQ_{Experiment3}}$, we hypothesize that neither of the two placement strategies would be singularly superior for information acquisition in all contexts, as their advantages and disadvantages can vary depending on the specific context. To explore the impact of different contextual components on the effectiveness of these placement strategies for information acquisition, we formulated and tested three hypotheses denoted as **H1, H2, and H3**.

H1 *Placement strategy effectiveness varies across different contexts.*

When utilizing the appropriate contextual information for adaptation, the *Environment-referenced* placement facilitates rapid access to information, by placing each XR object where it is needed. Within this experiment, this adaptive strategy can lead to intuitive placement and optimal access to information with no added cognitive load for navigation between category panels. On the other hand, *Body-fixed* placement facilitates constant and rapid access to all category panels at all times, regardless of the *RW Setting* or *User State*, or the

suitability of the intermediaries. However, this strategy's non-adaptive placement increases the decision-making and mental load associated with navigating the correct category panel.

H2 *The effectiveness of Environment-referenced placement strategy is influenced by RW Setting.*

The people, objects, events, and elements within the RW setting can affect user behavior, information needs, and interests. We evaluated the performance of adaptive and non-adaptive placement strategies within various *RW Settings*, hypothesizing that *Environment-referenced*:

H2-a *outperform Body-fixed in Dynamic contexts.*

In social contexts, information acquisitions are usually associated with their conversation and interlocutor with whom the users typically maintain eye contact [68]. Similarly, in both *Dynamic* contexts within this experiment, the information acquisition inquiries are associated with the same trivia host with whom the user is conversing and visually engaged. Additionally, the user's engagement in social and conversational interactions, along with the dynamic interactions between the trivia hosts, can increase the user's mental load in our *Dynamic* contexts. By considering the conversation partner as the intermediary in these contexts, *Environment-referenced* intuitively displays solely the required category panel where needed. We hypothesize that in these *Dynamic* contexts, the *Environment-referenced* strategy facilitates the cognitive and physical load of finding and accessing the relevant panels and enhances information access.

H2-b *underperforms Body-fixed in Static contexts.*

The advantages of *Environment-referenced* rely on its potential to place the AR content precisely where needed. On the other hand, the suboptimal selection of intermediaries can make this placement unintuitive and lead to increased cognitive load and physical effort for information access. Unlike the *Dynamic* contexts, in our *Static* ones, the user's visual

attention within the environment varies, posing challenges in anticipating the optimal intermediaries. Additionally, due to the unchanging environment and the *solo* social setting, our *Static* contexts impose a lower mental load on the user. Considering the limited number of category panels, this higher cognitive capacity can facilitate decision-making associated with the *Body-fixed* placement. We speculate that the *Body-fixed* placement's consistency and effortless access to all documents through a glance, outweigh its cognitive decision-making in *Static* contexts. Consequently, we hypothesize that in *Static* context, the *Body-fixed* placement consistent glanceable access enhances information acquisition, compared *Environment-referenced.*

H3 *Environment-referenced underperforms Body-fixed in Mobile contexts.*

The user's physical, cognitive, and emotional states influence their capability to interact with and perceive the RW environment, including their perspective, spatial awareness, directional perception, and distance from different objects. While within our *Stationary* contexts, the user's perception and interaction remain unchanged, in our *Mobile* contexts, users move across different locations within the environment.

When utilizing *Environment-referenced* placement, depending on the user's position, the relative position of each category panel changes, potentially appearing on their left side at one time and on their right side at another. The user's varying perspective and directional perception of the environment can lead to inconsistency and confuse the user. In our *Mobile* contexts, the *Environment-referenced* placement requires spatial recall for the users to determine their current perspective and the relative position of each intermediary, increasing their cognitive load, especially within unfamiliar environments. Conversely, regardless of the user's mobility and location, *Body-fixed* placement ensures a persistent approach to position each category panel relative to the user. We evaluate the effectiveness of each placement strategy and hypothesize that the consistency of the non-adaptive *Body-fixed* placement

enhances information acquisition in *Mobile* contexts.

4.4.3 Apparatus

Throughout the experiment, participants utilized a Microsoft HoloLens (2nd gen) Augmented Reality Head-Worn Display (AR HWD). This wireless apparatus has a display density of 47 pixels per degree and a diagonal field of view spanning 52 degrees. Our virtual environment and task-specific panels and documents were developed using the Unity 2022.3.4f1 game engine in conjunction with Microsoft's MRTK toolkit.

All sessions started with the user positioned at the *Start Point*, situated at an equal distance from the intermediaries (Figure 4.3: **1-a and 1-b**). The participants directly faced the sports intermediary, with the food intermediary to their left and the movies intermediary to their right. We designed eight unique trial sets, each serving as the script for one session. Each *Stationary* session comprised three trials, with one trial for each category. In all *Stationary* trials, the user was far (farther than $\mathbf{1.5}\ \boldsymbol{m}$) from all intermediaries. In specific *Mobile* trials, the user was positioned near the intermediary relevant to the trial's question. To ensure the independence of our trial results from this distance, the *Mobile* sessions consisted of six trials, with two trials for each category, one when the user was near its intermediary, and another one when far.

The interface provided three panels, each containing 28 documents corresponding to one of our three trivia categories. These documents were organized alphabetically in a grid format, with four rows and seven columns within each panel. All three panels shared a uniform appearance, except for the panel header, which showcased the category title along with an icon representing that category. Participants utilized hand interactions facilitated by the HoloLens device to select and interact with documents, enabling them to open and

close them as needed. Each document displayed a single word providing the answer to the question, closing once the user clicked anywhere on it.

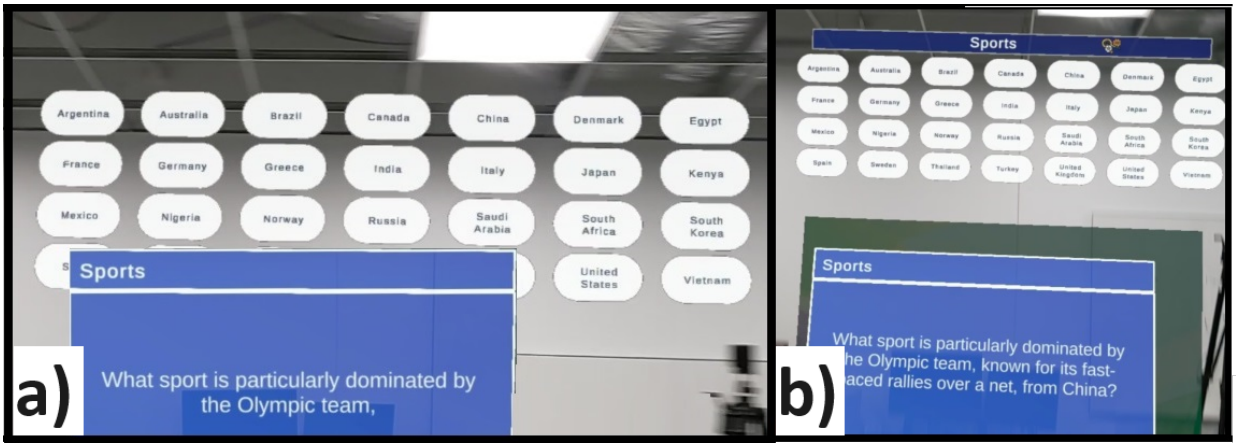


Figure 4.7: a) The questions are presented word-by-word. b) The header for each category panel reappears after the question is asked.

In the *Static* contexts, a screen was displayed in front of the user at eye level to present the trivia question (Figure 4.7). In the *Dynamic* contexts, a trivia host engaged in a conversation with the user and verbally asked the question. Consequently, the questions were presented differently across contexts, providing text in the *Static* contexts while audio in the *Dynamic* ones. We employed multiple methods to prevent participants from initiating document searches during the question in specific contexts and ensure consistency across sessions. First, all questions were formulated to reveal the name of the country as the concluding word of the question. Second, in *Dynamic* contexts, participants were instructed to maintain eye contact during questions and to respond only after the entire question was asked. Third, we matched the rate at which the question was presented across various *RW Settings*. Using the duration of each question's audio within the *Dynamic* contexts, its equivalent question was revealed one word at a time in the *Static* contexts. Fourth, while presenting the question, the panel headers were rendered transparent.

Additionally, within *Dynamic* contexts, to enhance the ecological validity of social interac-

tions and alleviate any biases arising from individual differences and social behavior among participants, we instructed all participants to engage with the trivia hosts as if they were real people and to maintain eye contact during conversations. These trivia hosts were designed using avatars and animations sourced from the Microsoft Rocketbox library and Maximo [49, 51].

4.4.4 Procedure

Following approval from our university's Institutional Review Board, we enlisted participants from our local university. The experiment lasted 120 minutes and was divided into four parts, each session corresponding to a distinct context. Within each context, each participant underwent two sessions, each employing one of our placement strategies. Overall, each participant underwent eight sessions.

The presentation order of the sessions was counterbalanced to mitigate any potential bias arising from a specific trial set or the order in which participants experienced the session. To counterbalance the ordering of the sessions, participants were presented with the four contexts in a Latin Square ordering. Within each context ordering, half of the participants began with *Body-fixed* placement, while the other half started with *Environment-referenced*. We recruited 24 subjects, ensuring three participants per session-ordering (3 x 4 x 2).

Upon their arrival, participants were asked to carefully review and sign the consent form. Subsequently, we gathered demographic details and information regarding their prior experience with XR. Once this process was completed, we provided a concise overview of the study, acquainted participants with the device, and instructed them on its operation and interaction methods. They then proceeded to engage in the four parts of the experiment.

We started each part by introducing the particular task associated with that context. Within

each part, participants experienced each *Placement Strategy* in a separate session. Prior to each session, they underwent a training session, completing a minimum of six trials to acquaint themselves with the environment, tasks, and specific content arrangement offered by the *Placement Strategy* within that context. Following each session, participants were directed to complete the post-session survey on a computer. During the post-context interview conducted after trying both placement strategies within a context, participants were prompted to compare and rank the two *Placement Strategy* options based on speed, ease of use, intrusiveness, and preference.

At the conclusion of the study, all participants underwent interviews regarding their overall experiences across all contexts, as well as the advantages and drawbacks of each *Placement Strategy* within each combination of *RW Setting* and *User State*. Participants had the freedom to take breaks between sessions as needed.

4.4.5 Evaluation Metrics

The impact of XR on mental workload and task performance is frequently assessed through objective metrics like task completion time and error rate, as well as subjective evaluations like NASA TLX [72]. To explore information acquisition using our proposed placement strategies within these contexts, we evaluated the Accuracy, Efficiency, and Usability of each *Placement Strategy* for information acquisition. To assess these metrics, we collected objective task-specific data from each trial and subjective data from survey and interview responses.

Accuracy

A critical aspect in determining the effectiveness of an AR placement strategy for information acquisition is its potential to maintain task performance accuracy. In this experiment, a trial continued until the participant opened the correct document. We regard every instance of opening the wrong document as an error, adversely affecting accuracy. To assess accuracy, we evaluated ***Error Rate*** based on the total number of trials in which any errors occurred.

Efficiency

An effective placement strategy facilitates efficient and swift information acquisition through a large quantity of AR documents and minimizes the associated physical and mental load. We assessed information acquisition efficiency through various metrics evaluating the time and difficulty of finding the correct document within each session.

Navigation Time: utilized the logged time-stamped gaze data to measure the duration of time for answering a question. This time was measured from the moment a question was completely presented to the user for the first time until the user's gaze fixated on the correct document. Unfortunately, we observed that the accuracy of the device in recognizing hand gestures was significantly influenced by the participant's race and skin color, resulting in longer times to open a document for certain participants. To remove this bias from our data, we measured *Navigation Time* using the gaze data rather than the time at which a document was opened.

Gaze Switch: measured the frequency of gaze switches among the category panels within each trial. An increase in the frequency of gaze switches suggests a higher head-turn rate and physical load for accessing the document. Additionally, this measurement can be indicative of the unintuitiveness of the placement and the user's difficulty in navigating the correct

category panel.

Perceived Speed: quantified the participants' response to the post-context interview question about ranking the fastest placement strategy for information acquisition within that context.

Perceived Ease: quantified the participants' post-context ranking of the easiest placement strategy for information acquisition.

NASA TLX: collected the post-session response to the short version of the Standard NASA Task Load Index (TLX) questionnaire [62].

Usability

To assess the usability of our placement strategies within these contexts:

Perceived Intrusiveness: quantified the post-context ranking of the most intrusive placement strategy within each context.

Preferred Placement: quantified the ranking of the most-preferred strategy within each context.

SUS: collected participants' responses to the short version of the Standard System Usability Survey (SUS) [18] in our post-session survey.

4.4.6 Participants

We enrolled **24** participants. This population aged between **18** and **45** years ($\boldsymbol{M = 25.62, SD = 5.24}$), and included eight female and thirteen graduate students. **13** participants had little to no experience with XR, and only **5** had used XR more than ten times.

4.4.7 Results

To test our hypotheses, we conducted a series of analyses on our quantitative data to assess the effects of our three independent variables: *RW Setting*, *User State*, and *Placement Strategy*. Our data, collected from 24 participants in a within-subjects study comprising eight sessions, represents a sizeable sample size. Considering our large sample size (n>50), we conducted Kolmogorov-Smirnov (KS) goodness-of-fit tests on our measurements, revealing the non-normal distribution of our data.

We opted to conduct nonparametric statistical analysis using Aligned Rank Transform (ART) [143]. ART enables parametric statistical analyses, such as Repeated Measures Analysis of Variances (RM-ANOVA), for both continuous and ordinal responses, without the assumption of normality or limiting the number of involved factors. For all metrics, we applied ART to the data and conducted a factorial non-parametric ART RM-ANOVA to test the significant effect of our independent variables. For all pairwise comparisons on the interactions that showed a significant effect, we applied the ART-C algorithm for multifactor post-hoc contrast tests with Bonferroni adjustments. We used an $\boldsymbol{\alpha}$ level of $\mathbf{0.05}$ in all significance tests.

Accuracy

Our analyses indicate that *Error Rate* is significantly influenced by *RW Setting* X *Placement Strategy* ($\boldsymbol{F_{(1,161)} = 5.25, p \sim 0.02}$). Pairwise comparisons on the effect of *RW Setting* on the *Error Rate* of each *Placement Strategy* are illustrated in Figure 4.8. These results indicate significantly lower *Error Rate* when using *Environment-referenced*, compared to *Body-fixed* in both **a)** in *Dynamic* contexts ($\boldsymbol{p \sim 0.01}$, supporting **H2-a**), and **b)** in *Static* contexts ($\boldsymbol{p < 0.0001}$, rejecting **H2-b**).

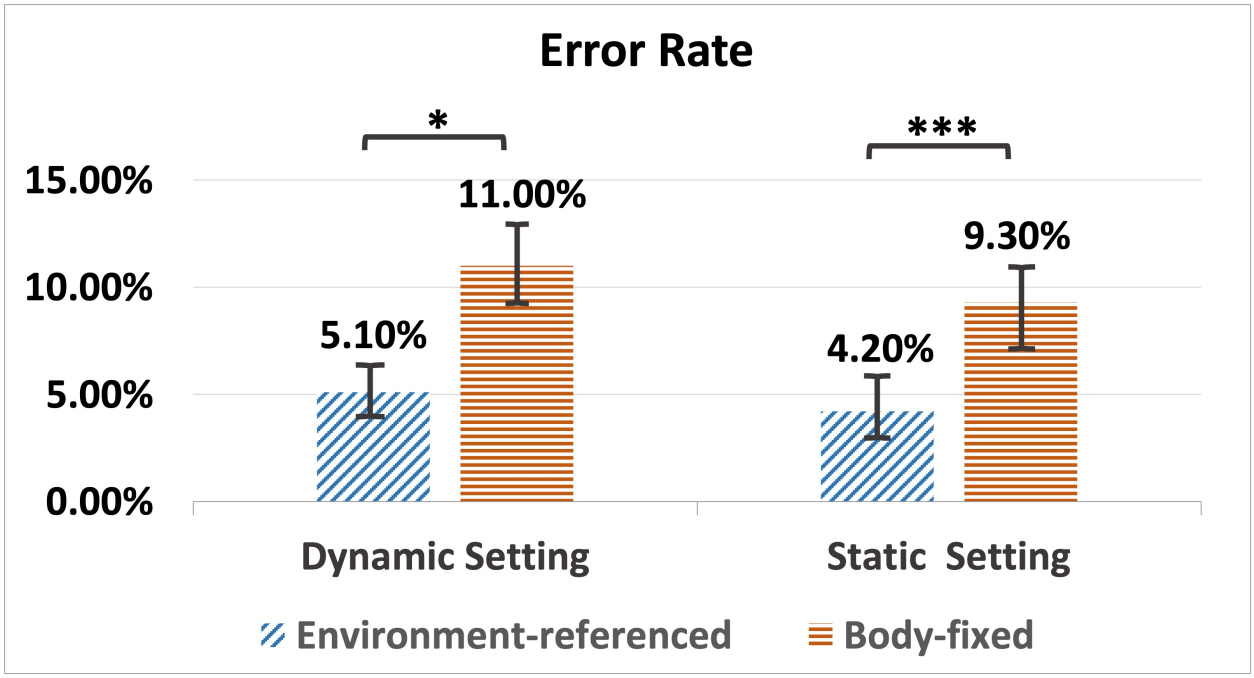


Figure 4.8: Accuracy within various RW settings

Efficiency

Overall analysis of our quantitative and qualitative data on the *Placement Strategy* efficiency within various contexts, points toward a significant influence of the interaction between *RW Setting* and *Placement Strategy*, supporting **H2**.

Navigation Time:

Participants' information acquisition within correct trials, where no errors occurred, showed a significant effect of *RW Setting* X *Placement Strategy* on *Navigation Time* ($\boldsymbol{F_{(1,769.53)} = 22.32,\ p < 0.0001}$). Our pairwise comparison on this interaction indicated that *Environment-referenced* performed **a)** statistically faster than *Body-fixed* in *Dynamic* contexts ($\boldsymbol{p < 0.0001}$, supporting **H2-a**). **b)** significantly faster in *Dynamic* contexts compared to *Static* ones ($\boldsymbol{p < 0.0001}$ supporting **H1**) (See Figure 4.9).

Gaze Switch:

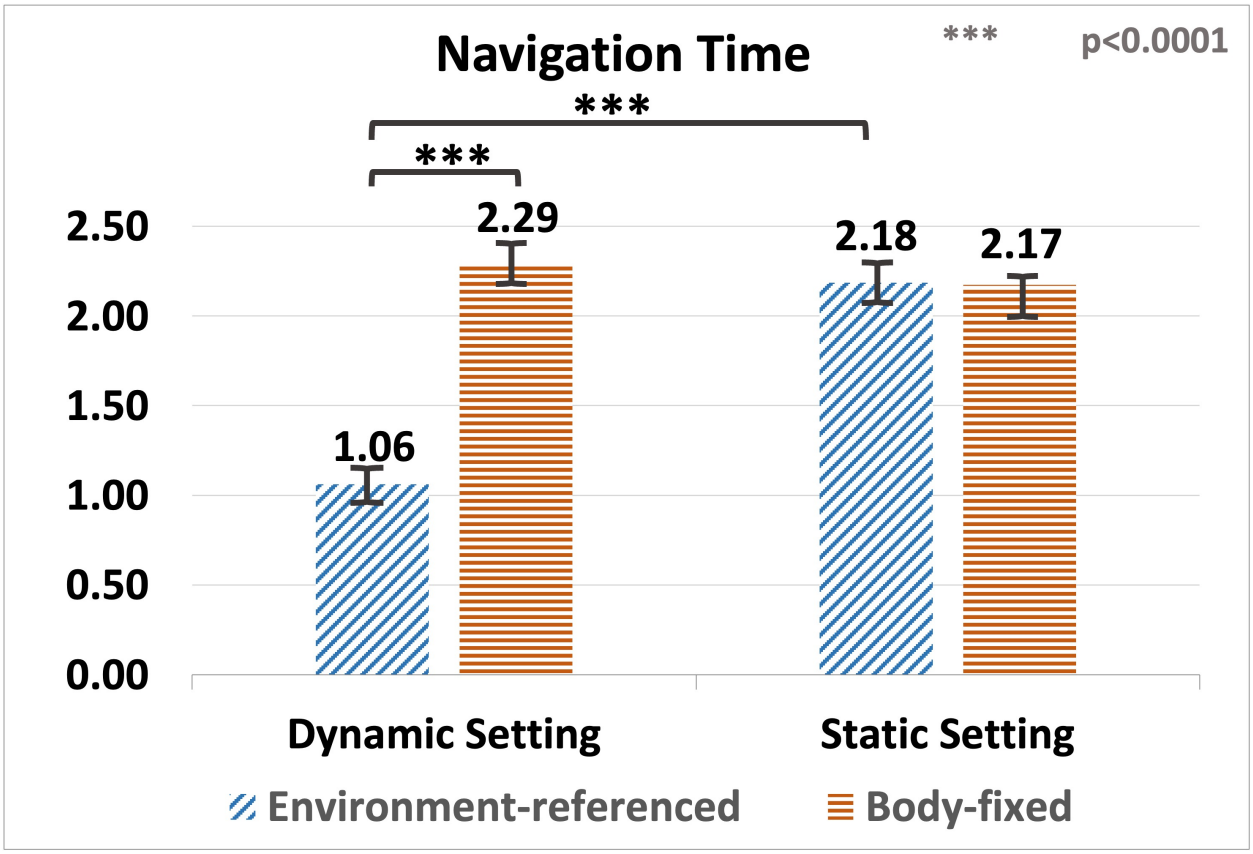


Figure 4.9: Time efficiency within various RW Settings

As shown in Figure 4.10, our effect analysis indicates that the number of *Gaze Switch* fluctuations within correct trials, is significantly affected by *RW Setting* X *User State* X *Placement Strategy* ($\boldsymbol{F_{(1,770.32)} = 18.6,\ p < 0.0001}$), *RW Setting* X *Placement Strategy* ($\boldsymbol{F_{(1,770.91)} = 154.24,\ p < 0.0001}$), and *User State* X *Placement Strategy* ($\boldsymbol{F_{(1,769.8)} = 56.69,\ p < 0.0001}$).

Pairwise comparison on *RW Setting* X *User State* X *Placement Strategy* revealed that gaze switches between the category panels when using *Environment-referenced* placement were significantly fewer **a)** compared to *Body-fixed* in the *Dynamic Mobile* ($\boldsymbol{p < 0.0001}$), *Static Mobile* ($\boldsymbol{p \sim 0.001}$), and *Dynamic Stationary* ($\boldsymbol{p \sim 0.009}$) contexts, and **b)** in the *Dynamic Mobile* context compared to the *Static Mobile* one ($\boldsymbol{p < 0.0001}$). Additionally, this analysis

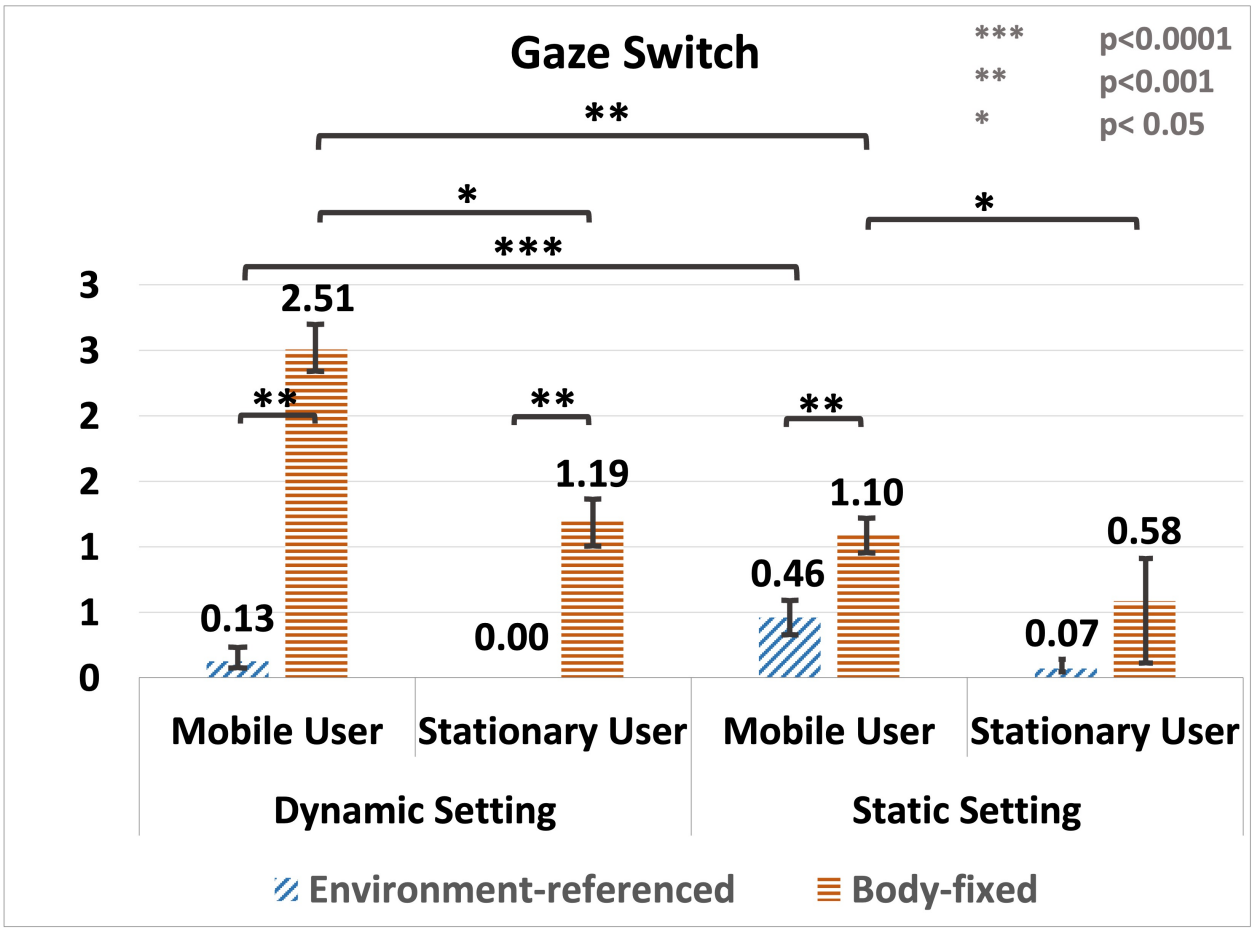


Figure 4.10: Gaze switch efficiency within various contexts

indicated that gaze fluctuations when using *Body-fixed* placement were significantly fewer in **c)** the *Dynamic Stationary* context rather than the *Dynamic Mobile* ($\boldsymbol{p} \sim \mathbf{0.005}$), **d)** the *Static Stationary* context rather than the *Static Mobile* ($\boldsymbol{p} \sim \mathbf{0.035}$), **e)** in the *Static Mobile* context compared to the *Dynamic Mobile* ($\boldsymbol{p} \sim \mathbf{0.006}$).

These results revealed that compared to *Body-fixed*, the *Environment-referenced* placement reduces gaze switches in *Dynamic* contexts ($\boldsymbol{p} < \mathbf{0.0001}$, supporting **H2-a**), *Static* onces ($\boldsymbol{p} < \mathbf{0.0001}$, rejecting **H2-b**), and *Mobile* contexts ($\boldsymbol{p} < \mathbf{0.0001}$, rejecting **H3**).

Perceived Speed:

Participants' post-context ranking of the fastest placement strategy revealed a significant

interaction effect of *RW Setting* X *Placement Strategy* ($\boldsymbol{F_{(1,161)} = 135.01,\ p < 0.0001}$). Pairwise comparison of the effect of these interactions revealed that *Environment-referenced* was ranked **a)** faster than *Body-fixed* in the *Dynamic* contexts by 86% (supporting **H2-a**), **b)** slower than *Body-fixed* in the *Static* contexts by 84% ($\boldsymbol{p < 0.0001}$) (supporting **H2-b**), and **c)** faster in the *Dynamic* contexts compared *Static* ones ($\boldsymbol{p < 0.0001}$, supporting **H1**). Additionally, **d)** *Body-fixed* was ranked higher in *Static* contexts compared to *Dynamic* ones ($\boldsymbol{p < 0.0001}$, supporting **H1**).

Perceived Ease:

Participants' post-context ranking of the easiest placement strategies revealed a significant interaction effect of *RW Setting* X *Placement Strategy* ($\boldsymbol{F_{(1,161)} = 231.4,\ p < 0.0001}$). *Environment-referenced* was ranked **a)** easier than *Body-fixed* in the *Dynamic* contexts by 93% (supporting **H2-a**), **b)** less easy than *Body-fixed* in the *Static* contexts by 80% ($\boldsymbol{p < 0.0001}$) (supporting **H2-b**). Indicating that **c)** *Environment-referenced* placement was perceived as easier in *Dynamic* contexts compared to *Static* ones ($\boldsymbol{p < 0.0001}$, supporting **H1**), while **d)** *Body-fixed* placement was periceved significantly more difficult in *Static* contexts compared to *Dynamic* ones ($\boldsymbol{p < 0.0001}$, supporting **H1**).

NASA TLX:

Our pairwise comparison on the effect of *RW Setting* X *Placement Strategy* on the **TLX Score**($\boldsymbol{F_{(1,161.06)} = 11.68,\ p < 0.001}$) indicates significantly lower overall task loads, **a)** in *Dynamic* contexts, when using *Environment-referenced* compares to *Body-fixed* ($\boldsymbol{p \sim 0.001}$, supporting **H2-a**), and **b)** when utilizing *Body-fixed* in *Static* contexts rather than *Dynamic* ones ($\boldsymbol{p \sim 0.002}$, supporting **H2-b**).

Analyzing the metrics from within the TLX questionnaire identified that *Effort* significantly varied dependent on the specific *RW Setting* X *User State* X *Placement Strategy*

($\boldsymbol{F_{(1,161.08)} = 4.51,\ p \sim 0.03}$, providing support for **H1** and **H2**), and *RW Setting* X *Placement Strategy* significantly impact *Physical Demand* ($\boldsymbol{F_{(1,161.08)} = 6.25,\ p \sim 0.01}$), *Mental Demand* ($\boldsymbol{F_{(1,161.10)} = 6.24,\ p \sim 0.01}$), and *Effort* ($\boldsymbol{F_{(1,161.08)} = 10.4,\ p \sim 0.001}$). A pairwise comparison of these metrics shows that:

Significantly less **Effort** was required **a)** in *Dynamic* contexts, when using *Environment-referenced* compared to *Body-fixed* ($\boldsymbol{p < 0.001}$, supporting **H2-a**), **b)** when utilizing *Body-fixed* in *Dynamic* contexts rather than *Static* ones ($\boldsymbol{p \sim 0.04}$, supporting **H1**), and **c)** when using *Environment-referenced* in *Static Stationary* contexts, rather than *Static Mobile* ones ($\boldsymbol{p \sim 0.02}$, providing partial support for **H3**).

Mental Demand is significantly decreased, **a)** in *Dynamic* contexts, when using *Environment-referenced* compared to *Body-fixed* ($\boldsymbol{p \sim 0.001}$, supporting **H2-a**), and **b)** when using *Body-fixed* in *Static* contexts rather than *Dynamic* ones ($\boldsymbol{p \sim 0.01}$, supporting **H1**).

Physical Demand is significantly lower, **a)** in *Static* contexts, when using *Body-fixed* compared to *Environment-referenced* ($\boldsymbol{p \sim 0.02}$, supporting **H2-b**), and **b)** when utilizing *Body-fixed* in *Static* contexts rather than *Dynamic* ones ($\boldsymbol{p \sim 0.01}$, supporting **H1**).

Usability

Our Analysis indicates a significant interaction effect of *RW Setting* X *Placement Strategy* for *Perceived Intrusiveness* ($\boldsymbol{F_{(1,161)} = 34.82,\ p < 0.0001}$), *Preferred Placement* ($\boldsymbol{F_{(1,161)} = 246.45,\ p < 0.0001}$), and *Overall SUS Scores* ($\boldsymbol{F_{(1,161.05)} = 11.6,\ p < 0.001}$).

Perceived Intrusiveness:

RW Setting X *Placement Strategy* significantly influenced *Perceived Intrusiveness* ($\boldsymbol{F_{(1,161)} = 34.82,\ p < 0.0001}$). The majority of our participants (79.2% within the *Static* contexts,

and 56.3% within the *Dynamic* ones), insisted that both placement strategies were non-intrusive and refused to choose any one as the most intrusive. In *Static* contexts, only 6.3% ranked *Body-fixed* as intrusive. On the contrary, only 4.2% ranked *Environment-referenced* as intrusive for *Dynamic* contexts.Our pairwise comparison indicates that **a)** in *Dynamic* contexts, *Environment-referenced* is significantly less intrusive than *Body-fixed* ($\boldsymbol{p < 0.0001}$, supporting **H2-a**), **b)** *Environment-referenced* is significantly less intrusive in *Dynamic* contexts compared to the *Static* ones ($\boldsymbol{p < 0.0001}$, supporting **H1**), and **c)** *Body-fixed* is significantly less intrusive in *Static* contexts compared to the *Dynamic* ones ($\boldsymbol{p < 0.0001}$, supporting **H2-b**).

Preferred Placement:

ART RM-ANOVA results on *Perceived Intrusiveness* indicate a significant interaction effect of *RW Setting* X *Placement Strategy* ($\boldsymbol{F_{(1,161)} = 246.45,\ p < 0.0001}$). 90% of our participants ranked *Environment-referenced* as the most preferred placement strategy for information acquisition in *Dynamic* contexts, and 82% ranked *Body-fixed* as the most preferred for *Static* contexts. Our pairwise comparison indicates that **a)** in *Dynamic* contexts, significantly more participants preferred *Environment-referenced* over *Body-fixed* ($\boldsymbol{p < 0.0001}$, supporting **H2-a**), **b)** in *Static* contexts, significantly more participants preferred *Body-fixed* over *Environment-referenced* ($\boldsymbol{p < 0.0001}$, supporting **H2-b**),
c) *Environment-referenced* is significantly more preferred for *Dynamic* contexts compared to *Static* ones ($\boldsymbol{p < 0.0001}$, supporting **H1**), and **d)** using *Body-fixed* is significantly more preferred for *Static* contexts compared to *Dynamic* ones ($\boldsymbol{p < 0.0001}$, supporting **H1**).

SUS:

Our Analysis of the SUS responses indicate significant effect of *RW Setting* X *Placement Strategy* on the **overall Usability Score** ($\boldsymbol{F_{(1,161.05)} = 11.6,\ p < 0.001}$). Delving deeper into the metrics from within SUS, found that other than the *overall Usability Score*, more

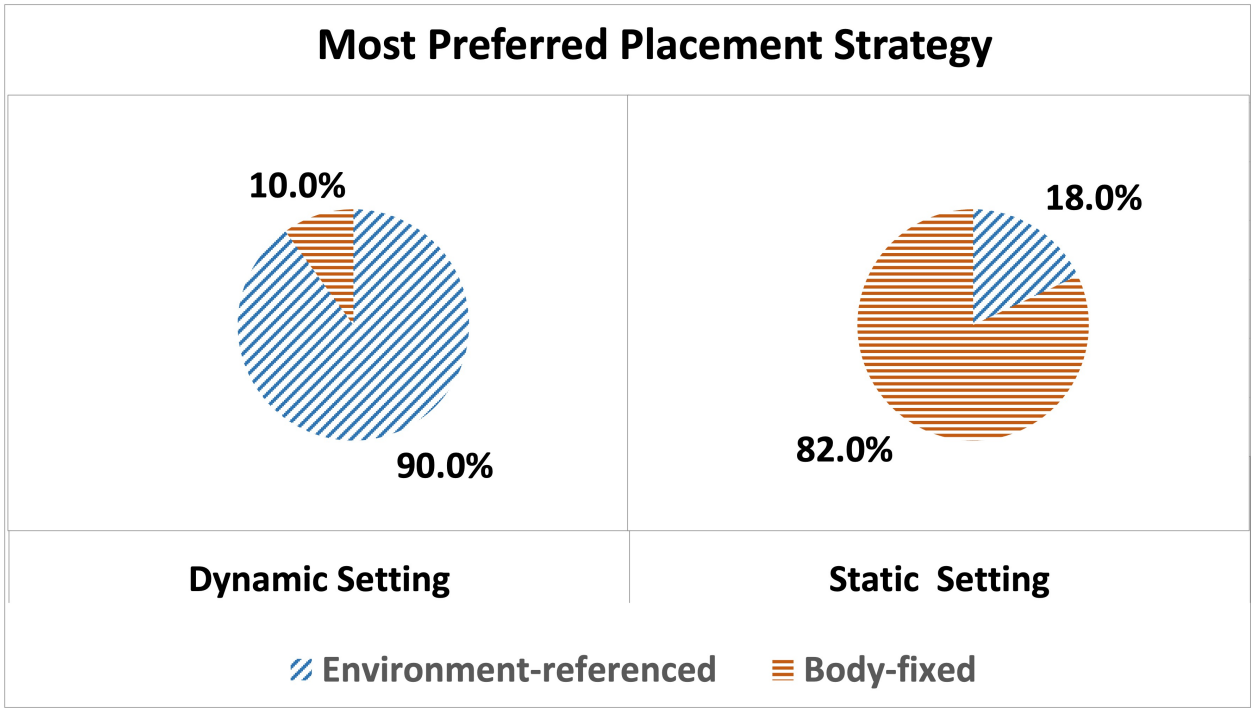


Figure 4.11: User preference on placement strategy in various RW settings

specifically, *User Satisfaction* ($\boldsymbol{F_{(1,161.12)} = 16.37,\ p < 0.0001}$), *Synergy* ($\boldsymbol{F_{(1,161.1)} = 8.62,\ p \sim 0.003}$), *Consistency* ($\boldsymbol{F_{(1,161.07)} = 7.4,\ p \sim 0.007}$), *Learnability* ($\boldsymbol{F_{(1,161.09)} = 3.62,\ p \sim 0.05}$), and *Effortlessness* ($\boldsymbol{F_{(1,161.13)} = 5.55,\ p \sim 0.02}$) were significantly affected by *RW Setting* X *Placement Strategy*. Pairwise comparisons on these measurements provide support for **H2-a**, indicating that in *Dynamic* contexts, *Environment-referenced* significantly improves overall usability score ($\boldsymbol{p < 0.0001}$) and provides a significantly more satisfactory ($\boldsymbol{p < 0.0001}$), well-integrated ($\boldsymbol{p < 0.0001}$), and consistent ($\boldsymbol{p \sim 0.04}$) interface that better facilitates learning ($\boldsymbol{p < 0.0001}$) and effortless document navigation ($\boldsymbol{p \sim 0.004}$) compared to *Body-fixed*. These analyses also support **H1** by showing that **a)** utilizing *Environment-referenced* resulted in significantly higher user satisfaction in *Dynamic* contexts compared to the *Static* ones ($\boldsymbol{p \sim 0.03}$), and **b)** utilizing *Body-fixed* results in significantly higher user satisfaction in *Static* contexts compared to the *Dynamic* ones ($\boldsymbol{p \sim 0.005}$).

4.4.8 Discussion

In this section, we evaluate the user study results, examining their alignment with our hypotheses. We explore alternative hypotheses to explain any discrepancies between our initial hypotheses and the findings, offering insights into the study's limitations and proposing directions for future research.

Evaluation of Our Hypotheses

Our results revealed that the effectiveness of *Environment-referenced* varies across different contexts, excelling in certain scenarios while falling short compared to *Body-fixed* in others. For each of our measurements, the factorial analysis revealed at least one significant interaction effect involving *Placement Strategy* and one or both of our contextual components, i.e., *RW Setting* and *User State*. Considering *Environment-referenced* placement, the participants switched their gaze between panels less often, navigated the documents faster, and expressed overall better feedback on the usability and efficiency of the placement strategy, when in *Dynamic* contexts, compared to *Static* ones. *Body-fixed* placement increased the user's gaze switches within the *Dynamic* contexts compared to the *Static* contexts. Additionally, considering the *User State*, when *Mobile*, this placement increased gaze switches within a *Dynamic* setting but reduced it within the *Static* setting. This underscores the impact of contextual scenario on the efficacy of placement strategy, thereby supporting **H1**.

To evaluate **H2**, we investigated the performance of *Environment-referenced* information acquisition, compared to *Body-fixed* in our *Dynamic* (**H2-a**) and *Static* (**H2-b**) contexts. All of our measurements consistently support **H2-a**, indicating that in *Dynamic* contexts *Environment-referenced* offers a similar level of physical efficiency as *Body-fixed*, with significantly lower cognitive load, optimizing accuracy, speed, ease, and usability of information

acquisition. This underscores the importance of context awareness, suggesting that in contexts where the relevant intermediary objects are dynamic, applying *Environment-referenced* placement in a context-aware interface could enhance the user experience.

Our data yielded inconclusive outcomes regarding **H2-b** and **H3**.While participants' post-context ranking of the placement strategies and the Overall TLX score support **H2-b**, indicating that users perceive *Body-fixed* to outperform *Environment-referenced* in *Static* contexts, the *Error Rate* and *Gaze Switch* measures contradict this hypothesis, indicating lower accuracy and efficiency of *Body-fixed* placement. Similarly, according to our results, the *User State* only influences placement strategy performance in terms of *Gaze Switch* and the *Effort* subscale of the TLX. Both these metrics indicate the inferior performance of *Body-fixed* in the *Mobile* contexts and refute **H3**.

Influence of Info-focus Relevance

Our results strongly confirm **H1** and **H2-a** while offering partial support for **H2-b** and partially rejecting **H3**. The inconsistency in our results can be attributed to a subtle interplay among our contextual components, which influences the suitability of the intermediaries. Across these contexts and even within specific ones, the user's focus on specific objects within the room varied. We used the user's primary conversation partner or location within each trial to extract the user's primary focus. While the choice of intermediaries ensures that the information access needs are relevant to the intermediaries, within each trial, the relevance of the information being accessed to the user's primary focus differs. This discrepancy impacted the efficacy of the *Environment-referenced* placement strategy placement strategy across diverse contexts. We use the term *Info-focus Relevance* to refer to the degree of relevance between information access need and the user's current focus within each trial. The following discussion delves deeper into *Info-focus Relevance*'s effect on placement strategy performance

in our contexts.

***Dynamic Contexts*:** Within our *Dynamic* contexts, the user was always focused on their conversation partner. Additionally, in these contexts, regardless of the *User State*, the information needs within a trial were always *Relevant* to the same trivia host the user was interacting with. Due to the high *Info-focus Relevance*, the intermediaries are suitable for all trials in the *Dynamic* contexts. This enables *Environment-referenced* placement to always provide effortless access to the required document through a forward glance and eliminate the cognitive load associated with navigating to the correct category panel. This can explain our results regarding **H3**, as the *Environment-referenced* performs significantly better than *Body-fixed* in *Dynamic* contexts, even when the *User State* was *Mobile*.

***Static Contexts*:** In the *Static* contexts, the user's focus and, therefore, *Info-focus Relevance* vary based on the *User State*. The *Static Mobile* context involved the user moving to one of the intermediary objects and maintaining visual attention to it. In this context, half of the trials require information relevant to the same category as the nearby intermediary object. Consequently, only half of the trials in this context were *Relevant* in terms of *Info-focus Relevance*. On the other hand, within all the *Static Stationary* trials, the user was facing forward and focused on the Sports category's poster. In this context, only **1/3** of the trials require information relevant to the Sports category and were considered as *Relevant* in terms of *Info-focus Relevance*.

To better understand *Info-focus Relevance*, we conducted a factorial non-parametric ART RM-ANOVA to test the interaction effects of *User State*, *Placement Strategy* and *Info-focus Relevance* on the quantitative trial data within the Static contexts. The results indicate significant influence of *Placement Strategy* and *Info-focus Relevance* on *Gaze Switch* ($\boldsymbol{F_{1,372.59} = 40.02, p < 0.0001}$) (See Figure 4.12) and significant effect of *User State* X

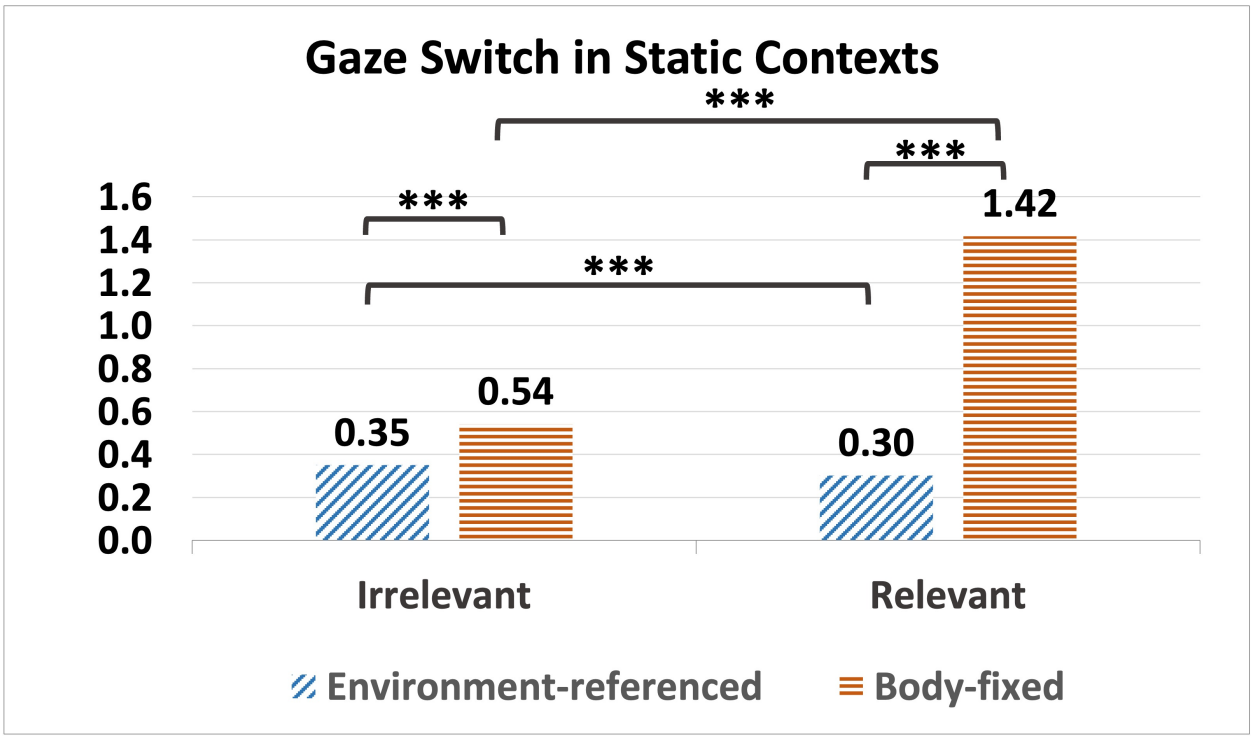


Figure 4.12: The influence of Info-focus Relevance on gaze switch efficiency within static contexts

Placement Strategy X *Info-focus Relevance* on *Navigation Time* ($\boldsymbol{F_{1,372.39} = 11.65, p < 0.001}$) (See Figure 4.13).

***Static Stationary*:** The *Static Stationary* context entails a stationary user positioned between the three posters (intermediaries), maintaining a consistent orientation and location. Consequently, *Environment-referenced* appears almost identical to *Body-fixed* and imposes equivalent physical and mental load levels. This is contradictory to our **H2-b** and partially explains the lack of support for this hypothesis in our findings. Additional nonparametric analyses using Wilcoxon Rank Sums tests on our objective data, NASA TLX, and SUS measurements within this context indicate no significant effect of *Placement Strategy* on any of our metrics.

***Static Mobile*:** We conducted nonparametric analyses using Wilcoxon Rank Sums tests

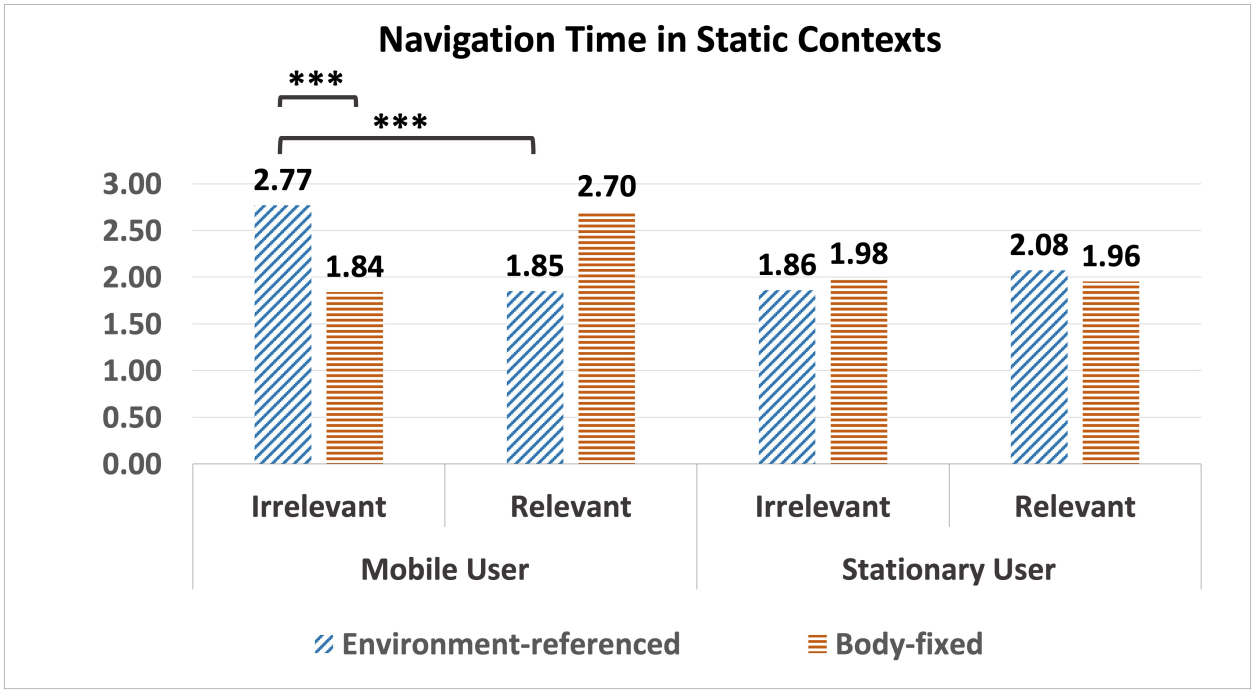


Figure 4.13: The influence of Info-focus Relevance on time efficiency within static contexts

to examine the effect of *Placement Strategy* within this context. The results indicated that *Environment-referenced* placement in this context resulted in significantly lower *Gaze Switch* compared to *Body-fixed* ($\boldsymbol{\chi^2 = 11.84, Z = 3.44, p \sim 0.0006}$). This helps further explain the contradictory results to **H2-b** and **H3**.

Limitations

In each combination of our contextual components (*RW Setting* and *User State*), the user's focus within the environment varied. This resulted in the involvement of *Info-focus Relevance* as an independent variable, influencing the effectiveness of *Environment-referenced* placement strategy. On the other hand, *Body-fixed* offers reliable and effortless access to all documents through a glance, irrespective of the particular context and *Info-focus Relevance*. While this placement strategy imposes an additional mental load to select the required infor-

mation, this mental load is consistent across different contexts. Additionally, as mentioned by our participants, due to the limited number of category panels within this study, this mental load is relatively easy.

Significance of iAR

Our results indicate that contingent upon the user's focus and interactions with their setting, the choice of optimal placement strategy within each context varies. Within our experiment, the *Environment-referenced* strategy utilized relevant people and objects from the environment as intermediaries to position the XR content. On the other hand, the *Body-fixed* strategy provided a consistent glanceable interface, offering effortless access to all content, regardless of the specific context.

These results indicate the significance of context in the choice of optimal placement strategy and highlight the significance of *Info-focus Relevance* on the effectiveness of *Environment-referenced* placement strategy across different contexts. In contexts with more *Relevant Info-focus* trials, *Environment-referenced* intelligently positioned each piece of AR content precisely where needed, thereby enhancing efficiency while mitigating information overload. However, with *Irrelevant Info-focus*, *Environment-referenced* diminished performance as it necessitated excessive head and body movements towards the intermediaries and was inconsistent when the user or the environment was dynamic. These observations indicate the significance of *RW Setting* and *User State* in determining *Info-focus Relevance*. More specifically, within our contexts, the interlocutor, the user's proximity to the intermediary objects, and the mobility of the user influenced the user's focus and information needs.

In a context-switching scenario, an iAR interface is necessary to evaluate the cognitive and physical impact associated with various placement strategies and determine the optimal

placement for the specific context. Such iAR interfaces must infer information such as *Info-focus Relevance* from the context. In addition, future work must explore various ways for such interfaces to provide feedback to users when the placement strategy changes.

4.5 Conclusions

Towards context-aware adaptations in iXR, in Section 4.2, we explored various XR design elements that can be adapted, providing a comprehensive design space for XR and addressing *RQ5.*.

To address *RQ3.*, we concentrated on the content's spatial layout from our design space and explored context-aware placement strategies for XR objects. Section 4.3 proposed utilizing a relevant intermediary from the environment within a Hybrid Frame of Reference (FoR) for each XR object. Section 4.4 investigated the effectiveness of this context-aware placement strategy and a non-adaptive *Body-Fixed* placement strategy in four contextual scenarios varying in *RW Setting* and *User State* (*RQ4.*). Our findings underscore the significance of iXR interfaces that determine the optimal *Placement Strategy* through inferring information such as *Info-focus Relevance* from the context.

This work was submitted to TVCG and is currently under review:

- [27] Shakiba Davari, and Doug A. Bowman. "Towards Context-Aware Adaptation in Extended Reality: A Design Space for XR Interfaces and an Adaptive Placement Strategy." IEEE Transactions on Visualization and Computer Graphics, Jun 2024, 12 pages

Chapter 5

Taxonomy of Context & iAR Framework

5.1 Overview

iAR systems aim to utilize context and optimize the AR design dimensions identified in Section 4.2. To effectively utilize context, two criteria must be met: First, the context taxonomy should encompass all quantifiable contextual components that influence AR effectiveness. Second, there should be a framework for inferring implicit contextual information using this taxonomy.

Previous research has defined context, identifying primary categories such as location, identity, time, and activity [2, 110, 111]. These categories form the foundation for understanding and detecting context in diverse applications. Within the field of AR, Grubert et al. expanded on this by categorizing context into user, physical environment, and social/cultural context [56]. This taxonomy offers a more nuanced understanding of context in Real-World(RW) applications of AR, thereby facilitating the development of context-aware and adaptive AR interfaces. Building on these frameworks, in this chapter, we refine and expand the taxonomy of context, focusing on quantifiable components and providing a framework for inferring implicit contextual information within iAR systems.

Section 5.2 presents a comprehensive taxonomy of quantifiable contextual components, including those related to the user, the user's setting, and their interplay (*RQ6.*). Utilizing this taxonomy, Section 5.3 outlines our proposed framework to design iAR systems that receive context as their input, infer the optimal interface, and make the proper adaptations accordingly (*RQ7.*).

5.2 Taxonomy of Context

To inform the design and development of iAR, it is essential to describe the context comprehensively through data that can serve as input for these systems. The context dictates the user's specific limitations, wants, and needs, influencing their preferences for the AR interface at each time. An iAR interface must use the input data it receives to identify these user preferences for the interface in each context and make the proper adaptations accordingly (Figure 5.1). We classify the input data provided to iAR into three input types:

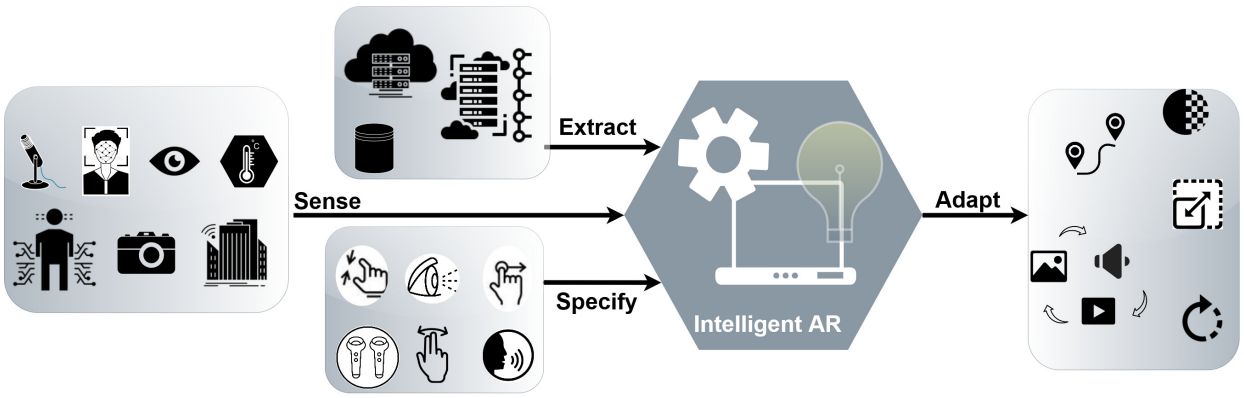


Figure 5.1: To identify context and provide the proper adaptations as output, an intelligent AR interface uses quantifiable data from the user, sensors, and personal or publicly-available data storage and databases.

Specified input are quantifiable information *provided manually* by the user. iAR

aims to limit using this type of input when possible. That is because these inputs are limited to specific human-understandable data and prone to human error. They can also be slow and inaccurate, distract the user, and increase their physical and mental load, which contradicts these interfaces' objective of increasing efficiency and effectiveness.

Sensed input are raw or processed quantifiable information *directly received* from sensors.Sensed inputs from various sensors like a camera, microphone, accelerometer, gyroscope, and magnetometer can gather information about the environment, the user, the device's pose, etc. This input type can also include processed sensor data for recognizing and tracking faces and objects or detecting voice commands, gestures, and gaze.

Extracted input are quantifiable information *retrieved from previously stored* data on various personal or publicly-available data storage and databases through in-network and web queries. These data storage and databases are collections of information that were *specified* or *sensed* in the past. For example, extracted inputs can provide information such as the GPS location and time of day.

Context: We describe *Context* as the knowledge derived from all the available *contextual components* that can affect the AR interface's content, presentation, or interaction. These ***contextual components*** are considered as any piece of *specified*, *extracted*, or *sensed* input and provides *dynamic* or *persistent* information about the context.

Persistent contextual components are static or change infrequently and are quantified by *extracted* inputs. Information about the user's education, local time zone, or family members is a few examples of persistent contextual components. While recording a change in these contextual components is through *specified* or *sensed* inputs, they will be stored on

an internal or external data storage and *extracted* when accessed for context detection.

Transient contextual components are prone to more frequent and unexpected changes and need to be quantified more frequently through *sensed* or *extracted* input, e.g., user's body gesture, speech content, environmental lighting.

To design a taxonomy of explicit and quantifiable contextual components that influence the effectiveness of AR interfaces, we adopted a comprehensive and systematic methodology. Initially, to avoid biases present in existing literature, we conducted multiple brainstorming sessions to enumerate all conceivable contextual components. During this process, we explored the XR design space and identified design elements that can adapt the content and presentation of an XR interface [27]. We thoroughly evaluated the challenges and capabilities associated with these design elements and compiled a list of contextual components that could influence them. For instance, when considering the color theme as a design element, a pertinent challenge is that blue colors are not suitable for nighttime use and can cause eye fatigue. This underscores the importance of "the time of day" as a crucial contextual component. As another example, the position, size, and opacity design elements of an AR object can result in the occlusion of important objects in RW. This emphasizes the vitality of contextual components providing information about the user's interplay with objects and people within the setting.

Next, to ensure the inclusion of all pertinent aspects from the literature, we reviewed well-established taxonomies for describing context and analyzed the contextual information utilized in various proposed context-aware AR interfaces. These steps enabled us to compile a comprehensive list of contextual information. Subsequently, we engaged in an iterative refinement process, continuously reassessing the relevance of each contextual component, ensuring a robust and effective taxonomy.

Abowd et al. establish location, identity, time, and activity as the main context categories [111], Grubert et al., categorize context into user, physical environment, and social/cultural context, other research has focused on adaptations based on objects [70, 140], background [95], depth perception [44], and physical environment [22, 52] from the context. We identify three main categories of quantifiable contextual components that influence AR effectiveness either directly or indirectly through inferences (e.g., see the scenario of Sanaz working from home in Section 5.3). While the Setting category provides information about the state of the world, excluding the user, the User category focuses solely on the user, and the Setting-User Interplay category provides information about the interactions between context components from these categories (Figure 5.2).

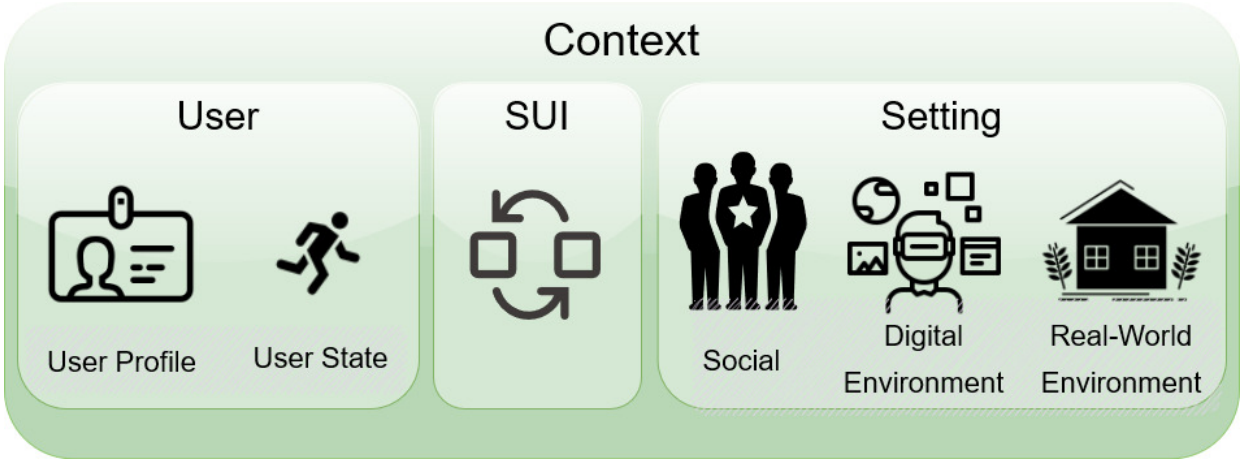


Figure 5.2: Our taxonomy identifies three principal categories of components within context: User, Setting, and Setting-User Interplay.

5.2.1 User

Who the user is, what they are currently doing, how they feel, and other information about the user themselves affect the type of information they need and their preferences, resources, and limitations for accessing and interacting with this information. For example,

a chef may need less detailed instructions from a recipe, while a beginner requires the same recipe to be presented in more detail and include visualizations. In our taxonomy, we classify the contextual components only about the user, isolated from their setting, as the user category of context. We classify the contextual components within the User category into two subcategories: User Profile and User State (Figure 5.3).

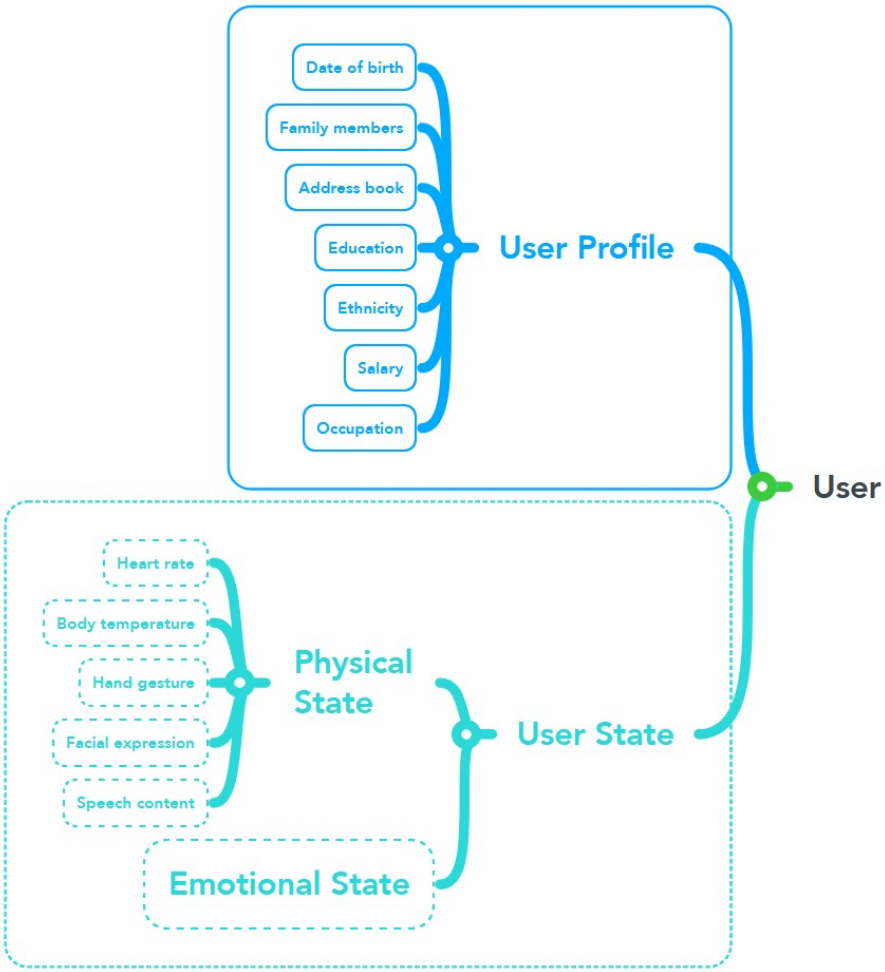


Figure 5.3: User is described by persistent context components about the user profile and dynamic context components about their current state.

Persistent User Profile

Persistent information about users and their characteristics, such as their values and

responsibilities, sociocultural factors, socioeconomic status, psychological and physiological abilities, and traits, can affect their interface. For example, a visually impaired person may require a different font size than a person with normal vision.

Persistent User profile provides established information about the user through *persistent contextual components*. These are contextual components such as meaningful dates, locations, and people, user's physical limitations, assets, properties, weekly schedule, salary, occupation, and responsibilities that can be *extracted* from a user's digital profiles, health records, legal documents and contracts, bank accounts, etc.

For example, a student specifies their demographic and education information and the classes they are taking on their digital profile. Inputs such as their class schedules and course syllabus *extracted* from this digital profile and the learning management system allow their AR interface to provide the class material for them automatically.

Transient User State

Adaptations to a user's interface can also be affected by transient factors such as their current physical activity, conversation topic, and mental and emotional state. For example, the user's emotional state can affect content adaptation, such as presenting the user's anxiety management tool during a panic attack. *User state* provides transient information about the user through *Transient contextual components* that change continually. These are contextual components such as heart rate, body temperature, gestures, and speech content that can be obtained using sensor data from an altimeter, 3-axis accelerometer, optical and temperature sensor, microphone, camera, gesture sensors, etc.

Since the user state is subject to unpredictable and frequent changes, these components must be *sensed*. However, with accuracy at stake, current technological limitations may require combining or replacing *sensed inputs* with other contextual component types to remove or

reduce reliance on specific sensors. For instance, when proper sensors are unavailable, costly, or require high power consumption, a combination of *specified* and *extracted* inputs can be used as an alternative.

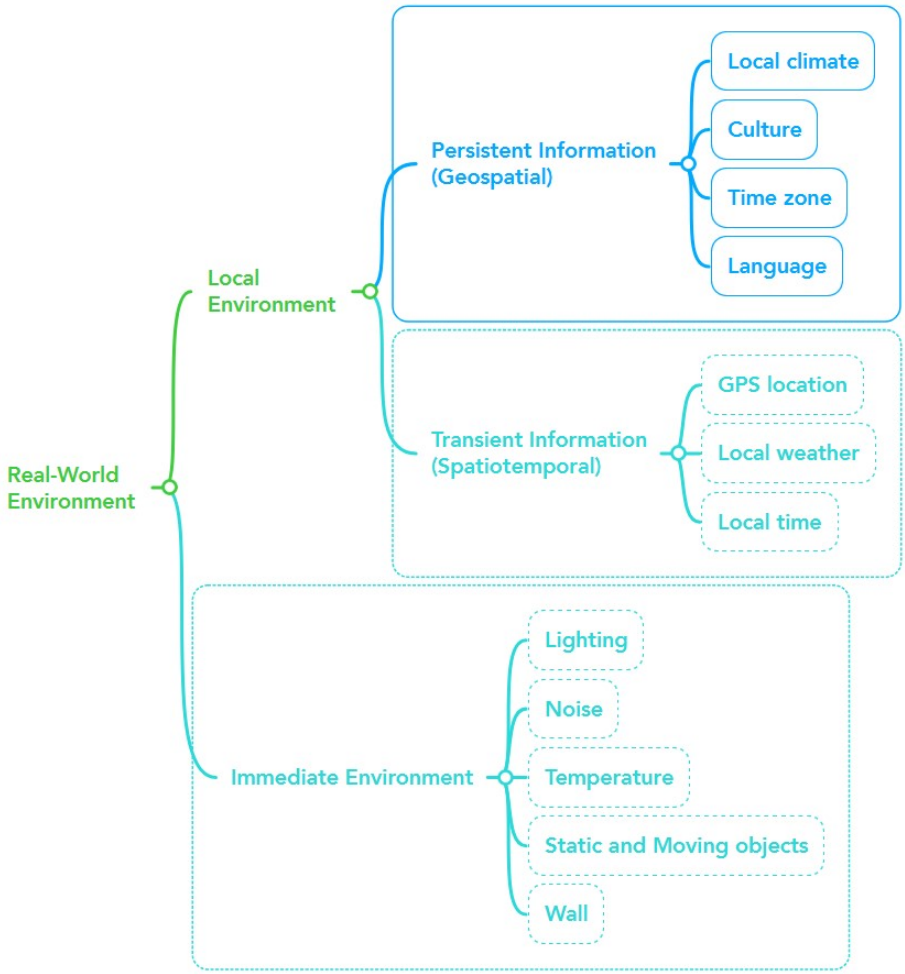


Figure 5.4: The real-world environment, consists of persistent and dynamic information about the local environment, and dynamic context components about the structure and state of the physical items and phenomena in the user's immediate environment.

5.2.2 Setting

By definition, AR integrates the virtual information within their environment [5]. Consequently, in addition to affecting the user's efficiency and effectiveness in their digital data acquisition, AR interfaces influence their experience interacting and perceiving the RW and other people [125]. The state of the user's environment and other people, objects, and digital devices within it constitute a crucial part of the context influencing interface design elements that can cause or prevent conflicts to the user's interaction and awareness of social, RW, and digital information. For instance, to adapt the AR content's layout and avoid interfering with a mobile user's visual attention and interpersonal interactions, contextual components providing information about semantic changes in the environment and social setting are crucial [22, 108].

The user category isolated the user from the context, providing all the contextual information about them. The setting category describes the context without any knowledge of the user's profile and state. The setting is provided through three main subcategories of contextual components describing the: **Real-World Environment**, *Digital Setting*, and **Social Setting** (See Figure 5.5).

Real-World Environment

These contextual components describe the immediate surroundings of the user, the current and local phenomena, and the geographical area (Figure 5.4).

Local Environment is described by *extracted* spatiotemporal and geospatial contextual components about the user's current location.

Persistent Local Environment includes *persistent contextual components* such as geopolitical data, local climate, culture, time zone, language, health resources, and social factors.

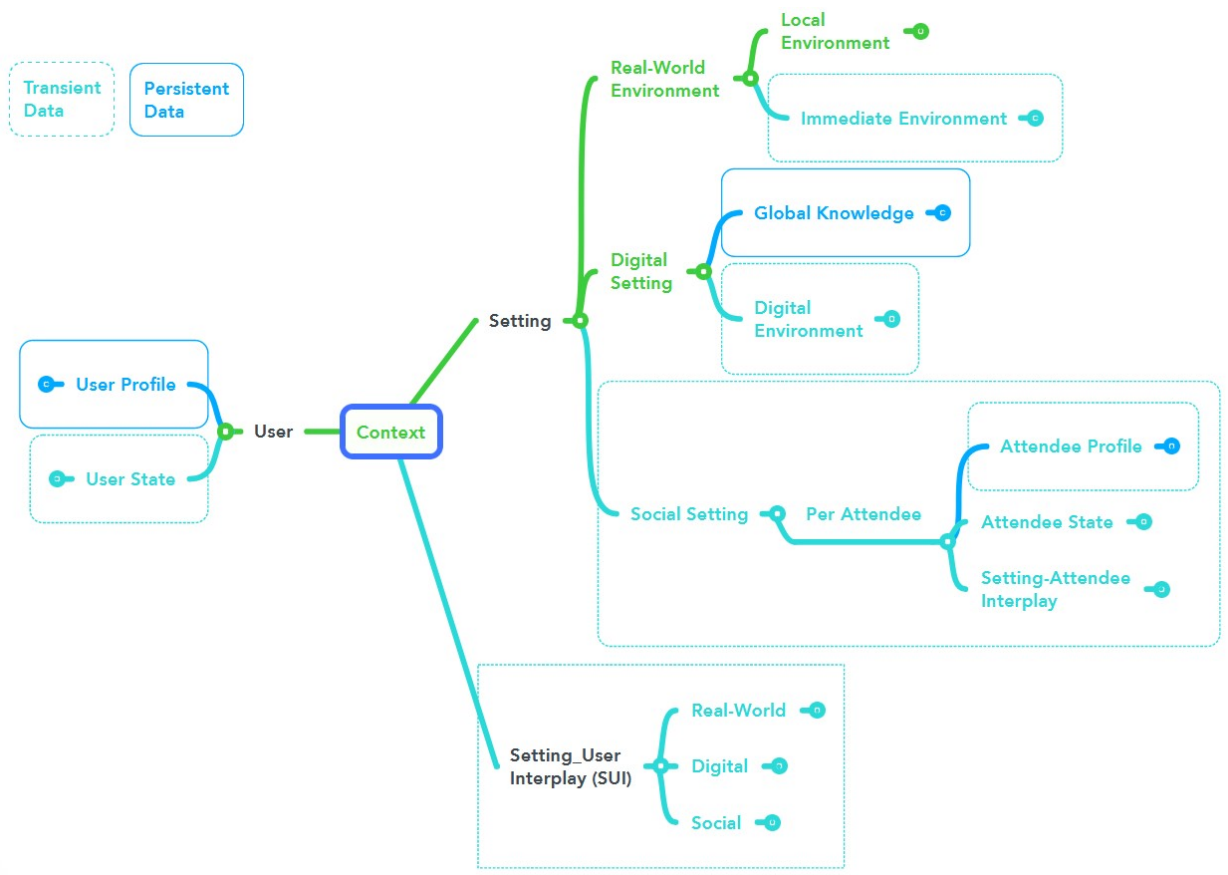


Figure 5.5: The principal subcategories of context components within User, Setting and SUI in our taxonomy of context.

Transient Local Environment includes *Transient contextual components* such as GPS location, the local weather and time, the natural, cultural, and political phenomena, and other spatiotemporal data.

The user's ***Immediate Environment*** changes as they move throughout the day. The immediate environment provides transient information to describe the user's exact surroundings. It includes *Transient contextual components* that must be *sensed* from the sensors in the user's immediate environment. contextual components such as temperature, lighting, noise and audio, spatial map, smoke, nearby objects, their state, and their motions are a few of these components.

Digital Setting

The digital setting includes contextual components that, *extract* information from the

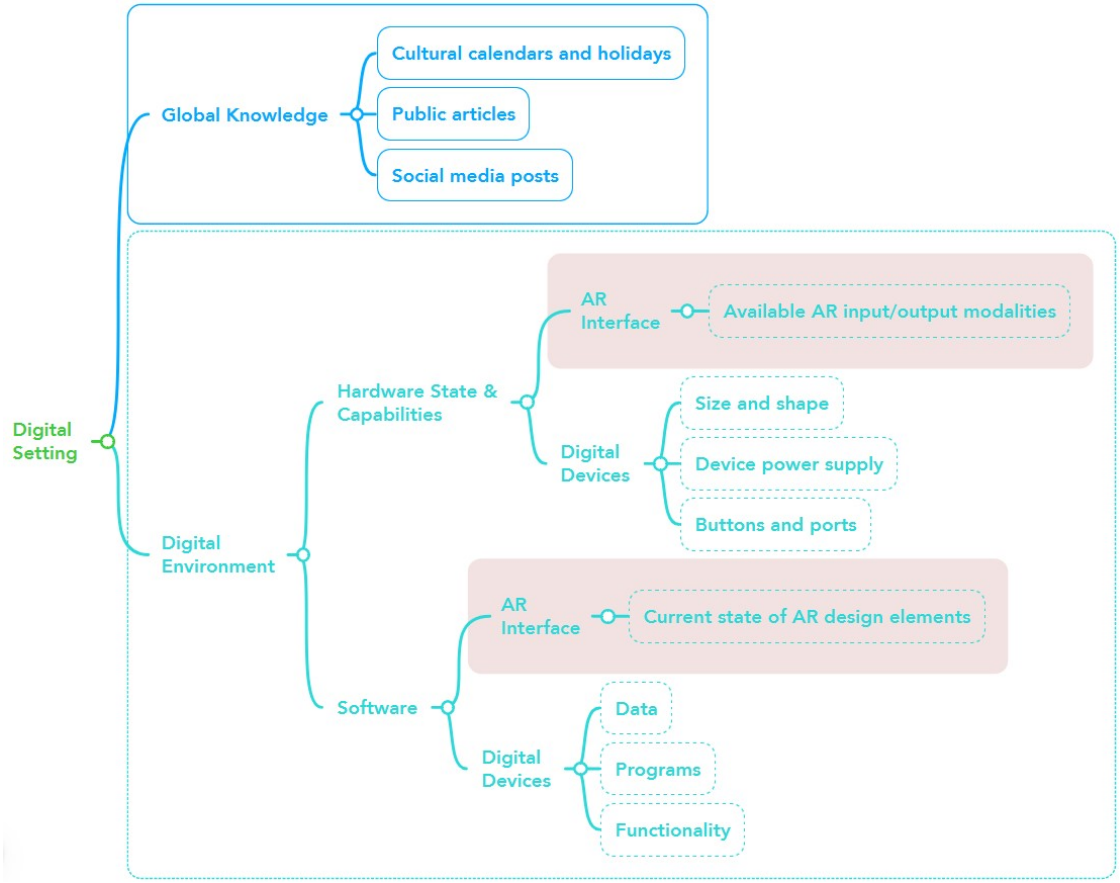


Figure 5.6: Digital setting is a collection of persistent and publicly available global information, and dynamic context components representing the hardware and firmware state and capabilities of available digital devices and the AR interface, including the current state of the AR interface's design elements.

available global knowledge and the immediate digital and virtual interfaces and devices (Figure 5.6).

Persistent Global Knowledge includes *persistent contextual components* extracted from global and common knowledge such as LLMs, search engines, public articles, social media posts, and web pages.

Transient Digital Setting includes *Transient contextual components* that describe the current content, state, and activities of the available digital interfaces and devices within the user's immediate surroundings. The transient digital setting includes contextual components such as the AR interface's available input and output modalities, the current state of its design dimensions, and the nearby devices and their states. For example, when driving an electric car, contextual components, such as the battery level and the car's state of movement, can affect the AR interface.

Social Setting

Throughout the day, changes in the number of people around a user, who they are, and what they do affect the user's AR interface. For instance, when the user is in a public setting, concerns such as social stigma, privacy, and security can affect their preferred interaction modality. As part of the Setting, Social Setting describes the other people involved in a context, without any knowledge of the user. The social setting involves *Transient contextual components* that provide information about the virtual or physical presence, state, and profile of other people, i.e., attendees, in the user context and their interactions with each other and the context(Figure 5.7).

Attendee Profile includes *extracted contextual components* from publicly available data or internal databases that identify the persistent profile of the present attendees. While the user profile data are persistent for each attendee, these attendee profile contextual components are transient due to changes in the presence of each specific attendee. The attendee profile includes each attendee's social status, occupation, relationships, values, responsibilities, etc.

Attendee State provides *sensed inputs* that describe the transient state of each attendee. The attendee state includes contextual components such as face and speech detection, the physical and emotional state of each person, and their motions.

Attendee-Setting Interplay provides *sensed* or *extracted* to provide information about how the attendee's profile and state links with other people, the RW environment, and digital setting within the user's context The attendee sui includes contextual components such as the personal relationships of attendees with each other and their conversation and interactions with other people and the setting.

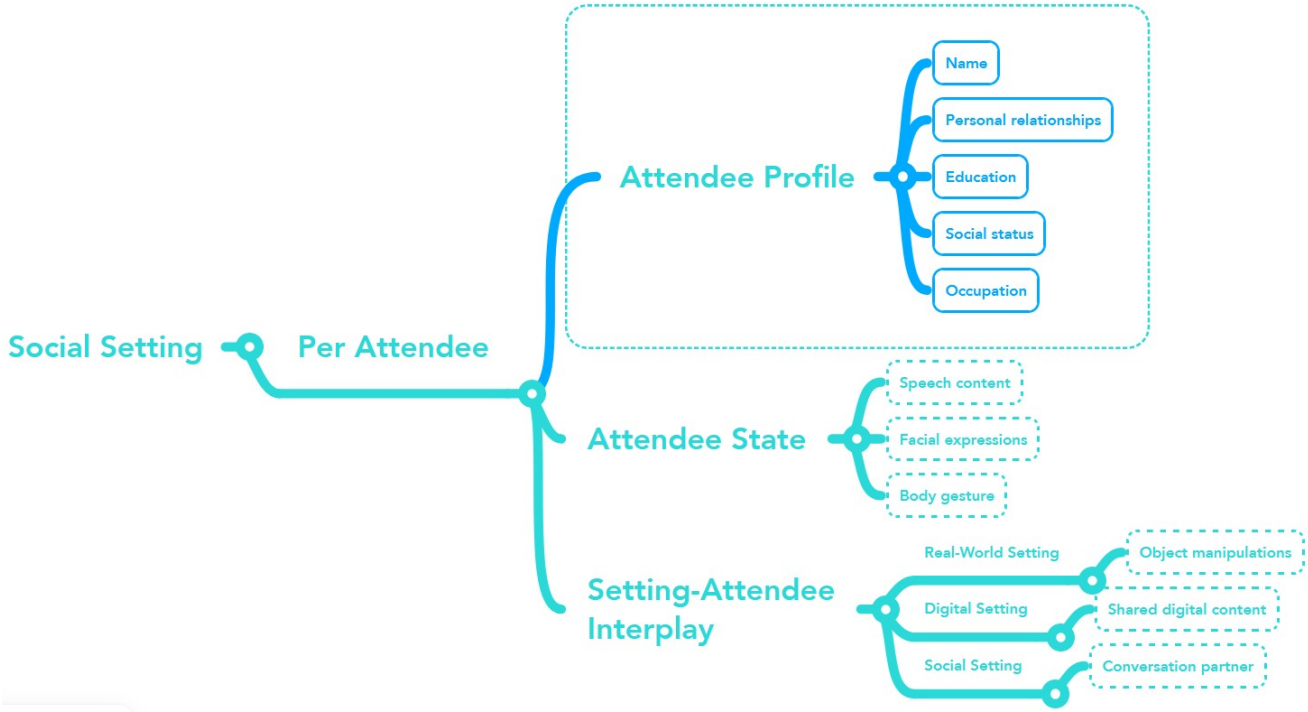


Figure 5.7: Social setting is presented by dynamic context components providing accessible information from the profile and current state of other people (attendees) present in the user context and their interactions with each other and with RW and digital setting.

5.2.3 Setting-User Interplay

Providing effective, efficient, and unobtrusive AR interfaces requires knowledge about the user's needs, preferences, resources, and limitations. Such information can be learned from the user's background, characteristics, current state, activities, and setting. However, the

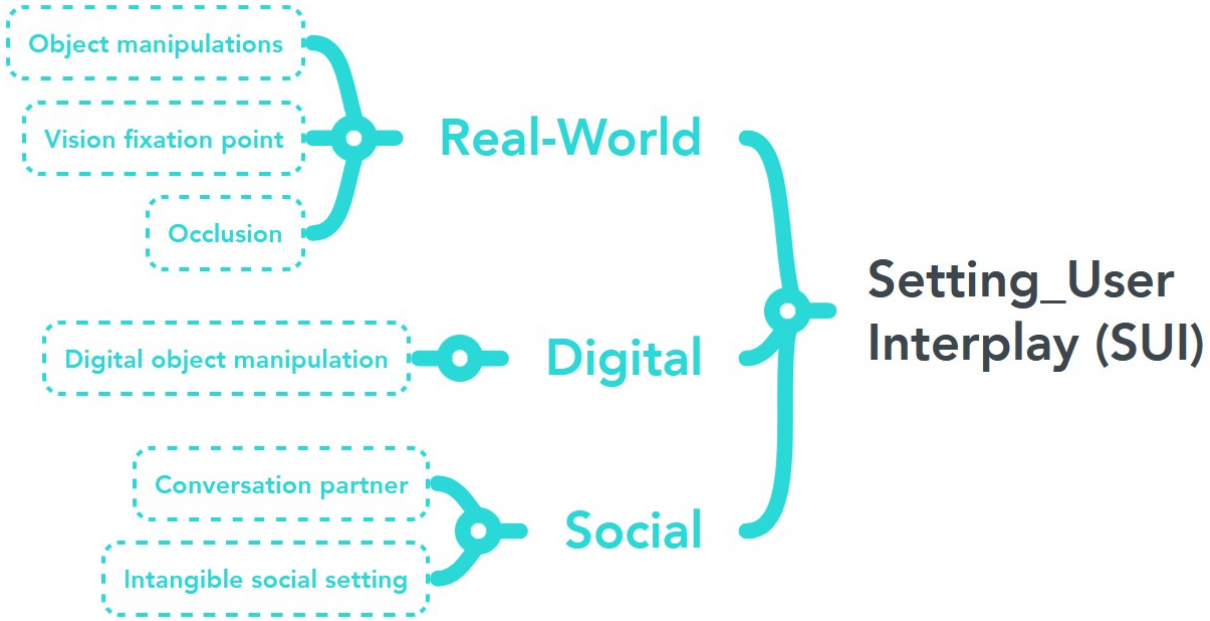


Figure 5.8: SUI includes context components that represent the user's connections and interplay with other people and the RW and digital setting.

interplay between the User and the Setting can lead to conflicts and associations within the context, influencing the effectiveness of the interface. Setting-User Interplay (SUI) includes Transient contextual components that can be *sensed* or *extracted* to provide information about how the user's profile and state links with other people and the RW environment and digital setting (Figure 5.8). SUI includes contextual components such as physical and digital objects, locations, dates, and people that are occluded, related to the user, or interacted with (through physical, verbal, gaze, gesture, digital, social, or emotional interactions). For example, SUI provides *sensed contextual components* on the user's fixation point in the setting. This contextual component can *indirectly* provide information about the people and virtual and physical objects of interest to the user. Similarly, SUI contextual components from face recognition *indirectly* provide information to distinguish the optimal interface for a formal social interaction and an intimate one. To equip the iAR system with all necessary information, our framework proposes an approach to infer the impact of such implicit information

on interface effectiveness from contextual components.

5.3 Framework

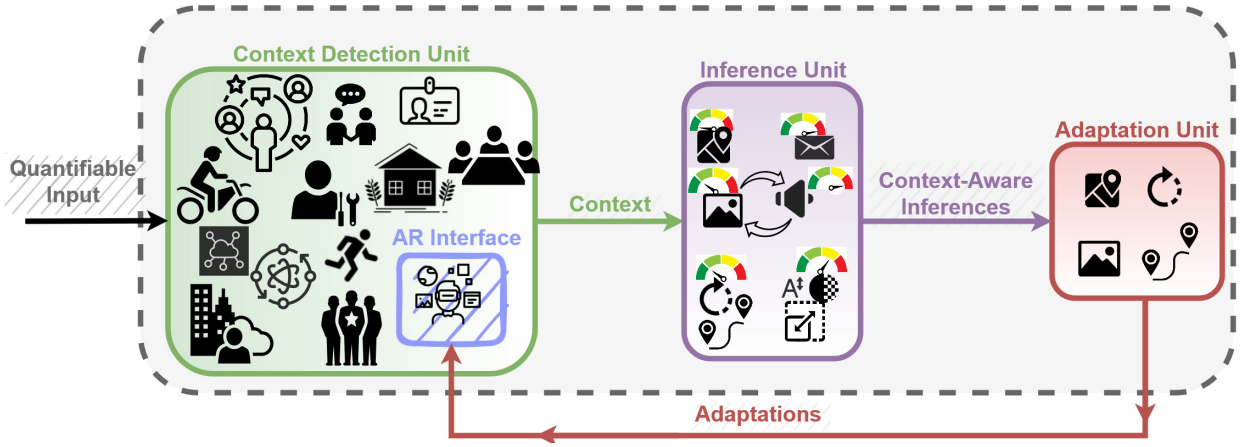


Figure 5.9: Our framework infers potential adjustments for interface performance using the context identified by our taxonomy. The Adaptation unit optimizes these inferences to adapt the interface.

We propose a framework for iAR to utilize context-aware inferences and optimize adaptations that enhance AR effectiveness. This framework defines an **Inference** as the level of the contingency of an interface design dimension on the context (Figure 6.1). Since the AR interface is part of the virtual setting, an inference may involve correlating multiple interface design dimensions. We refer to these inferences, where only interface design dimensions are involved, as *Design Inferences*.

Inferences indicate the potential impact of specific **Adaptations** on the performance of the interface. However, each design dimension can be involved in numerous inferences. Within a context, multiple inferences from different sources can provide duplicated or contradictory adaptations to a specific design dimension. To automatically adapt and optimize the interface's effectiveness, this framework proposes an *Adaptation* process to integrate these

inferences and identify the optimal non-conflicting set of adaptations for the interface.

Given the multitude of contextual components and their potential variations, it is infeasible to replicate all possible contexts, much less to determine the optimal adaptation that the iAR system must apply for each unique context. We define a ***scenario*** as a collection of distinct contexts with at least one common contextual component. For example, a walking scenario encompasses contexts such as walking in the street, walking on a treadmill, playing soccer, and any other contexts in which the user's physical state is walking. *Inferences* enables an iAR interface to generalize the known optimal adaptations for a specific context to other contexts within the same scenario. Additionally, *Inferences* enable an iAR interface to deduce information such as *User Task* and *User Objective* that are influenced by multiple contextual components and cannot be captured through a single contextual component. While such information can be included within context through *specified* contextual components, they are usually limited or inaccurate (e.g., in a task-switching scenario). For example, imagine Sanaz working from home and contacting the IT specialist through non-work-related social media to resolve an issue with the company's Slack channel. Without inferences, a specified *User Task* contextual component may indicate she is "working", leading her AR interface to close non-work related social media and interrupt her communication with the IT specialist. *Inferences* enable her iAR to correlate her *Location*, the *Disconnected Slack*, and *the company's IT specialist* to prevent disruptions to their communication.

5.4 Conclusion

To optimize its effectiveness, an iAR system must collect all relevant contextual information from the user that can influence the AR design dimensions identified in Section 4.2. In this chapter, we presented our approach to addressing *RQ6.* and *RQ7.*. Section 5.2 presented

our taxonomy of context that encompasses the explicit knowledge about the User, Setting, and Setting-User Interplay, obtained from contextual components that can affect the AR interface's content, presentation, or interaction methods. In Section 5.3, we proposed a framework detailing how our taxonomy can be integrated into iAR systems, enabling context-aware inference and adaptations. This chapter was presented at the IEEE VR 2022 Doctoral Consortium and is currently under review as part of our recent submission to TVCG.

- [26] Shakiba Davari "[DC] context-aware inference and adaptation in augmented reality" 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pages 938-939. IEEE, 2022.
  DOI:10.1109/VRW55335.2022.00320.

- [30] Shakiba Davari, Daniel Stover, Alexander Giovannelli, Cory Ilo, and Doug A. Bowman. "Towards Intelligent Augmented Reality (iAR): A Taxonomy of Context, An Architecture for iAR, and an Empirical Study." IEEE Transactions on Visualization and Computer Graphics, Aug 2024, 12 pages

Chapter 6

An Architecture for iAR & An Empirical Study

6.1 Overview

To address *RQ7.*, we propose an architecture for implementing iAR based on our proposed framework, design space (Chapter 4), and context taxonomy (Chapter 5). This architecture delineates our envisioned implementation of the Context, Inference, and Adaptation Units, including input (*RQ6.*), output (*RQ5.*), and the data flow and communication mechanisms within iAR systems (*RQ7.*). This serves as a validation of our proposed solutions to these research questions.

In Section 6.2, we provide a detailed description of these comprising modules and their internal data flow and external communications, as depicted in Figure 6.2. Next, we utilize this Architecture to explore the coverage and comprehensiveness of our proposed XR Design Dimensions, Taxonomy of Context, and Framework in the design of the context-aware interfaces proposed in Section 3.3 (Section 6.3).

In Section 6.4, we present an empirical AR experiment to observe user behavior and record user performance, context, and user-specified adaptations to the AR interfaces within a context-switching Real-World(RW) scenario. We explore the nuanced relationships between

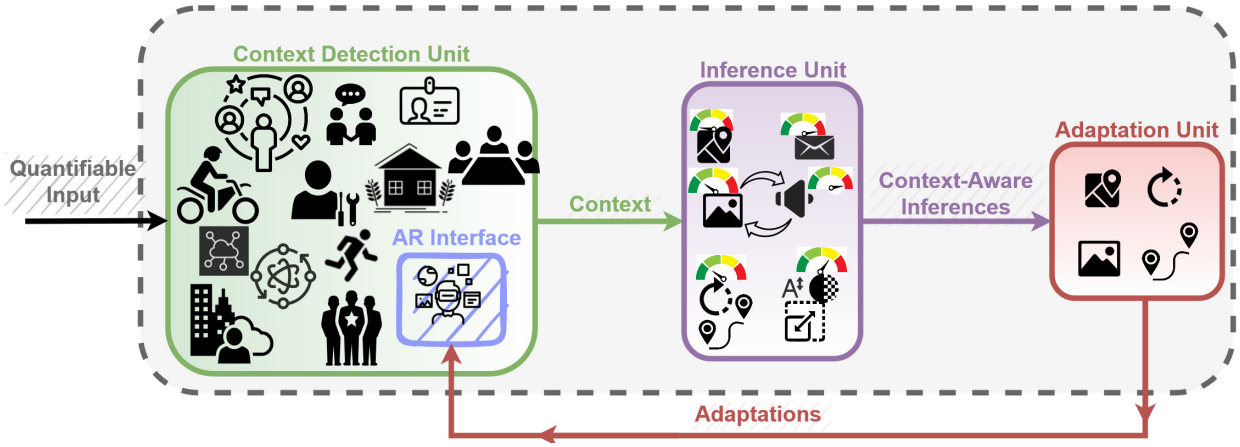


Figure 6.1: Our framework infers potential adjustments for interface performance using the context identified by our taxonomy. The Adaptation unit optimizes these inferences to adapt the interface.

context and user adaptations in this scenario and discuss the significance of our architecture in implementing iAR that identifies such patterns. This experiment emphasizes the significance of context-awareness in iAR and provides a preliminary training dataset for this specific Scenario (*RQ4.*).

6.2 Architecture

To illustrate the implementation of our solution to *RQ5.*, *RQ6.*, and *RQ7.* within iAR, we propose an architecture for these systems (Figure 6.2). We use examples of data structures for each data type used in the architecture. These examples are for demonstration purposes only and may vary based on the specific implementation of the architecture in the real world. For instance, while we use a dictionary as an example data structure for contexts, depending on the system implementation, other data structures, such as a graph or a tree, may be more appropriate.

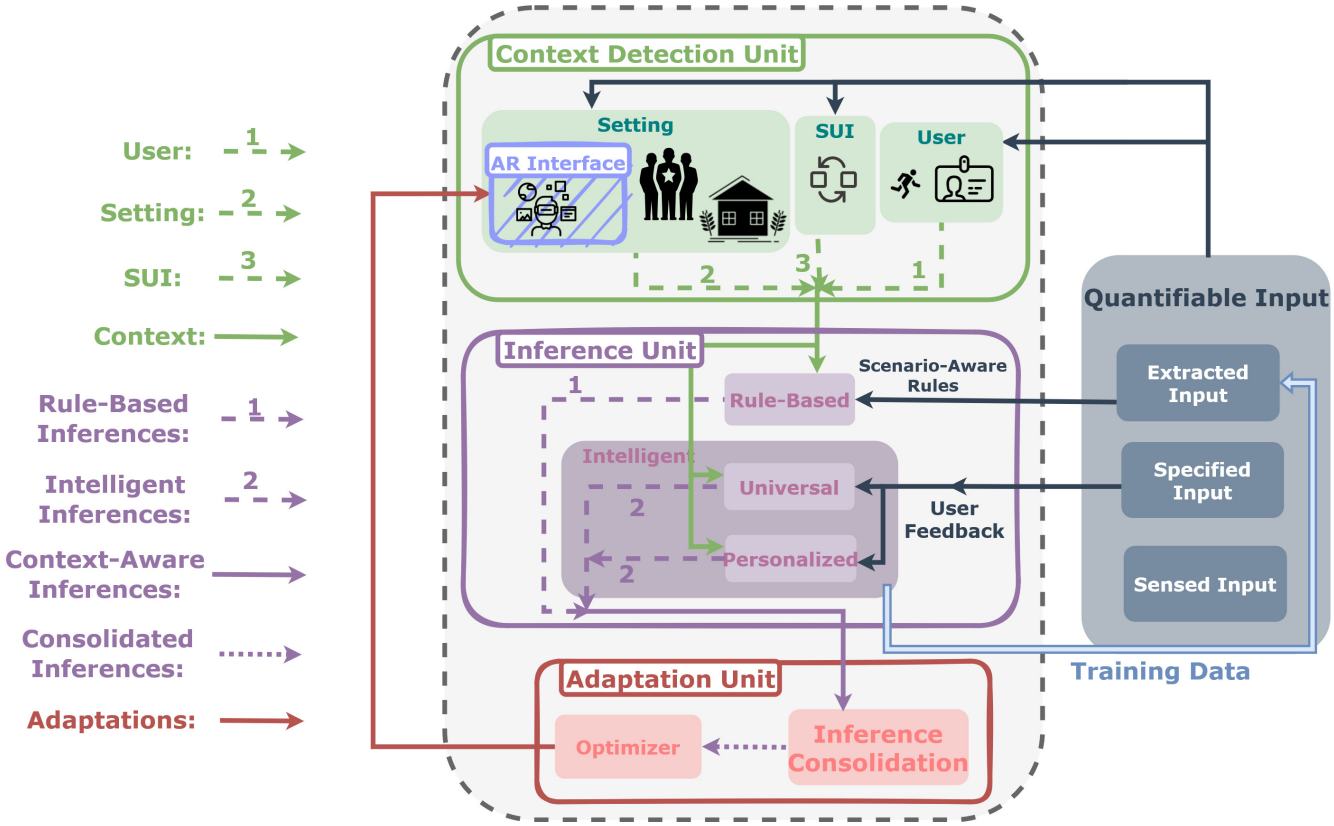


Figure 6.2: The architecture of intelligent AR systems.

6.2.1 Context Unit

The Context Unit identifies the value of contextual components relevant to the user, their setting, and their interplay with it (SUI) from the input data specified by the user or derived from sensors and internal and external databases. Here we represent context as a dictionary with a key for each contextual component ($\boldsymbol{CC}$). For example, the present context can be represented as:

$$\boldsymbol{Context_{present} = \{CC_1 : val_{CC1}, CC_2 : val_{CC2}, ...\}}$$

A scenario can be represented by a sub-dictionary of context, including only the common contextual components within that scenario.

Since the interface is part of the virtual setting, iAR directly affects the Context Unit

(Figure 6.2). We represent an adaptation by a heterogeneous dynamic list that includes three items: the design dimension that will be modified, the value assigned to this element by the adaptation, and a dynamic list of the entities to which the adaptation applies. For instance:

$$\boldsymbol{Adaptation_{bodyFixed} = ['Frame_of_Reference','body_fixed',}$$
$$[\boldsymbol{Map, NotificationMenu, ...}]]$$

6.2.2 Inference Unit

The Inference Unit extracts information regarding the potential impact of specific adaptations from the context and provides it to the Adaptation Unit. The iAR interface deduces inferences through Rule-Based and Intelligent Inference modules. Here we represent an inference by a heterogeneous list comprised of two items: a dynamic list of adaptations and an impact score indicating its effectiveness:

$$\boldsymbol{Inference_{bodyFixed} = [[Adaptation_{bodyFixed}, Adaptation_{transparent}], 0.85]}$$

Rule-based Inference Module

The rule-based module utilizes a database of predefined Design Principles to identify inferences. This module uses these rules and the present context to identify the present scenarios for which a rule exists and infers the rule-based inferences.

A scenario generalizes multiple contexts, making it easier to replicate. By considering a relatively small number of contexts within a scenario, patterns can be identified. These patterns can provide generalized Design Principles about the impact of specific adjustments

in that scenario. For instance, a Design Principle for a walking scenario, $\boldsymbol{Rule_{walking}}$, may indicate a highly favorable impact for adapting the virtual content's frame of reference to be body-fixed, as it creates a more natural and intuitive interaction. Since each unique context belongs to numerous scenarios, utilizing a collection of Design Principles congruent with the present context can enhance the effectiveness of adaptations.

A Design Principle can be represented with a heterogeneous list of two items: a scenario and a dynamic list of inferences for that scenario. A few examples are:

$$\boldsymbol{Rule_{walking} = [Scenario_{walking}, [Inference_{bodyFixed}]]}$$

$$\boldsymbol{Rule_{crossStreet} = [Scenario_{crossStreet}, [Inference_{multi_1}, Inference_{headFixed}]]}$$

Currently, the utilization of rule-based inference constitutes the primary method for the design of context-aware interfaces. As interface designers, we manually established rules (i.e., design principles/considerations/choices) that specify the impact of particular adaptations for specific scenarios. We refer to these design choices, design guidelines, rules, etc. as *Design Principles*. In this work, we distinguish iAR from context-aware AR. We consider **Context-aware AR** as the existing contemporary, proposed approaches where the Design Principles are the only source for providing inferences to the adaptation unit. For instance, consider two Design Principles for a scenario $\boldsymbol{Scenario_{Occlusion}}$ in which the user's view of a child they are monitoring is occluded by $\boldsymbol{App_{occluding}}$: $\boldsymbol{Rule_{repositionMechanism}}$ infers moving up the occluding entity, while $\boldsymbol{Rule_{OpacityMechanism}}$ infers making the occluding entity transparent [28].

$$\boldsymbol{Adaptation_{Move} = [position, position.y + 10, App_{occluding}]}$$

$$\boldsymbol{Inference_{Move} = [Adaptation_{Move}, 0.99]}$$

$$\boldsymbol{Rule_{repositionMechanism} = [Scenario_{Occlusion}, [Inference_{Move}]]}$$

$$\boldsymbol{Adaptation_{Transparent} = [color, color.\alpha = 0.0, App_{occluding}]}$$

$$\boldsymbol{Inference_{Transparent} = [Adaptation_{Transparent}, 0.99]}$$
$$\boldsymbol{Rule_{OpacityMechanism} = [Scenario_{Occlusion}, [Inference_{Transparent}]]}$$

Rule-based inferences offer iAR the potential to automatically adapt even when a limited number of contextual components can be identified and extracted. This approach also allows iXR to utilize rules from the currently available interface design knowledge for specific scenarios. Additionally, the rule-based approach offers the potential for adaptability to a wide range of contexts through a small set of rules. However, despite its capabilities, rule-based inference is not without its limitations and challenges.

One key limitation of this approach is the manual process of extracting and designing these Design Principles, which can be prone to human error, demanding, time-consuming, and limited to the existing knowledge. On the other hand, the over-generalization of the adaptations through a limited number of rules can negatively impact the performance. For example, a rule mandating the Opacity of virtual content in a conversation potentially enhances the user experience by reducing distraction when the conversation primarily aims to establish social connections and the virtual content presents supplementary information. However, if the primary objective of the conversation is to convey information from the virtual content, this rule may degrade the interface's performance. As a result, an interface that lacks the appropriate rules for the latter scenario may fail in such contexts.

Furthermore, the effectiveness of this approach is highly dependent on the reliability and accuracy of the Design Principles. For instance, incorporating a rule pertaining to the scenario of $[\boldsymbol{Scenario_{Y2023} = CC_{year} : 2023}]$ would impact all conceivable scenarios within the year 2023 and may prove to be detrimental if not appropriately considered. Another key aspect of iAR is its ability to tailor the experience to the individual user. However, it is difficult for the interface designer to create Design Principles that take into

account the unique preferences, values, and limitations of a user who deviates from the typical, average user [45]. While limited user-specified rules can be introduced, the rule-based inference approach is constrained to personal experiences.

Intelligent Inference Module

Intelligent Inference can enhance iAR's capabilities by leveraging Machine Learning (ML) techniques to learn from the user's behavior and automatically extract inferences without any prior knowledge or rules. This approach allows the iAR system to go beyond the limitations of traditional context-aware AR interfaces. By observing the user's manual adaptations in different contexts, ML models can be trained to predict the impact of specific adaptations in those contexts and their comprising scenarios.

This architecture envisions the implementation of distinct ML models for each AR design dimension.These models generate Universal and Personalized Inferences derived directly from the context. *Universal Inferences* are general inferences that can be applied to a wide range of users, while *Personalized Inferences* are tailored to the specific user's behavior.

The Intelligent inferences generated by this unit are a collection of all these Universal and Personalized Inferences, providing a more comprehensive and accurate set of inferences.This approach enables iAR to adapt to the user's needs and preferences in real-time, resulting in a more personalized and seamless user experience.

Universal Inferences Module

Universal Inferences are generated through static ML models, which are trained using batch learning on a large dataset collected from a diverse user population. Utilizing a large user population enables training models on a greater number of data points, resulting in a higher level of accuracy. These models are trained on a fixed dataset and do not update their

parameters after the training process, allowing them to predict universal inferences with a relatively low computational cost. However, it should be noted that the generation of universal inferences relies heavily on the prior collection and availability of data from the implementation of AR technology. This poses several challenges, including the necessity of widespread adoption of AR technology, addressing user privacy concerns, establishing a unified infrastructure for data collection, and ensuring the availability of adequate resources for data storage and transfer. Further research is necessary to explore solutions to these challenges. One potential solution is the use of Generative Adversarial Networks (GANs) to generate training data using the available Design Principles.

Personalized Inference Module

Personalized Inferences are generated through online machine-learning models that continuously process the user's feedback on each adaptation in real-time. Online learning is particularly useful in this case as it not only addresses privacy concerns but also eliminates the need for a unified infrastructure for data collection and storage, thus reducing the technological requirements for data storage and transfer. As new adaptations are made, these models update their parameters and adapt to the user's preferences. The data used for training these models includes both the user's manual interventions and the interface's automatic adaptations in each context. This feedback provides the online models with information on which adaptations the user accepted or rejected, enabling them to adjust their internal parameters to better suit the user's needs. For instance, receiving the user's manual intervention on the interface, $\boldsymbol{Adaptation_{manual}}$, after an automatic adaptation, $\boldsymbol{Adaptation_{automatic}}$, can serve as feedback for the machine learning model to adjust its internal parameters. Specifically, this feedback can reinforce a negative assessment of the effectiveness of $\boldsymbol{Adaptation_{manual}}$, while concurrently reinforcing a positive assessment of the effectiveness of the $\boldsymbol{Adaptation_{automatic}}$ in that particular context. This type of

real-time learning allows the model to adapt to changing data distributions and provide personalized inferences tailored to the specific user. The use of transfer learning to transfer the knowledge acquired by these online learners from different users to the batch learners in the Universal Inference Module provides another potential solution to the challenges of training those static models.

6.2.3 Adaptation

The adaptation unit utilizes the *inferences* to optimize the AR interface design dimensions [27] through two primary operational modules: Inference Consolidation and Optimization.

Inference Consolidation Module

The Inference Consolidation module is responsible for determining the actual impact of each proposed adaptation. In each context, multiple rule-based and intelligent inferences may provide duplicate or contradictory information about the potential impact of specific adaptations. The inference consolidation module integrates various inferences involving a given adaptation and synthesizes a consensus about its potential impact. This module then provides the Optimization module with a single, consolidated inference for each potential adaptation.

Optimization Module

The optimization module combines the inferences generated by the inference consolidation module and optimizes these adjustments to select a set of adaptations that enhance the overall effectiveness of the interface.

As an example, let us consider utilizing the social context-aware interface from [29] in a conversation in which the user is being asked about the weather app, currently placed in front of the interlocutor's face. The inference consolidation module assesses the potential impact of making the weather app fully opaque. In this occlusion scenario, the Design Principle, *Support for viewing social cues*, infers a fully opaque weather app would have a negative effect on the user's experience. However, the same rule also infers that simultaneously adapting the weather app's opacity to fully opaque and moving it above the user's face would reverse this negative effect. On the other hand, the weather app provides socially relevant information. Thus, the inference rule, Support for socially relevant information access, infers that the weather app's fully opaque opacity would have a relatively significant positive impact. The inference consolidation module integrates these inferences and subsequently presents the optimization module with a single comprehensive inference, which suggests a significant positive impact on the simultaneous adjustment of the weather application's opacity to fully opaque and its placement above the user's face. In this example, the Adaptation Unit also receives an inference suggesting a small positive impact on adapting the opacity to be fully transparent for all content, including the weather app. This inference, inferred from the Design Principle, **RW prioritization**, in this social scenario, will be directly provided to the optimization module without consolidation since it is the only inference on adapting the opacity to fully transparent. Finally, the optimization module will enhance the overall performance of the interface by rendering the weather app fully opaque and manipulating its positioning.

6.3 Reproducing the Design of Occlusion-Aware AR Interfaces for Real-World Monitoring Task

In this section, we aim to demonstrate the crucial elements utilized for context detection, adaptation, and implementation of the interfaces presented in Section 3.3 through our architecture. We reproduce the design principles and interfaces introduced in this work using our contextual components and AR design elements and identify the Design Principles that enable our framework to reproduce and expand upon these interfaces. The context-aware AR interfaces in this study can be designed using contextual components and adaptations to AR interface design elements from our taxonomy.

$\boldsymbol{CC_1}$ is the position of the RW item, categorized as the immediate environment in our taxonomy. In the scenario of this study, the position of the child is specifically used.

$\boldsymbol{CC_2}$ is the user priorities, categorized as SUI contextual components. In this scenario, the interplay between the user's occupation (babysitter) and the RW setting (child's home) identifies the priority of the RW and the child.

$\boldsymbol{CC_3}$ is a specific virtual entity (app), categorized under digital setting in our taxonomy. This is the occluding object in this user study.

$\boldsymbol{A_1}$ adapts the position of a virtual entity. Consider $\boldsymbol{A1^{up}_{all}}$ to represent adjusting the $\boldsymbol{position}$ design element of $\boldsymbol{all}$ of the virtual entities to the value $\boldsymbol{up}$:

$$\boldsymbol{A1^{up}_{all} = [position, up, all - entities]}$$

$$\boldsymbol{A1^{down}_{occluding} = [position, down, occluding - entity]}$$

$\boldsymbol{A_2}$ adapts a virtual entity's visibility using transparency, and $\boldsymbol{A2^{transparent}_{manual}}$ represent adjusting the $\boldsymbol{transparency}$ design element of the virtual entity that was selected manually,

$\boldsymbol{manually-selected-entity}$, to the value $\boldsymbol{transparent}$:

$$\boldsymbol{A2^{transparent}_{manual} = [transparency, transparent, manually-selected-entity]}$$

$$\boldsymbol{A2^{opaque}_{occluding} = [transparency, opaque, occluding-entity]}$$

$\boldsymbol{A_3}$ adapts the dynamic visuals of a virtual entity using a $\boldsymbol{blinkingredoutline}$.

$$\boldsymbol{A3^{on}_{occluding} = [dynamic-visuals, blinking-red-outline, occluding-entity]}$$

$$\boldsymbol{A3^{off}_{occluding} = [dynamic-visuals, none, occluding-entity]}$$

Based on $\boldsymbol{CC_2}$, all the contexts within this user study are within the $\boldsymbol{Scenario_{RW_Monitoring}}$. $\boldsymbol{CC_1}$ and $\boldsymbol{CC_3}$ can identify whether the context is within the $\boldsymbol{Scenario_{Occlusion}}$ or not. This work introduced eight interfaces, including six context-aware interfaces. We identify the inferences and Design Principles introduced in this study to describe the implementation of these interfaces based on our framework. The interfaces were designed using three design principles.

Adaptation Mechanism enables users to access the RW information and virtual content using $\boldsymbol{A_1}$ or $\boldsymbol{A_2}$.

Content Prioritization enables prioritizing real-world content over the virtual and vice versa using $\boldsymbol{A_2}$.

Level of Automation utilizes $\boldsymbol{CC_1}$ and $\boldsymbol{CC_3}$ to identify occluding objects and designs three types of AR interfaces. *Fully Automatic* interfaces are context-aware AR interfaces that apply their adaptation mechanism to resolve the occlusion. *Automatic Detection* interfaces are context-aware AR interfaces that use $\boldsymbol{A_3}$ to draw the user's attention and awareness to the occluding object and allow them to control whether to resolve the occlusion or maintain access to the virtual content. *Fully Manual* interfaces are non-adaptive AR interfaces.

Utilizing these design principles within our architecture can result in the following inferences:

$\boldsymbol{Inference_{Reposition_RW-Prioritized} = [A1^{up}_{all}, 1.0]}$

$\boldsymbol{Inference_{Transparency_RW-Prioritized} = [A2^{transparent}_{all}, 1.0]}$

$\boldsymbol{Inference_{Reposition} = [A1^{up}_{occluding}, 1.0]}$

$\boldsymbol{Inference_{Transparency} = [A2^{transparent}_{occluding}, 1.0]}$

$\boldsymbol{Inference_{Notification} = [A3^{on}_{occluding}, 1.0]}$

Each proposed interface in this study can be reproduced using one of the following Design Principles.

Automatic-Reposition Interface applies $\boldsymbol{A_1}$ to automatically move any occluding virtual entity when the context is within the $\boldsymbol{Scenario_{Occlusion}}$.

$\boldsymbol{Rule_{Automatic_Reposition} = [Scenario_{Occlusion}, [Inference_{Reposition}]]}$

Automatic-Transparency Interface applies $\boldsymbol{A_2}$ to automatically make any occluding virtual entity transparent when the context is within the $\boldsymbol{Scenario_{Occlusion}}$.

$\boldsymbol{Rule_{Automatic_Transparency} = [Scenario_{Occlusion}, [Inference_{Transparency}]]}$

Detection-Reposition Interface and **Detection-Transparency Interface** apply $\boldsymbol{A_3}$ to automatically blink any occluding virtual entity when the context is within the $\boldsymbol{Scenario_{Occlusion}}$.

$\boldsymbol{Rule_{Detection} = [Scenario_{Occlusion}, [Inference_{Notification}]]}$

RW-Prioritized Reposition Interface applies $\boldsymbol{A_1}$ to automatically move all virtual entities when the context is within the $\boldsymbol{Scenario_{RW_Monitoring}}$.

$\boldsymbol{Rule_{RW-Prioritized_Reposition} = [Scenario_{RW_Monitoring}, [Inference_{Reposition_RW-Prioritized}]]}$

RW-Prioritized Transparency Interface applies $\boldsymbol{A_2}$ to automatically make all virtual entities transparent when the context is within the $\boldsymbol{Scenario_{RW_Monitoring}}$.

$\boldsymbol{Rule_{RW-Prioritized_Transparency} = [Scenario_{RW_Monitoring}, [Inference_{Transparency_RW-Prioritiz}}$

Using these inferences and Design Principles, our framework can be applied to reproduce each of the six context-aware interfaces designed in this study. Our architecture enables iAR to detect the context using $\boldsymbol{CC_1}$, $\boldsymbol{CC_2}$, and $\boldsymbol{CC_3}$, providing the rule-based inferences for the present scenarios to the adaptation unit for optimization.

6.4 Experiment 4: User-Specified AR Adaptations in Eight Contexts

Designing an effective iAR interface necessitates understanding the intricate relationship between the context and users' preferences regarding various design dimensions of the interface. This requires training data involving user adaptations across diverse contexts. We conducted an experiment to observe and record context, user performance, and user-specified adaptations to an AR interface within a context-switching scenario. This experiment aims to address one research question:

$\boldsymbol{RQ_{Experiment4}}$: How do users wish to adapt their AR interface in a context-switching scenario?

To address this $\boldsymbol{RQ_{Experiment4}}$, we conducted an empirical study involving a dynamic scenario with various context switches in a library. We observed each participant's behavior and interactions and collected data about their context and their adaptations to the AR interface.

6.4.1 Experimental Design

In this experiment, participants utilized an AR interface equipped with five apps necessary for their task completion. Each participant encountered multiple contexts, varying in the

immediate RW environment and the user's state and objective. They were instructed to manually adapt the availability, spatial layout, and appearance of each app at any time according to their preferences.

AR Interface

The AR interface provided participants with five apps to complete their tasks: a *Library Map app*, displaying a picture of the library's floor plan and the call number ranges on each bookshelf, helping participants find shelves with specific call numbers; two *Book app*s, each presenting a picture displaying a book cover along with its corresponding library call number; a *Messaging app*, displaying simulated text messages, including a message from a friend, inviting the user to join them at a specific study area, along with an image of the area; and a *Stock app*, allowing participants to trade simulated stocks and providing visual and auditory alerts for favorable trading opportunities.

Contexts

The session involved a context-switching scenario in which changes occurred to three contextual components: *RW Environment*, *Mobility*, or *RW Objective*.

RW Objective

The participants were instructed to sequentially attend four objectives in the RW. Each participant started by searching for two books indicated within the *Book apps* in the library (*Locate Book1*, *Locate Book2*). Once both of these objectives were achieved, they received a text message on the AR device. They were instructed to use an image from the *Messaging app* to locate their friend in a specific study area (*Locate Friend*). Finally, upon reaching

the study area, they were asked to read a page from one of the books(*Read*).

Mobility

The participant's physical activity changed throughout the session. At times, they were *Mobile*, walking within the environment, while at other times, they were *Stationary*, either reading or searching through a shelf of books.

Environment

When searching for a book, at times, the spatial characteristics of the participant's immediate environment remained *Confined* within the surrounding shelves. However, at other times their immediate environment was *Unconfined*, e.g., as they were walking between open areas and bookshelves.

During each session, the participants encountered eight unique contexts:

$\boldsymbol{C_1}$: In the first context that participants encountered, their *RW Objective* involved *Locating Book1*. In this context, to navigate within the library and locate the correct isle of bookshelves, the participant was *Mobile*, searching within a *Unconfined Environment*.

$\boldsymbol{C_2}$: In the second context, to *Locate Book1*, the participant was navigating the correct shelf within an aisle of shelves, being *Mobile* in a *Confined Environment*.

$\boldsymbol{C_3}$: In the third context, to *Locate Book1*, the participant was *Stationary* in a *Confined Environment*, searching for a book within a shelf.

With the difference in the *RW Objective*, when *Locating Book2*, the participant encountered $\boldsymbol{C_4}$, $\boldsymbol{C_5}$, $\boldsymbol{C_6}$ which were identical to $\boldsymbol{C_1}$, $\boldsymbol{C_2}$, $\boldsymbol{C_3}$ respectively, in terms of the participant *Mobility* and *Environment*.

$\boldsymbol{C_7}$: After locating both books, in this context, to *Locate their Friend*, the participants were *Mobile*, navigate within the *Unconfined Environment* of the library.

$\boldsymbol{C_8}$: In the final context they encountered, the participant sat on a chair and *Read a Book*, while *Stationary* within a *Unconfined Environment*.

Throughout the session, within all these contexts, the participants were directed to utilize the *Stock app* as a secondary objective to maximize their revenue (*Virtual Task*). We incorporate this *Virtual Task* to investigate whether there is a consistent pattern across all participants' adaptations to the *Stock app* as the AR content utilized for a solely Virtual Task.

AR Adaptations

To investigate the interaction between these contexts and the visual presentation of the AR interface, the AR interface provided various functionalities to adapt various design dimensions of each AR app at any time [27].

Visibility: Each app featured a ***Minimize*** button, allowing participants to remove apps from their environment or reopen them at their previous state from an app menu according to their preference.

Spatial Layout: Three buttons labeled ***head-fixed***, ***body-fixed***, and ***world-fixed*** within each app permitted participants to switch between these three fixed FoRs at any time. Participants had the ability to drag each app to their desired ***position*** within its FoR and adjust its ***size*** using a bimanual pinch-scale gesture.

Appearance: Participants used a percentage slider to adjust the ***Opacity*** of each app whenever desired.

6.4.2 Apparatus

During the experiment, participants utilized a Microsoft HoloLens (2nd gen) Augmented Reality Head-Worn Display (AR HWD). This wireless device features a holographic density of 47 pixels per degree and a 52-degree diagonal field of view. Our AR interface was developed using Unity and Microsoft's Mixed Reality Toolkit 2 (MRTK2).

We utilized pictures, as detailed in Section 6.4.1, for both our *Library Map app* and the *Book apps*. To create the *Stock app*, we developed a custom application that featured the participant's total "profit" along with two stock prices, each accompanied by a red "buy" button and a green "sell" button. Participants aimed to increase their profit by purchasing stocks displayed in red text, denoting a price decrease, and selling those in green text, signaling a price increase. Price changes were announced by a notification sound, occurring randomly every 40 to 60 seconds. Consistency across participants was ensured via predefined intervals stored in a text file. Upon locating *Book 2*, our interface issued both auditory and visual notifications to indicate a message from a "friend". Subsequently, the *Messages app*, initially showing an empty chat window, revealed the received text and image.

App adaptations were exclusively enabled within the "Edit Mode," accessible via a hand-attached edit button. Upon activation, users gained access to a hand-attached app menu, as well as UI buttons, sliders, and drag and scale gestures on individual apps, enabling app adaptations as outlined in Section 6.4.1.

Implementing the *Body-fixed* FoR in current AR devices poses challenges in accurately detecting the user's body orientation.In our approach, we inferred the body's forward direction based on the movements of the headset. We identified changes in body direction by comparing the current movement vector at each frame($\vec{\boldsymbol{x}}_{\boldsymbol{curr}}$) with the previous normalized movement vector ($\vec{\boldsymbol{x}}_{\boldsymbol{prev}}$). To mitigate the influence of head jitter and minor movements,

we applied a threshold ($\vec{x}_{prev} \cdot \vec{x}_{curr} < \mathbf{0.85}$) and focused on linear movement along the floor plane, disregarding vertical Y-axis components. This method addressed the detection of the user's body orientation in mobile contexts. To handle body orientation changes while the user is stationary, we introduced a "center" voice command. This command utilized the head orientation at the time of activation as the reference for the body's orientation in stationary contexts.

6.4.3 Procedure

After receiving approval from our university's Institutional Review Board, we recruited participants from our local university. The experiment lasted 60 minutes.

Upon arrival, participants were requested to carefully review and sign the consent form. Following this, we collected demographic information and their prior experience with AR. After this pre-study questionnaire, we provided a brief overview of the study and introduced the device, its operation, and its interaction methods.

Prior to the main session, participants conducted eye calibration on the HoloLens device and underwent a training session to become acquainted with the experimental tasks and the provided AR apps and adaptations. They then implemented their preferred adaptations to the AR apps and proceeded to the main session to carry out the study tasks. All participants experienced the eight contexts in the same sequence and were encouraged to adapt whenever necessary. Moreover, they were prompted to voice their thoughts freely (think-aloud protocol), articulating reasons for making adaptations, providing specifics about each adaptation, and identifying any desired adaptations that were unavailable. At the conclusion of the study, all participants engaged in an interview regarding their overall experiences during the session.

6.4.4 Participants

We recruited 23 participants from our local university community. This population was between 20 and 25 years old ($\boldsymbol{M = 21.35}$, $\boldsymbol{SD = 1.43}$), with 21 being male and two female. Among them, 16 participants had little to no experience with AR, while only three had used AR more than ten times.

6.4.5 Data Collection

We collected information about the user's context, time-stamped interface adaptations, and task completion time and accuracy throughout the session. As our contextual information, we collected CSV files logging spatial and velocity data from the user and their gaze during system use, as well as depth images and RGB videos from the forward-facing cameras on the HWD. The depth images had a resolution of 512x512 pixels and were recorded from the device's depth sensor at 1 frame per second (FPS). The videos had a resolution of 896x504 pixels and were recorded at 15 FPS from the device's external RGB camera. These videos included audio and digital holograms rendered in the headset's display.

Moreover, we collected each participant's time-stamped adaptations to each app's FoR (head, body, world), position (x, y, z), rotation (x, y, z), scale (x, y, z), Opacity percentage, and visibility in CSV files. We also recorded participant performance, aiming to provide this dataset as preliminary training data for future machine-learning models tailored to this specific scenario. To record each participant's performance, we also collected CSV files logging book discovery and study area navigation speed and user's speed and accuracy in responding to stock app recommendations. Metrics such as task completion time and accuracy can be utilized for the training of multi-objective optimization models.

6.4.6 Results

Our participants adjusted the design dimensions of 5 apps within 8 contexts. We considered each participant's final design of each app used within each context as a data point. Three participants' data were excluded from the analysis as they were disrupted and not collected properly for the entire session. Consequently, our data involved 800 user-specific design choices and adaptations (8 * 5 * 20). This section analyzes these logged data and investigates patterns within user design choices across different contexts.

The *Visibility* design dimension is represented by categorical data alternating between *Minimized* and *Visible* values. In **660** of the logs (82.5%), the apps were *Visible*. These apps were used to explore design choices regarding *FoR*, *Position*, *Scale*, and *Opacity*.

Since *Position* inherently relies on the *FoR*, we explored this design dimension within each *FoR* separately. From the **660** *Visible* design choices, **90%** used the *Body-fixed FoR*, while the *World-fixed FoR* was used only in **6** instances($\sim$ **1%**). Consequently, to remove any biases caused by a large number of *Body-fixed* apps, we analyzed patterns of *Scale* and *Opacity* separately for the *Body-fixed* and *Head-fixed FoRs*. Additionally, we removed the *World-fixed FoR* from our analysis due to insufficient data.

In these analyses, the *Opacity* values are continuous measurements from **0.15** (Nearly invisible) to **1.0** (fully Opaque), and the *Scale* values are continuous numbers from **0.25** to **1.5** indicating the relative size of each app compared to a baseline app size. Within the *Head-fixed FoR*, we binned the $\boldsymbol{(x, y, z)}$ app positions from the study into nine sectors, subdividing the HoloLens' FoV (**52°** diagonal) into three horizontal and three vertical categories. Within the *body-fixed FoR*, 11 sectors were considered, requiring varying degrees of head-turn and physical effort to be accessed. Once the user confirms their forward-facing direction as the *Center*, we subdivide the area covering the front of the user and their periphery ($<$ **45°**

head turns) into three horizontal and three vertical categories, resulting in nine sectors. The *Far Right* and *Far Left* position sectors cover the apps positioned to the right/left side of the user's body where more extreme head-turns ($> \mathbf{45°}$) are required to access the app. The study did not observe apps in other areas, such as directly behind or above the user.

We report the Likelihood Ratio for contingency analysis of our categorical design dimensions and explore correlations of *Scale* and *Opacity* with other categorical factors through non-parametric Kruskal-Wallis tests. For pairwise comparisons of *Scale* and *Opacity* within non-binary factors, non-parametric Wilcoxon was used.

Design Dimension Correlations

We explored various design dimensions of apps regardless of the specific context or app. Within our observations, only $\sim \mathbf{18\%}$ of the apps were *Minimized*, and $\sim \mathbf{90\%}$ of the apps were *Body-fixed*. Within both $\boldsymbol{FoRs}$, nearly $\mathbf{85\%}$ of the apps used $\geq \mathbf{0.9}$ opacity levels.

In the following, we present the results of our statistical analyses of these patterns and correlations.

Kruskal-Wallis tests revealed *FoR*'s significant influence on *Scale* ($\boldsymbol{\chi^2_{(647,1)} = 32.83,\ p < 0.0001}$), showing that the *Head-fixed* apps were significantly smaller ($\boldsymbol{M = 0.57,\ STD = 0.18}$) than the *Body-fixed* ones ($\boldsymbol{M = 0.8,\ STD = 0.32,\ Z = -5.73,\ p < 0.0001}$).

Figure 6.3 and Figure 6.4 depict the Highest Density Region Contours (HDS) of Scale vs. Density in each sector based on the user-specified design within the *Body-fixed* and *Head-fixed* FoRs. We analyzed the correlations between *Position*, *Opacity*, and *Scale* within each of these *FoRs*. The results indicated that Opacity correlated to position in both *FoRs*, and to Scale in *Head-fixed FoR*.

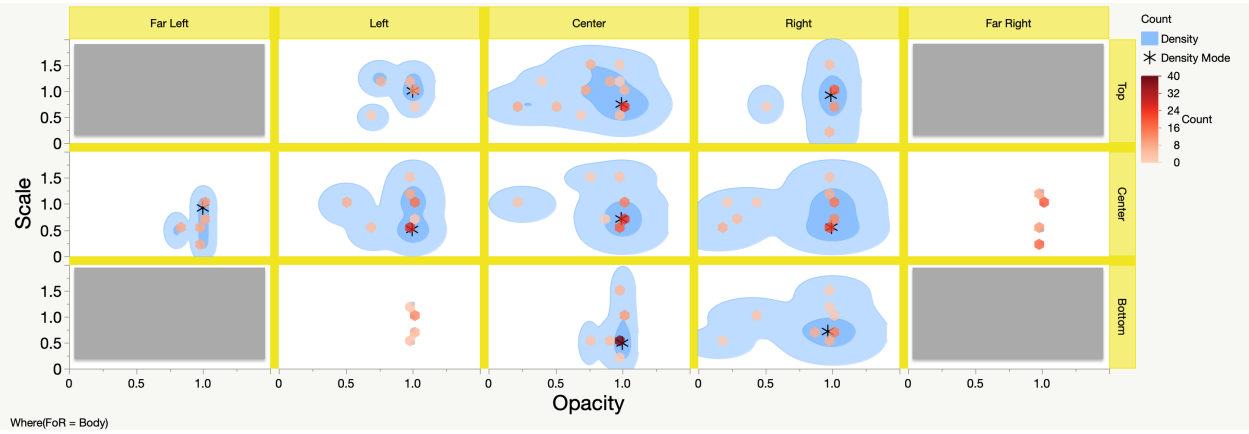


Figure 6.3: Highest Density Region Contours (HDS) of Scale vs. Opacity in each sector within the Body-fixed FoR

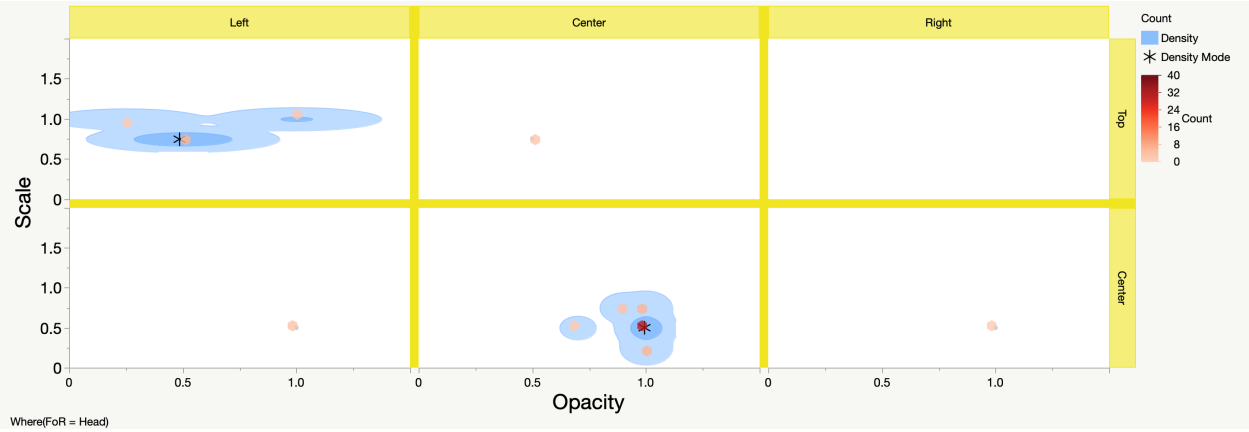


Figure 6.4: HDS of Scale vs. Opacity in each sector within the Head-fixed FoR

Body-fixed:

$\sim$ **85%** of the *Body-fixed FoR* apps were Opaque ($\boldsymbol{Opacity} \geq \mathbf{0.8}$) and $\sim$ **42%** were using larger scales $\geq \mathbf{1}$. In this *FoR*, *Opacity* was significantly influenced by *Sector* ($\boldsymbol{\chi^2_{(592,10)} = 87.4,\ p < 0.0001}$), *Horizontal Placement* ($\boldsymbol{\chi^2_{(592,4)} = 17.63,\ p \sim 0.0015}$), and *Vertical Placement* ($\boldsymbol{\chi^2_{(592,2)} = 24.64,\ p < 0.0001}$).

All of the apps placed on the *Far Right* sector and the *Bottom Left* sector were fully Opaque ($\boldsymbol{M = 1.0,\ STD = 0.0}$). The apps placed in the sectors *above eye-level* had *Lower Opacity* ($\boldsymbol{M = 0.86,\ STD = 0.22}$), compared to those that were placed *at* ($\boldsymbol{M = 0.94,\ STD =}$

$\mathbf{0.18,\ Z = -4.55,\ p < 0.0001}$) or *below* ($\mathbf{M = 0.96,\ STD = 0.12,\ Z = -3.66,\ p \sim 0.0003}$) eye-level.

Scale was also significantly correlated to *Sector* ($\boldsymbol{\chi^2_{(592,10)} = 78.94,\ p < 0.0001}$), *Horizontal Placement* ($\boldsymbol{\chi^2_{(592,4)} = 27.81,\ p < 0.0001}$), and *Vertical Placement* ($\boldsymbol{\chi^2_{(592,2)} = 42.88,\ p < 0.0001}$).

Compared to all other vertical categories, the body-fixed apps placed in the sectors to the *Left* of the user's periphery were the *Largest* ($\mathbf{M = 0.89,\ STD = 0.3}$), and those to the *Far Left* were *Smallest* ($\mathbf{M = 0.62,\ STD = 0.27}$). The apps *above eye-level* were also *Larger* ($\mathbf{M = 0.94,\ STD = 0.3}$) than those placed at eye-level ($\mathbf{M = 0.76,\ STD = 0.32,\ Z = 5.76,\ p < 0.0001}$) or below it ($\mathbf{M = 0.74,\ STD = 0.29,\ Z = 5.92,\ p < 0.0001}$).

Within the *Body-fixed Center Sector*, the *Opacity* and *Scale* were shown to be inversely correlated by a polynomial simple linear regression ($\mathbf{F(65,1) = 4.03,\ p \sim 0.049}$).

Head-fixed:

The *Size* of the Head-fixed apps was inversely correlated to their *Opacity*. $\mathbf{\sim 95\%}$ of these apps were either placed in the *Center* sector using high Opacity and scaled-down sizes ($\mathbf{M_{Scale} = 0.52,\ STD_{Scale} = 0.14,\ Z = 4.61,\ p < 0.0001}$) or had lower Opacity and relatively larger size and were placed at the *Top Left* sector ($\mathbf{M_{Scale} = 0.83,\ STD_{Scale} = 0.13,\ M_{Opacity} = 0.58,\ STD_{Opacity} = 0.25}$). $\mathbf{\sim 70\%}$ of these apps had $\boldsymbol{Scale \leq 0.5}$, and $\boldsymbol{Scale > 1}$ was never applied to any apps.

Our statitictical analyses reveal that *Opacity* was significantly influenced by *Sector* ($\boldsymbol{\chi^2_{(62,4)} = 34.8,\ p < 0.0001}$), *Horizontal Placement* ($\boldsymbol{\chi^2_{(62,2)} = 15,\ p \sim 0.0006}$) , and *Vertical Placement* ($\boldsymbol{\chi^2_{(62,1)} = 34.24,\ p < 0.0001}$). With one app placed at the Right side of the user's FoV (specifically at the *Right Center sector*), two placed in the *Top Center*

Sector, and three in the *Left Center Sector*, all the other apps were placed either in the *Center Sector* (**75%**) or the *Top Left* one (**15%**).

Similarly, *Scale* correlated to *Sector* ($\boldsymbol{\chi^2_{(62,4)} = 26.08,\ p < 0.0001}$), *Horizontal Placement* ($\boldsymbol{\chi^2_{(62,2)} = 13.85,\ p \sim 0.001}$), and *Vertical Placement* ($\boldsymbol{\chi^2_{(62,1)} = 25.93,\ p < 0.0001}$).

Polynomial Simple Linear Regression also found a linear fit between *Scale* and *Opacity*, indicating an inverse correlation berween the two within Head-fixed ($\boldsymbol{F(60,1) = 23.18,\ p < 0.0001}$).

Contextual component Correlations to Design Dimensions

The study scenario involved three contextual components and eight unique combinations of these components (i.e., Contexts). Our results indicate that Context influenced when users *Minimized* apps, and how they positioned their *head-fixed* apps. The following delineates these analyses of the influence of each contextual component or context on design dimensions.

As shown in Figure 6.5, *Visibility* correlates to *Context*. Our user-specified design choices indicated that the users *Minimize* more apps, when **a.** in *Unconfined Environments*(**23%** of the apps compared to **14%** of the apps within the contexts involving *Confined Environments*), **b.** they were *Stationary* (**21%** of the apps compared to **15%** of the apps within the contexts in which they were *Mobile*), and **c.** their objective in the RW involved *Reading*(**39%** of the apps) rather than *Locating Book1*(**9%** of the apps), *Book2*(**18%** of the apps), or their *Friend*(**27%** of the apps). Likelihood Ratio tests indicated the significance of correlations between *Visibility* and **a.** *Environment* ($\boldsymbol{\chi^2_{(800,1)} = 12.8,\ p \sim 0.0003}$), **b.** *Mobility* ($\boldsymbol{\chi^2_{(800,1)} = 4.14,\ p \sim 0.04}$), **c.** *RW Objective* ($\boldsymbol{\chi^2_{(800,3)} = 48.5,\ p < 0.0001}$), and **d.** *Context*($\boldsymbol{\chi^2_{(800,7)} = 48.5,\ p < 0.0001}$)

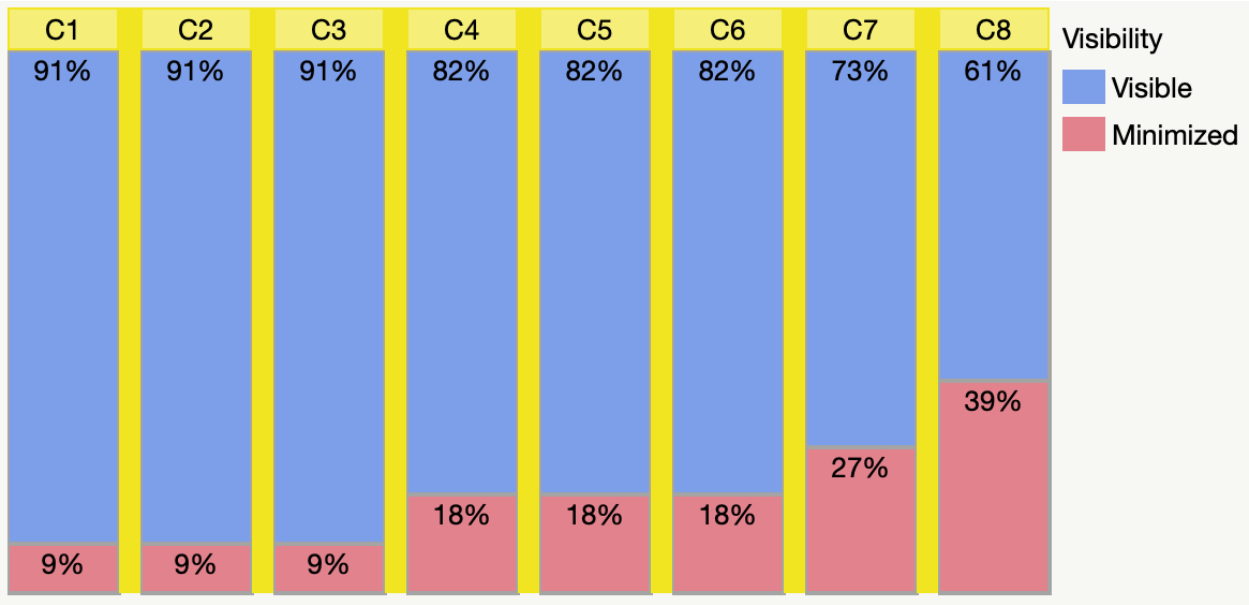


Figure 6.5: User-specified app Visibility within various Contexts.

Our observations did not correlate context and contextual components to the choice of *FoR* or *Position*, *Opacity*, and *Scale* within the **Body-fixed** and **Head-fixed** *FoRs*. In our analyses of the *Head-fixed* design dimensions, we disregard the correlations of *Context* and *RW Objective*. This is because these results are misleading considering the small number of head-fixed samples, 62 instances, leading to only 6 data points for Find Friend ($\boldsymbol{C_7}$) and 4 for Read ($\boldsymbol{C_8}$).

App Role Correlations to Design Dimensions

The analyses presented in Section 6.4.6 and Section 6.4.6 do not consider the influence of individual apps and their content. Depending on the user's task, the relevance of a particular app, and therefore the chosen design for that app, may vary. Our findings indicate that the Stock app was never *Minimized*, and makes up **40%** of the uses *Head-fixed* FoR. Within the contexts $\boldsymbol{C_1}$, $\boldsymbol{C_2}$,and $\boldsymbol{C_3}$, Book1 was never *Minimized* and **20%** of its uses in these contexts employed *Head-fixed* FoR. Conversely, in all other contexts, Book1 was *Minimized*

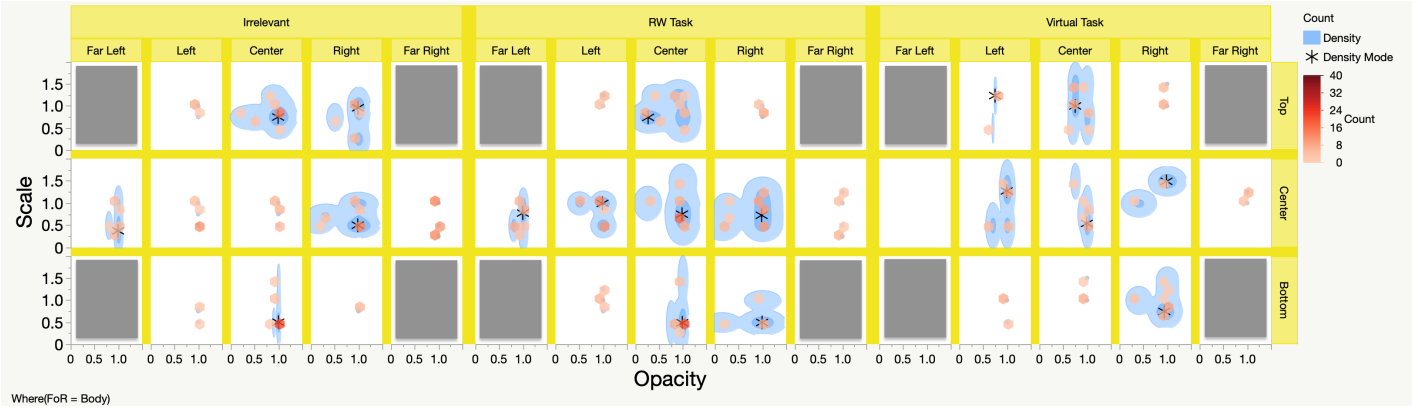


Figure 6.6: Influence of App Role on the user-specified design dimensions within the Body-fixed FoR

approximately **70%** of the time and always employed the *Body-fixed* FoR when visible. The map app was *Minimized* only in $\boldsymbol{C_8}$ and consistently used *Body-fixed FoR* when visible. Notably, except for the Stock app, each app, $\boldsymbol{C_8}$ exhibited the least usage of *Body-fixed* FoR and the highest frequency of *Minimization*.

The role of each app in the user's task within various contexts can be different. Depending on the *Environment*, and the user's *Mobility* and *Objective* in the RW, an app could be *Irrelevant* in a context, *Assistive* to the user's RW task, or *Primary* to their virtual task.

Regardless of the context, the Stock app was always *Primary* for the user's virtual task. The other four apps were *Assistive* to the user's RW task in some contexts and *Irrelevant* in others. In $\boldsymbol{C_8}$, since the user's RW task of reading a physical book did not require any information from the AR interface, all four apps were *Irrelevant*. The Messaging app was considered *Irrelevant* in all contexts except for $\boldsymbol{C_7}$, where it was *Assistive* toward finding their friend. While *Assistive* in finding Book1 in $\boldsymbol{C_1}$, $\boldsymbol{C_2}$, and $\boldsymbol{C_3}$, the Book1 app was *Irrelevant* in all other contexts. Similarly, the Book2 app was *Assistive* in $\boldsymbol{C_4}$, $\boldsymbol{C_5}$, and $\boldsymbol{C_6}$, and *Irrelevant* in all other contexts. The map app in Contexts $\boldsymbol{C_1}$, $\boldsymbol{C_4}$, and $\boldsymbol{C_7}$ was *Assistive* for navigating through a changing and *Unconfined Environment* and was *Irrelevant* in the other contexts.

In our observations, **33%** of the *Irrelevant* apps were *Minimized*. Of the remaining *Irrelevant*

apps that were *Visible*, only **5%** used *Head-fixed*. The apps that were *Assistive* or *Primary* to the user's task were never *Minimized*. *Head-fixed FoR* was applied to **15%** of the time to the app that was *Primary* and **12%** to those that were *Assistive* to the user's task. We explored the statistical significance of patterns involving *App Role* and participants' choices of design dimensions. Our analyses indicate that *App Role* significantly correlates to *Visibility* ($\boldsymbol{\chi^2_{(800,2)} = 202.66,\ p < 0.0001}$) and choice of **FoR** ($\boldsymbol{\chi^2_{(647,2)} = 15.67,\ p \sim 0.0004}$). Within each *FoR*, we evaluated the correlations between *App Role* and *Position*, *Opacity*, and *Scale*.

Body-fixed:

As depicted in Figure 6.6, the *Far Left* sector was never used for the *Primary* app and used only for **9%** of the *Irrelevant* and **9%** of the *Assistive*. When *Body-fixed*, compared to the other apps, the *Primary* app was less frequently placed at the Vertical Center ($\boldsymbol{VirtualTask : 37\%,\ RWTask : 63\%,\ Irrelevant : 55\%}$) and more in the Top sectors ($\boldsymbol{VirtualTask : 33\%,\ RWTask : 20\%,\ Irrelevant : 24\%}$) or Bottom sectors ($\boldsymbol{VirtualTask : 30\%,\ RWTask : 17\%,\ Irrelevant : 21\%}$). These *Body-fixed Position* patterns were confirmed by Likelihood Ratio tests indicating that *App Role* significantly correlates to: *Sector*($\boldsymbol{\chi^2_{(585,20)} = 89.5,\ p < 0.0001}$), *Horizontal Placement*($\boldsymbol{\chi^2_{(585,8)} = 35.4,\ p < 0.0001}$), and *Vertical Placement*($\boldsymbol{\chi^2_{(585,4)} = 20.5,\ p \sim 0.0004}$).

Additionally, in our observations of the *Body-fixed* apps, the *Primary* app was often larger and less opaque than the other apps. Kruskal-Wallis tests revealed a significant influence of *App Role* on both the *Opacity* ($\boldsymbol{\chi^2_{(2)} = 20.8,\ p < 0.0001}$) and the *Scale* ($\boldsymbol{\chi^2_{(2)} = 62.86,\ p < 0.0001}$) of the *Body-fixed* apps.

As shown in Figure 6.7, pairwise comparisons using the Wilcoxon method indicated that

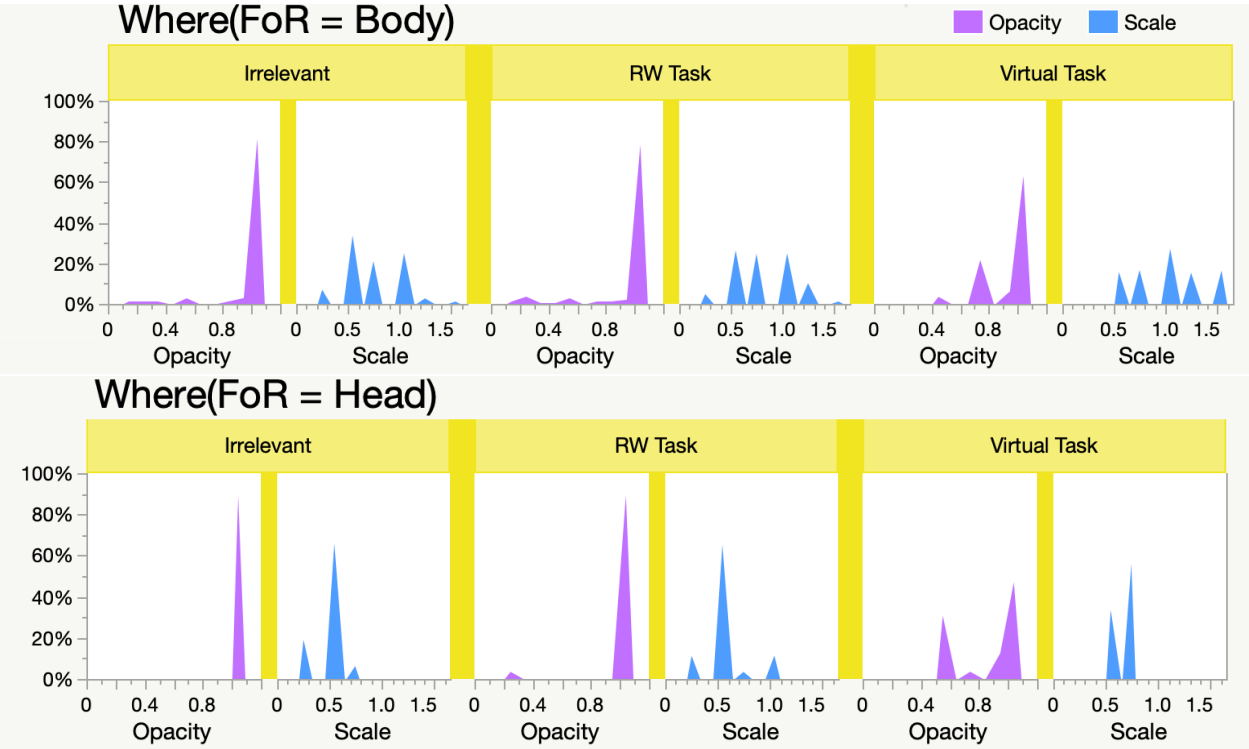


Figure 6.7: The histogram of the Opacity and Scale of apps with different App Roles within Body-fixed (Top) and Head-fixed (Bottom) FoRs

Lower *Opacity* values were used more often for the *Primary* app ($\boldsymbol{M = 0.91,\ STD = 0.15}$) compared to the *Assistive* ($\boldsymbol{M_{Assistive} = 0.92,\ STD_{Assistive} = 0.21,\ Z_{Assistive} = -2.98,\ p_{Assistive} \sim 0.003}$) or the *Irrelevant* ones ($\boldsymbol{M_{Irrelevant} = 0.94,\ STD_{Irrelevant} = 0.19,\ Z_{Irrelevant} = -4.46,\ p_{Irrelevant} < 0.0001}$). This *Primary* app was also more frequently *Scaled* up to larger sizes ($\boldsymbol{M = 0.1,\ STD = 0.33}$) compared to the *Assistive* apps ($\boldsymbol{M_{Assistive} = 0.78,\ STD_{Assistive} = 0.29,\ Z_{Assistive} = 5.4,\ p_{Assistive} < 0.0001}$) or the *Irrelevant* ones ($\boldsymbol{M_{Irrelevant} = 0.71,\ STD_{Irrelevant} = 0.28,\ Z_{Irrelevant} = 7.8,\ p_{Irrelevant} < 0.0001}$). In this FoR, the *Irrelevant* apps were also found to be often smaller than the *Assistive* ones($\boldsymbol{Z = -2.52, p \sim 0.012}$)(Refer to Figure 6.7).

Head-fixed:

Within the *Head-fixed FoR*, *App Role* influenced *Position* ($\boldsymbol{\chi^2_{(62,8)} = 20.71,\ p \sim 0.008}$) and *Vertical Placement*($\boldsymbol{\chi^2_{(62,2)} = 9.33,\ p \sim 0.009}$). The Center Sector was used for placing the *Primary* app **67%** of the time, **79%** for the *Irrelevant* apps, and **83%** for the *Assistive* ones. While the *Primary* app was placed in the *Top Left* sector **25%** of the time, the *Irrelevant* apps never used this sector, and only **13%** of the *Assistive* ones used it.

Kruskal-Wallis tests revealed a significant influnce of *App Role* on *Opacity* ($\boldsymbol{\chi^2_{(2)} = 18.4,\ p < 0.0001}$) and *Scale* ($\boldsymbol{\chi^2_{(2)} = 15.16,\ p \sim 0.0005}$). As shown in Figure 6.7, pairwise comparisons using the Wilcoxon method indicated that the *Primary* app often had *Larger Scale*($\boldsymbol{M_{(Scale,Primary)} = 0.66,\ STD_{(Scale,\ Primary)} = 0.12}$) and *Lower Opacity* ($\boldsymbol{M_{(Opacity,Primary)} = 0.81,\ STD_{(Opacity,Primary)} = 0.23}$) compared to the Scale ($\boldsymbol{M_{(Scale,Assistive)} = 0.54,\ STD_{(Scale,Assistive)} = 0.2,\ Z_{(Scale,Assistive)} = 2.81,\ p_{(Scale,Assistive)} \sim 0.0048}$) and Opacity ($\boldsymbol{M_{(Opacity,Assistive)} = 0.97,\ STD_{(Opacity,Assistive)} = 0.15,\ Z_{(Opacity,Assistive)} = -3.32,\ p_{(Opacity,Assistive)} \sim 0.0009}$) of the *Assistive* ones. Additionally, the *Irrelevant* apps were always fully Opaque ($\boldsymbol{Opacity_{Irrelevant} = 1}$) when *Head-fixed* and Scaled down to smaller sizes ($\boldsymbol{M_{(Scale,Irrelevant)} = 0.46,\ STD_{(Scale,Irrelevant)} = 0.13}$). Therefore, compared to the these *Irrelevant* apps, the *Primary* app was significantly *Larger* with *Lower Opacity* ($\boldsymbol{Z_{(Scale,Irrelevant)} = 3.59,\ p_{(Scale,Irrelevant)} \sim 0.0003}$, $\boldsymbol{Z_{(Opacity,Irrelevant)} = -3.09,\ p_{(Opacity,Irrelevant)} \sim 0.002}$)

In Section 6.4.6, we found an inverse correlation between the *Scale* and *Opacity* of the *head-fixed* apps. Further polynomial simple linear regressions on each App Role found that this inverse correlation does not apply to head-fixed apps that are *Irrelevant*.

6.4.7 Discussion

Our study explored the influence of various contextual components on the design dimensions of various AR apps, revealing notable patterns and correlations, and highlighting the complexity of designing context-aware AR and the significance of our framework in enabling iAR.

The patterns observed across all apps and contexts suggest:

- Body-fixed FoR is more commonly used than head-fixed, while world-fixed is almost never used in a mobile setting.
- The scale and opacity of head-fixed apps are inversely correlated.
- Head-fixed apps are smaller than body-fixed ones.
- Head-fixed apps are preferably placed closer to the center of the (FoV) or near the top left corner.
- Placing head-fixed apps below eye level is not preferred.
- Head-fixed apps placed near the center are typically small.

These findings can suggest several potential design principles for future AR interfaces. However, our results indicate inconsistencies in some design principles when applied to specific contexts or apps. For instance, the inverse correlation between the scale and opacity of head-fixed apps was significantly influenced by the App Role and various contextual components, highlighting the necessity of context awareness.

Additionally, the relevance of apps to the user's task also influenced their design dimensions, suggesting that:

- Relevant apps are always visible.
- The apps that are primary to the user are larger with lower opacity.
- Relevant body-fixed apps are preferably placed at eye level.
- Irrelevant apps rarely use head-fixed FoR.
- Irrelevant body-fixed apps are rarely placed in the Center sector.
- Irrelevant apps are smaller.

These observations underscore the importance of the App Role in determining the optimal design and placement of AR apps.

The study results demonstrated that *RW Objective* significantly influenced the position and visibility of head-fixed applications. These findings also suggest that *Mobility* and *Environment* play a role in how AR interfaces are managed, showing that users tended to minimize more apps when stationary or in an unconfined environment. The small number of direct correlations between design dimensions and the contextual components can be attributed to multiple factors. We observed that participants adapted their interfaces attentively when their RW Objective changed, but less meticulously across different contexts with the same RW Objective. This can explain the lack of correlation between *Environment*, *Mobility*, and design dimensions. Notably, many participants made little to no adaptations between consecutive mobile contexts, potentially contributing to inconsistencies within the contexts with similar *Environment*.

Additionally, all contexts in this experiment were within the same scenario. Common contextual components and implicit contextual factors across these contexts may have overshadowed the influence of the evaluated components. For instance, in this scenario, the user's task involves traveling to various locations within the library. Consequently, even in

contexts where the user is *Stationary*, this contextual factor remains constant. Such implicit contextual information could be the primary reason for not using world-fixed FoR, leading to our observation that *Body-fixed FoR* is more commonly used. Future work exploring these common implicit and explicit contextual factors can provide evidence for the independence of design dimension correlations from specific common contextual components.

While our study provides valuable insights into AR interface design, it also highlights the complexities and need for further research to develop robust context-aware AR. Identifying generalizable patterns for adaptations emphasizes the need for robust factorial analyses and logistic regression models to understand the interaction effects of all independent variables, i.e., contextual components. Even with three contextual components within this study, data from all 16 possible combinations would be essential in providing meaningful conclusions.

Exploring a large number of contexts requires extensive data collected from all these unique contexts. This poses not just a limitation of our work but a significant practical challenge in designing iAR.

As proposed in our framework, inferences allow for the integration of various contextual components, including AR apps and their design dimensions, to identify generalizable rules applicable to different contexts. In this study, the significance of the App Role in determining the optimal design of AR apps underscores the importance of such *Inference*.

Our framework proposes to use both rules and machine learning to make various inferences relevant to specific contexts and optimize over conflicting ones to provide the optimal AR interface. The rule-based inference module in our architecture can utilize findings from this and similar studies as rules to deduce various inferences relevant to specific contexts. On the other hand, the intelligent inference module in our architecture allows for intelligently handling these inferences and further refining the interface.

6.5 Conclusion

To adapt the presentation and interaction of virtual content and improve effectiveness, an iAR interface must collect and meaningfully integrate all relevant context information. In Section 6.2, we provided an architecture for designing iAR in accordance with our proposed framework (*RQ7.*). This architecture utilizes the contextual components from our proposed taxonomy to automatically optimize AR interface adaptations, identified in our design space (*RQ5.* and *RQ6.*).

The Rule-based Inference Module and the Intelligent Inference Module in our architecture enables the iAR system to infer the impact of certain AR adaptations on performance. These modules utilize the context to provide a collection of context-aware inferences that enable iAR to **1)** cover the context-aware interfaces designed and proposed in the previous studies in the field, and **2)** extend beyond the existing context-aware interfaces by learning to personalize and intelligently adapt to the users' behavior. The Adaptation Unit outlines how iAR consolidates and optimizes over these context-aware inferences.

In Section 6.3, we utilized this Architecture to design the context-aware interfaces proposed in Section 3.3, thereby validating the coverage and comprehensiveness of our XR Design Dimensions, Taxonomy of Context, and Framework (*RQ5.*, *RQ6.*, and *RQ7.*).

Section 6.4 further investigated *RQ4.* and presented an experiment conducted across multiple contexts with varying *Mobility*, *Environment*, and *RW Objectives* in a library. This study validated the application of our XR Design Dimensions and Taxonomy of Context for data collection, which is a necessary step towards implementing iAR (*RQ5.*, *RQ6.*). Our analysis of user-specified adaptations to the design elements of various apps in this experiment yielded several key insights and potential design principles. These results further clarified the significance of context on AR effectiveness (*RQ4.*) and underscored the importance of our iAR

framework and architecture in optimizing AR performance through inferences (*RQ7.*). This chapter presented part of our recent submission to TVCG and is currently under review:

- [30] Shakiba Davari, Danniel Stover, Alexander Giovannelli, Cory Ilo, and Doug A. Bowman. "Towards Intelligent Augmented Reality (iAR): A Taxonomy of Context, An Architecture for iAR, and an Empirical Study." IEEE Transactions on Visualization and Computer Graphics, Aug 2024, 12 pages

Chapter 7

Conclusions and Future Work

7.1 Summary

This dissertation outlines a roadmap towards intelligent AR (iAR) systems, aiming to enable effective, efficient, and unobtrusive AR.

In Section 3.3, Section 3.4, Section 4.4 and Section 6.4, we explored the design principles for context-aware AR within multiple contexts and demonstrated the significance of context on AR effectiveness through four user studies.

To enhance AR design and enable adaptations in iAR, Section 4.2 explored various AR design dimensions that influence its effectiveness and provided a comprehensive design space for the content and visual presentation of XR objects.

To provide contextual input to iAR and context-aware AR interfaces, in Section 5.2, we proposed a taxonomy of explicit and quantifiable contextual components. We then introduced a framework for context-aware inference and adaptation using this taxonomy in iAR to enable optimal AR effectiveness (Section 5.3).

Finally, Section 6.2 integrated these components by presenting an architecture that envisions how iAR can utilize our context taxonomy as input to infer the impact of adaptations on each design dimension from our design space and optimize these inferences. This architecture enables the application of design principles and findings from existing context-aware AR

studies within iAR, facilitated by its inclusion of a rule-based inference unit.

7.2 Insights and Reflections

Designing effective AR requires a clear understanding of the issues in AR and the factors that influence its effectiveness. Literature suggests that these issues arise from the context. But how does context and its integration into AR interface design influence AR effectiveness? We explored addressing this question by investigating AR design principles for maintaining efficiency while resolving occlusion, social, and content placement issues in AR (*RQ1.*, *RQ2.*, and *RQ3.*) in four user studies.

These studies involved designing multiple context-aware AR interfaces and exploring their effectiveness across various contexts. They provided valuable insights into the design of context-aware approaches that enhance overall usability and user experience within specific contexts. For example, we found that AR effectiveness improves by **(a)** prioritizing the real world when the user is primarily monitoring or interacting with it (Experiments 1, 2, and 4), **(b)** placing content relevant to a social interaction near the conversation partner (Experiments 2 and 3), and **(c)** avoiding consistent visibility of irrelevant AR content, e.g. head-fixed or centrally body-fixed placements (Experiments 1 and 4). The results from these studies highlight the impact of context on AR effectiveness and demonstrate the potential benefits of context-aware AR.

While context-aware AR provides invaluable insights and *enhances* AR effectiveness, our results show that it may not necessarily achieve *optimal* performance, as it is limited to changing only one or a few design elements based on one or a few context components. For instance, in Experiment 1, participants frequently suggested an interface that prioritizes the real world and automatically provides AR content only when needed. The practicality of

such interfaces depends on their ability to detect the user's virtual content needs, which rely on a combination of numerous contextual factors. Such optimal interfaces can only be realized within iAR systems that consider all context components and can combine them to infer user needs at any given time.

In Experiment 4, we showed that various AR design dimensions are interrelated. For example, in this experiment participants set head-fixed AR content to smaller sizes compared to body-fixed content. This indicates that designing the optimal iAR interface requires consideration of all design dimensions. However, context-aware AR design and evaluations typically explore only one or a few design dimensions while arbitrarily fixing the rest.

Given the vast number of context components and design elements in AR interfaces, achieving optimal AR effectiveness may only be possible through iAR systems. Implementing iAR necessitates a clear understanding of its inputs and output and how to use the input towards achiving the optimal output.

Identifying the output of iAR involves determining the interface design elements that can be adapted to create an effective interface for specific contexts. To provide a clear understanding of iAR outputs, in Chapter 4, we identified these adaptable AR design elements and proposed a comprehensive design space for the content and visual presentation of XR objects (*RQ5.*). Our studies demonstrate the comprehensiveness of this approach.

For instance, in Experiment 4, participants were asked to describe any missing adaptations they wished to implement. Multiple participants expressed a desire to combine the functionality of various FoRs, specifically by integrating body position references with world orientation references. Our design space accommodates this need by proposing a hybrid FoR, illustrating its versatility and comprehensiveness.

To provide a clear understanding of the input to iAR, in Chapter 5, we propose a taxonomy

of explicit context components that can be detected by the iAR system (*RQ6.*). Our studies demonstrate the comprehensiveness of our taxonomy. For example, the context components from our taxonomy provide all the necessary information for implementing the design principles of a context-aware interface in Experiment 2: **(a)** the "interacting people" context component from the SUI category identifies the interlocutor; **(b)** the user's "conversation content" is included within the user physical state context components of our taxonomy; **(c)** the "interlocutor's face position" and "conversation content" are included as the attendee physical state context components within the social setting part of our taxonomy.

Similarly, in Experiment 1, the context involved a user babysitting a child, where the virtual content led to the occlusion of an important real-world object (the child trying to get the user's attention). This context can be indicated through: **(a)** the "occupation" context component from the user profile category, **(b)** the "attendee personal relationship" from the social setting category, and **(c)** the child's "position" and "waving" as attendee physical state context components within the social setting category of our taxonomy.

The results from all these experiments point toward the importance of AR content and its role and relevance within each context. The virtual setting category of our taxonomy includes "AR content", providing the necessary information for implementing iAR and context-aware interfaces effectively.

Our studies underscored the critical importance of information regarding the urgency, relevance, and role of AR content on AR effectiveness and optimal adaptations. In Experiment 1, we discovered that participants' preferences heavily depended on the urgency and relevance of AR content to real-world tasks. Participants generally considered this study's AR content, such as social media, irrelevant and not urgent, preferring to reduce virtual clutter by using transparency and prioritizing the real world.

Similarly, Experiment 4 highlighted the importance of *App Role*, while Experiment 3 emphasized *Info-focus Relevance* in adaptations. These implicit pieces of contextual information, although significantly influencing AR effectiveness, are not quantifiable. Our comprehensive taxonomy includes the necessary context components to identify such information, enabling context-aware AR interfaces to consider such implicit contextual information for specific scenarios through its design principles. However, identifying such design principles, determining which implicit information is essential in each specific scenario, and identifying the appropriate context components to use for detecting this implicit information is challenging.

For instance, in Experiment 3, determining *Info-focus Relevance* varied across contexts. In social contexts, it was indicated by the social setting and "attendee" components of our taxonomy. In the Mobile_Individual context, it relied on the SUI and the user's "distance from the intermediary", while in the Stationary_Individual context, it depended on the "user's visual focus" component of the SUI category in our taxonomy. This demonstrates that even when *Info-focus Relevance* is considered within a context-aware AR's design principles, the interface cannot rely on a single combination of contextual components for detecting *Info-focus Relevance* across different contexts.

This necessitates designing unique and separate context-aware AR for each context within a scenario and underscores the significance of iAR to recognize all context components. These results demonstrate that to effectively utilize the explicit context components from our taxonomy, iAR systems must consider the impact of implicit contextual information on AR effectiveness. This underscores the significance of context-aware inference and adaptation in iAR proposed by our framework in Section 5.3.

To incorporate implicit contextual information, our proposed framework for iAR employs machine learning models to infer the impact of various adaptations and select the optimal AR design dimensions within the current context (*RQ7.*). Experiment 4 revealed that a

single context component did not show a strong correlation and was insufficient to predict the optimal adaptation. However, specific combinations of context components influenced the *App Role*, which significantly affected the optimal adaptations. This underscores the importance of the inferences proposed by this framework.

In Chapter 6, to validate our solutions to (*RQ5.*, *RQ6.*, *RQ7.*), we proposed an architecture elucidating how our design space, taxonomy, and framework can be utilized within iAR. The insights into AR design and the influence of context gained from our user studies enabled us to extract and offer design principles for context-aware approaches in specific contexts. The rule-based inference unit proposed in our architecture allows for the utilization of these design principles and the results from our context-aware AR studies within iAR.

As the complexity and number of context components increase, patterns that may not be human-understandable can be detected more effectively. For example, in Experiment 4, our library scenario distinguished between confined and unconfined spaces. While considering spatial layout as a binary feature might be the only human-understandable approach in context-aware AR, we also gathered depth images and RGB videos to provide more detailed features for understanding the spatial layout of the real-world environment and its correlations with adaptations in iAR. This has been previously demonstrated in the field of computer vision, where researchers, after grappling with human-understandable inferences for solving complex problems, ultimately turned to neural networks. The intelligent inference unit proposed in our architecture enables such advanced inferences for iAR systems.

Within the specific scenario of Experiment 4, our results demonstrated that generalizable rules can be established across all contexts and apps. However, these general rules may contradict those applicable within individual contexts. For instance, considering the size of a head-fixed primary app, the primary app was shown to be larger with lower opacity, while head-fixed apps were generally smaller. The inference consolidation in our architecture

provides a method to integrate such contradictory inferences and optimize the interface accordingly.

7.3 Future Work Directions

While context-aware AR provides invaluable insights and enhances AR effectiveness, achieving optimal AR effectiveness may only be possible through iAR systems.

Achieving a fully intelligent iAR system depends on advancements across multiple domains, including data collection, hardware and sensor technology, and ML/AI algorithms. Capturing, storing, and processing the full spectrum of contextual components, as well as inferring optimal adaptations, require sophisticated algorithms and substantial computational resources. Our proposed taxonomy addresses some of these challenges by enabling iAR systems to focus on quantifiable contextual components that directly impact AR effectiveness and disregard undetectable or irrelevant information.

Although developing an iAR system is ambitious and challenging, it is increasingly attainable given the rapid technological advances. The primary obstacle is not technological feasibility but the need for more focused research and attention to realize this potential. A fully intelligent iAR system could revolutionize AR applications, offering highly personalized, context-aware experiences that seamlessly adapt to users' environments. This dissertation lays the groundwork for the development of iAR interfaces. However, there remains much work for future researchers to undertake. This work highlights the potential of iAR and outlines a roadmap toward this vision, encouraging further research and exploration to bring this concept to reality.

7.3.1 Input and Non-visual Presentation in iAR

In Section 4.2, we explored the design dimensions of XR objects that influence AR effectiveness, briefly introducing non-visual and input design dimensions. However, this work primarily focused on providing a comprehensive design space for the content and visual presentation of AR interfaces.

We briefly examined common design elements for delivering information through non-visual modalities and encouraged researchers to expand upon this line of inquiry. Additionally, interaction design within various input modalities and contexts in AR are interdependent, constantly influencing each other and affecting the user experience. Extensive research to explore the design space of interactions within each input modality can provide invaluable insights for designing iAR systems that automatically adapt interactions to offer the most efficient yet least intrusive input methods.

7.3.2 Frequent Context Switch and Adaptation in iAR

As iAR brings the potential to predict optimal adaptations within each context, several research questions regarding the transition from one interface design to another arise, particularly in scenarios with frequent context-switching.

For example, in Experiment 3, we demonstrated that *Info-focus Relevance* significantly influenced the effectiveness of the placement strategy and varied for each question asked within the same context. iAR can potentially predict the optimal placement strategy for each question. This raises the critical research question: how frequently should adaptations occur? Research is needed to explore the trade-off between a frequently-adapted optimal interface and a consistent sub-optimal one in various contexts. Additional questions include investigating how to transition from one strategy to another, how to indicate to the user that the

interface's strategy has changed, and how to communicate the current strategy to the user effectively.

7.3.3 Data Collection and Implementation

Although our proposed framework and architecture lay the groundwork for future research and development in this field, implementing and evaluating iAR in real-world scenarios will be critical. Future work must focus on the practical implementation of these concepts. In contemporary context-aware interfaces, designers deduce human-understandable inferences based on empirical data to create appropriate interfaces. Given the vast number of context components and design dimensions, manually processing all conceivable contexts is impractical. Leveraging Neural Networks (NNs), Reinforcement Learning (RL), and Machine Learning (ML) models within our architecture allows iAR systems to automatically adapt to a broader set of contexts, including those not previously encountered. This highlights the urgent need for further research to provide a clear understanding of the ML models required to enable such iAR systems.

Additionally, implementing iAR requires extensive data collection for training, which has not yet been fully realized. Our experiment in Section 6.4 provides preliminary data, including gaze, spatial, and video data, as well as interface design, adaptations, and performance metrics for a specific scenario. To enable iAR, more data from users' behavior, gathered from larger sets of participants across various contexts and scenarios, is needed. These efforts must ensure that all potentially changing context components and AR design dimensions are considered and detected.

Our taxonomy of context in Chapter 5 and our design space proposed in Chapter 4 provide a comprehensive list of input and output information necessary for training future iAR

interfaces. Future research should focus on collecting training data for iAR systems. We recommend starting with the implementation of iAR in specific scenarios. Highly dynamic scenarios, such as mobile and public environments, are less ideal because they require iAR to consider all context components, accommodating nearly limitless possible context switches. Initial implementation and data collection efforts should focus on a specific set of adaptations within commonly used and less dynamic scenarios. This approach helps collect a higher volume of data when the number of changing context components and adaptations is minimal, facilitating the attribution of adaptations to specific context components. For example, exploring applications like office work using the Apple Vision Pro could provide valuable insights due to the relatively larger number of power users in such scenarios.

Given the current scarcity of available data, until AR becomes widely accessible and more natural behavior data can be collected, alternative approaches, such as using Generative Adversarial Networks (GANs) and other models to generate training data, should be explored. While Apple Vision Pro has introduced AR to a broader audience, further efforts in hardware and optics research are needed to enhance the AR form factor and make it more affordable for everyone. This advancement will facilitate better data collection in AR.

7.3.4 Security & Privacy

iAR interfaces should be designed with caution regarding the user's physical safety, security, and privacy. Implementing iAR requires the collection of extensive data from individuals' personal lives. Ensuring the security and privacy of both users and bystanders in the implementation and data handling of iAR systems is paramount. Before integrating iAR into people's personal lives, future research must explore potential security and privacy risks and address these concerns to enable the safe and widespread use of iAR technologies.

Bibliography

[1] Mariek MP Vanden Abeele, Marjolijn L Antheunis, and Alexander P Schouten. The effect of mobile messaging during a conversation on impression formation and interaction quality. *Computers in Human Behavior*, 62:562–569, 2016.

[2] Gregory D. Abowd, Anind K. Dey, Peter J. Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a better understanding of context and context-awareness. In *Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing*, HUC '99, page 304–307, Berlin, Heidelberg, 1999. Springer-Verlag. ISBN 3540665501.

[3] Ralph Adolphs. Social cognition and the human brain. *Trends in cognitive sciences*, 3 (12):469–479, 1999.

[4] Ronald Azuma and Chris Furmanski. Evaluating label placement for augmented reality view management. In *Proceedings of the 2nd IEEE/ACM international Symposium on Mixed and Augmented Reality*, page 66. IEEE Computer Society, 2003.

[5] Ronald T Azuma. A survey of augmented reality. *Presence: teleoperators & virtual environments*, 6(4):355–385, 1997.

[6] Phil Baines and Andrew Haslam. *Type & typography*. Laurence King Publishing, 2005.

[7] Jared S Bauer, Sunny Consolvo, Benjamin Greenstein, Jonathan Schooler, Eric Wu, Nathaniel F Watson, and Julie Kientz. Shuteye: encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1401–1410. ACM, 2012.

[8] Blaine Bell, Steven Feiner, and Tobias Höllerer. View management for virtual and augmented reality. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 101–110, 2001.

[9] Klaus Bengler and Robert Passaro. Augmented reality in cars: Requirements and constraints. *ISMAR'06 industrial track*, 2006.

[10] M-O Berger. Resolving occlusion in augmented reality: a contour based approach without 3d reconstruction. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 91–96. IEEE, 1997.

[11] Verena Biener, Eyal Ofek, Michel Pahud, Per Ola Kristensson, and Jens Grubert. Extended reality for knowledge work in everyday environments. *arXiv preprint arXiv:2111.03942*, 2021.

[12] Mark Billinghurst and Thad Starner. Wearable devices: new ways to manage information. *Computer*, 32(1):57–64, 1999.

[13] Mark Billinghurst, Adrian Clark, and Gun Lee. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, 8(2-3):73–272, 2015. ISSN 1551-3955. doi: 10.1561/1100000049. URL http://dx.doi.org/10.1561/1100000049.

[14] Jeremy Birnholtz, Lindsay Reynolds, Eli Luxenberg, Carl Gutwin, and Maryam Mustafa. Awareness beyond the desktop: exploring attention and distraction with a projected peripheral-vision display. In *Proceedings of Graphics Interface 2010*, pages 55–62. Canadian Information Processing Society, 2010.

[15] Doug Bowman, Ernst Kruijff, Joseph J LaViola Jr, and Ivan P Poupyrev. *3D User interfaces: theory and practice, CourseSmart eTextbook*. Addison-Wesley, 2004.

[16] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. An introduction to 3-d user interface design. *Presence*, 10(1):96–108, 2001. doi: 10.1162/105474601750182342.

[17] Doug A. Bowman, Joseph Gabbard, Daniel Auerbach, Nazila Roofigari-Esfahan, Kathryn Britt, Cory I. Ilo, and Keerthana Adapa. Buildar: A proof-of-concept prototype of intelligent augmented reality in construction. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 508–512, 2022. doi: 10.1109/VRW55335.2022.00113.

[18] John Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[19] JJ Cadiz, Gina Venolia, Gavin Jancke, and Anoop Gupta. Sideshow: Providing peripheral awareness of important information. Technical Report MSR-TR-2001-83, Microsoft, September 2001. URL https://www.microsoft.com/en-us/research/publication/sideshow-providing-peripheral-awareness-of-important-information/.

[20] Wan-Jung Chang, Liang-Bi Chen, and Yu-Zung Chiou. Design and implementation of a drowsiness-fatigue-detection system based on wearable smart glasses to increase road safety. *IEEE Transactions on Consumer Electronics*, 64(4):461–469, 2018. doi: 10.1109/TCE.2018.2872162.

[21] Ishan Chatterjee, Robert Xiao, and Chris Harrison. Gaze+ gesture: Expressive, precise and targeted free-space interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 131–138, 2015.

[22] Yifei Cheng, Yukang Yan, Xin Yi, Yuanchun Shi, and David Lindlbauer. Semanticadapt: Optimization-based adaptation of mixed reality layouts leveraging virtual-

physical semantic connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pages 282–297, 2021.

[23] Karen Church, Mauro Cherubini, and Nuria Oliver. A large-scale study of daily information needs captured in situ. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(2):1–46, 2014.

[24] Enrico Costanza, Samuel A Inverso, Elan Pavlov, Rebecca Allen, and Pattie Maes. Eye-q: Eyeglass peripheral display for subtle intimate notifications. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, pages 211–218. ACM, 2006.

[25] James E Cutting and Peter M Vishton. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In *Perception of space and motion*, pages 69–117. Elsevier, 1995.

[26] Shakiba Davari. [dc] context-aware inference and adaptation in augmented reality. In *2022 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 938–939. IEEE, 2022. doi: 10.1109/VRW55335.2022.00320.

[27] Shakiba Davari and Doug. A. Bowman. Towards context-aware adaptation in extended reality: A design space for xr interfaces and an adaptive placement strategy. *IEEE Transactions on Visualization and Computer Graphics*, 2024.

[28] Shakiba Davari, Feiyu Lu, and Doug A. Bowman. Occlusion management techniques for everyday glanceable ar interfaces. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 324–330. IEEE, 2020. doi: 10.1109/VRW50115.2020.00072.

[29] Shakiba Davari, Feiyu Lu, and Doug A Bowman. Validating the benefits of glanceable

and context-aware augmented reality for everyday information access tasks. In *2022 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 336–444. IEEE, 2022. doi: 10.1109/VR51125.2022.00063.

[30] Shakiba Davari, Danny Stover, Alexander Giovannelli, Cory Ilo, and Doug. A. Bowman. Towards intelligent augmented reality (iar): A taxonomy of context, an architecture for iar, and an empirical study. *IEEE Transactions on Visualization and Computer Graphics*, 2024.

[31] David Dearman, Melanie Kellar, and Khai N Truong. An examination of daily information needs and sharing opportunities. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 679–688. ACM, 2008.

[32] Anind K Dey, Gregory D Abowd, and Daniel Salber. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human–Computer Interaction*, 16(2-4):97–166, 2001.

[33] Stephen Di Verdi, Daniel Nurmi, and Tobias Hollerer. Arwin-a desktop augmented reality window manager. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pages 298–299. IEEE, 2003.

[34] Stephen DiVerdi, Tobias Hollerer, and Richard Schreyer. Level of detail interfaces. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 300–301. IEEE, 2004.

[35] David Drascic and Paul Milgram. Perceptual issues in augmented reality. In *Stereoscopic displays and virtual reality systems III*, volume 2653, pages 123–134. International Society for Optics and Photonics, 1996.

[36] Andrew T Duchowski. Gaze-based interaction: A 30 year retrospective. *Computers & Graphics*, 73:59–69, 2018.

[37] Phillip S Dunston et al. Identification of application areas for augmented reality in industrial construction based on technology suitability. *Automation in Construction*, 17(7):882–894, 2008.

[38] Mica R Endsley. Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society annual meeting*, 32(2):97–101, 1988.

[39] Barrett Ens, Eyal Ofek, Neil Bruce, and Pourang Irani. Spatial constancy of surface-embedded layouts across multiple environments. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, pages 65–68, 2015.

[40] Barrett M Ens, Rory Finnegan, and Pourang P Irani. The personal cockpit: a spatial interface for effective task switching on head-worn displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3171–3180, 2014.

[41] S.K. Feiner. The importance of being mobile: some social consequences of wearable augmented reality systems. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, pages 145–148, 1999. doi: 10.1109/IWAR.1999.803815.

[42] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. Windows on the world: 2d windows for 3d augmented reality. In *Proceedings of the 6th annual ACM symposium on User interface software and technology*, pages 145–155, 1993.

[43] Steven K Feiner. Augmented reality: A new way of seeing. *Scientific American*, 286 (4):48–55, 2002.

[44] Qi Feng, Hubert PH Shum, and Shigeo Morishima. Bi-projection-based foreground-aware omnidirectional depth prediction. *DU*, 2021.

[45] Gerhard Fischer. User modeling in human–computer interaction. *User modeling and user-adapted interaction*, 11:65–86, 2001.

[46] Alexandra Frischen, Andrew P Bayliss, and Steven P Tipper. Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, 133 (4):694, 2007.

[47] Chris Frith. Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3453–3458, 2009.

[48] Nirit Gavish, Teresa Gutiérrez, Sabine Webel, Jorge Rodríguez, Matteo Peveri, Uli Bockholt, and Franco Tecchia. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, 23(6):778–798, 2015. doi: 10.1080/10494820.2013.815221. URL https://doi.org/10.1080/10494820.2013.815221.

[49] Alexander Giovannelli, Jerald Thomas, Logan Lane, Francielly Rodrigues, and Doug A Bowman. Gestures vs. emojis: Comparing non-verbal reaction visualizations for immersive collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 2023.

[50] Joseph H Goldberg and Jack C Schryver. Eye-gaze-contingent control of the computer interface: Methodology and example for zoom detection. *Behavior research methods, instruments, & computers*, 27(3):338–350, 1995.

[51] Mar Gonzalez-Franco, Eyal Ofek, Ye Pan, Angus Antley, Anthony Steed, Bernhard Spanlang, Antonella Maselli, Domna Banakou, Nuria Pelechano, Sergio Orts-Escolano,

et al. The rocketbox library and the utility of freely available rigged avatars. *Frontiers in virtual reality*, 1:20, 2020.

[52] Raphaël Grasset, Tobias Langlotz, Denis Kalkofen, Markus Tatzgern, and Dieter Schmalstieg. Image-driven view management for augmented reality browsers. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 177–186, 2012. doi: 10.1109/ISMAR.2012.6402555.

[53] Kristen Grauman, Margrit Betke, Jonathan Lombardi, James Gips, and Gary R Bradski. Communication via eye blinks and eyebrow raises: Video-based human-computer interfaces. *Universal Access in the Information Society*, 2(4):359–373, 2003.

[54] Jens Grubert. Mixed reality interaction techniques. *arXiv preprint arXiv:2103.05984*, 2021.

[55] Jens Grubert, Daniel Hamacher, Rüdiger Mecke, Irina Böckelmann, Lutz Schega, Anke Huckauf, Mario Urbina, Michael Schenk, Fabian Doil, and Johannes Tümler. Extended investigations of user-related issues in mobile industrial ar. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pages 229–230, 2010. doi: 10.1109/ISMAR.2010.5643581.

[56] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE transactions on visualization and computer graphics*, 23(6):1706–1724, 2016.

[57] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 23(6):1706–1724, 2017. doi: 10.1109/TVCG.2016.2543720.

[58] Jan Gugenheimer, Evgeny Stemasov, Harpreet Sareen, and Enrico Rukzio. Facedisplay: Towards asymmetric multi-user interaction for nomadic virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[59] Jan Gugenheimer, Christian Mai, Mark McGill, Julie Williamson, Frank Steinicke, and Ken Perlin. Challenges using head-mounted displays in shared and social spaces. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2019.

[60] Jeffrey A Hall, Nancy K Baym, and Kate M Miltner. Put down that phone and talk to me: Understanding the roles of mobile phone norm adherence and similarity in relationships. *Mobile Media & Communication*, 2(2):134–153, 2014.

[61] Chris Harrison and Anind K Dey. Lean and zoom: proximity-aware user interface and content magnification. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 507–510, 2008.

[62] Sandra G. Hart, Sandra G. Hart, Lowell E. Staveland, and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 1988. doi: 10.1016/s0166-4115(08)62386-9.

[63] Jeremy M Heiner, Scott E Hudson, and Kenichiro Tanaka. The information percolator: ambient information display in a decorative object. In *Proceedings of the 12th annual ACM symposium on User interface software and technology*, pages 141–148. ACM, 1999.

[64] Niels Henze. *Camera-based Mobile Interaction with Physical Objects*, volume 23. OIWIR Oldenburger Verlag für Wirtschaft, Informatik und Recht, 2012.

[65] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. A design space for gaze interaction on head-mounted displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[66] Tobias Höllerer, Steven Feiner, Drexel Hallaway, Blaine Bell, Marco Lanzagorta, Dennis Brown, Simon Julier, Yohan Baillot, and Lawrence Rosenblum. User interface management techniques for collaborative mobile augmented reality. *Computers & Graphics*, 25(5):799–810, 2001.

[67] David Holstius, John Kembel, Amy Hurst, Peng-Hui Wan, and Jodi Forlizzi. Infotropism: living and robotic plants as interactive displays. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 215–221. ACM, 2004.

[68] S. A. Hood, B. M. Beauchesne, T. A. Fahmie, and A. Go. Descriptive assessment of conversational skills: towards benchmarks for young adults with social deficits. *Journal of Applied Behavior Analysis*, 54:1075–1094, 2021. doi: 10.1002/jaba.831.

[69] Annette Horstmann and Klaus-Peter Hoffmann. Target selection in eye–hand coordination: Do we reach to where we look or do we look to where we reach? *Experimental Brain Research*, 167(2):187–195, 2005.

[70] Brandon Huynh, Jason Orlosky, and Tobias Höllerer. Designing a multitasking interface for object-aware ar applications. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 39–40, 2020. doi: 10.1109/ISMAR-Adjunct51615.2020.00025.

[71] Robert JK Jacob. What you look at is what you get: eye movement-based interaction

techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 11–18. ACM, 1990.

[72] N. F. S. Jeffri and S. Sulaiman. A review of augmented reality systems and their effects on mental workload and task performance. *Heliyon*, 7:e06277, 2021. doi: 10.1016/j.heliyon.2021.e06277.

[73] Roland S Johansson, Göran Westling, Anders Bäckström, and J Randall Flanagan. Eye–hand coordination in object manipulation. *Journal of Neuroscience*, 21(17):6917–6932, 2001.

[74] Richie Jose, Gun A Lee, and Mark Billinghurst. A comparative study of simulated augmented reality displays for vehicle navigation. In *Proceedings of the 28th Australian conference on computer-human interaction*, pages 40–48, 2016.

[75] Seungwon Kim, Mark Billinghurst, Gun Lee, and Weidong Huang. Gaze window: A new gaze interface showing relevant content close to the gaze point. *Journal of the Society for Information Display*, 28(12):979–996, 2020.

[76] Pascal Knierim, Paweł W Woźniak, Yomna Abdelrahman, and Albrecht Schmidt. Exploring the potential of augmented reality in domestic environments. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–12, 2019.

[77] R. v. Krevelen and R. Poelman. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 9:1–20, 2010. doi: 10.20870/ijvr.2010.9.2.2767.

[78] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A Lee, and Mark Billinghurst. Pinpointing: Precise head-and eye-based target selection for augmented

reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018.

[79] Wallace Lages and Doug Bowman. Adjustable adaptation for spatial augmented reality workspaces. In *Symposium on Spatial User Interaction*, page 20. ACM, 2019.

[80] Wallace S Lages and Doug A Bowman. Walking with adaptive augmented reality workspaces: design and usage patterns. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pages 356–366, 2019.

[81] Tobias Langlotz, Thanh Nguyen, Dieter Schmalstieg, and Raphael Grasset. Next-generation augmented reality browsers: Rich, seamless, and adaptive. *Proceedings of the IEEE*, 102(2):155–169, 2014. doi: 10.1109/JPROC.2013.2294255.

[82] Tobias Langlotz, Thanh Nguyen, Dieter Schmalstieg, and Raphael Grasset. Next-generation augmented reality browsers: rich, seamless, and adaptive. *Proceedings of the IEEE*, 102(2):155–169, 2014.

[83] Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and evaluation of a user experience questionnaire. In Andreas Holzinger, editor, *HCI and Usability for Education and Work*, pages 63–76, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. ISBN 978-3-540-89350-9. doi: https://10.1007/978-3-540-89350-9_6.

[84] E. Laviola, M. Gattullo, and V.M. et al. Manghisi. Minimal ar: visual asset optimization for the authoring of augmented reality work instructions in manufacturing. In *The International Journal of Advanced Manufacturing Technology*, volume 119, page 1769–1784, 2022. doi: doi.org/10.1007/s00170-021-08449-6.

[85] Youngho Lee, Choonsung Shin, Alexander Plopski, Yuta Itoh, Thammathip Piumsomboon, Arindam Dey, Gun Lee, Seungwon Kim, and Mark Billinghurst. Estimating gaze

depth using multi-layer perceptron. In *2017 International Symposium on Ubiquitous Virtual Reality (ISUVR)*, pages 26–29. IEEE, 2017.

[86] Vincent Lepetit and M-O Berger. A semi-automatic method for resolving occlusion in augmented reality. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*, volume 2, pages 225–230. IEEE, 2000.

[87] Yuan Li, Feiyu Lu, Wallace S Lages, and Doug Bowman. Gaze direction visualization techniques for collaborative wide-area model-free augmented reality. In *Symposium on spatial user interaction*, pages 1–11, 2019.

[88] Zhen Li, Michelle Annett, Ken Hinckley, Karan Singh, and Daniel Wigdor. Holodoc: Enabling mixed reality workspaces that harness physical and digital content. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–14, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359702. doi: 10.1145/3290605.3300917. URL https://doi.org/10.1145/3290605.3300917.

[89] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. Context-aware online adaptation of mixed reality interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pages 147–160, 2019.

[90] Weiyuan Liu. Natural user interface-next mainstream product user interface. In *2010 IEEE 11th International Conference on Computer-Aided Industrial Design & Conceptual Design 1*, volume 1, pages 203–205. IEEE, 2010.

[91] Mark A Livingston, J Edward Swan, Joseph L Gabbard, Tobias H Hollerer, Deborah Hix, Simon J Julier, Yohan Baillot, and Dennis Brown. Resolving multiple occluded layers in augmented reality. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pages 56–65. IEEE, 2003.

[92] Feiyu Lu and Doug A Bowman. Evaluating the potential of glanceable AR interfaces for authentic everyday uses. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 768–777. IEEE, 2021.

[93] Feiyu Lu, Shakiba Davari, Lee Lisle, Yuan Li, and Doug A. Bowman. Glanceable AR: Evaluating information access methods for head-worn augmented reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 930–939, 2020. doi: 10.1109/VR46266.2020.00113.

[94] Feiyu Lu, Shakiba Davari, and Doug Bowman. Exploration of techniques for rapid activation of glanceable information in head-worn augmented reality. In *Symposium on Spatial User Interaction*, SUI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450390910. doi: 10.1145/3485279.3485286. URL https://doi.org/10.1145/3485279.3485286.

[95] Blair MacIntyre, Elizabeth D. Mynatt, Stephen Voida, Klaus M. Hansen, Joe Tullio, and Gregory M. Corso. Support for multitasking and background awareness using interactive peripheral displays. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, UIST '01, page 41–50, New York, NY, USA, 2001. Association for Computing Machinery. ISBN 158113438X. doi: 10.1145/502348.502355. URL https://doi.org/10.1145/502348.502355.

[96] Paul P Maglio and Christopher S Campbell. Tradeoffs in displaying peripheral information. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 241–248. ACM, 2000.

[97] Koji Makita, Masayuki Kanbara, and Naokazu Yokoya. View management of annotations for wearable augmented reality. In *2009 IEEE International Conference on Multimedia and Expo*, pages 982–985. IEEE, 2009.

[98] Pavel Manakhov, Ludwig Sidenmark, Ken Pfeuffer, and Hans Gellersen. Gaze on the go: Effect of spatial reference frame on visual target acquisition during physical locomotion in extended reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. doi: 10.1145/3613904.3642915. URL https://doi.org/10.1145/3613904.3642915.

[99] Tara Matthews, Jodi Forlizzi, and Stacie Rohrbach. Designing glanceable peripheral displays. *UC Berkeley. Retrieved September*, 23:2015, 2006.

[100] Tara Lynn Matthews, Jennifer Mankoff, and John Canny. *Designing and evaluating glanceable peripheral displays*. University of California, Berkeley, 2007.

[101] Ann McNamara and Chethna Kabeerdoss. Mobile augmented reality: Placing labels based on gaze position. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 36–37. IEEE, 2016.

[102] Ann McNamara, Katherine Boyd, Joanne George, Weston Jones, Somyung Oh, and Annie Suther. Information placement in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1765–1769. IEEE, 2019.

[103] M. R. Miller, H. Jun, F. Herrera, J. Y. Villa, G. Welch, and J. N. Bailenson. Social interaction in augmented reality. *Plos One*, 14:e0216290, 2019. doi: 10.1371/journal.pone.0216290.

[104] Todd Miller and John Stasko. The infocanvas: information conveyance through personalized, expressive art. In *Conference on Human Factors in Computing Systems: CHI'01 extended abstracts on Human factors in computing systems*, volume 31, pages 305–306, 2001.

[105] Cecily Morrison, Edward Cutrell, Martin Grayson, Anja Thieme, Alex Taylor, Geert Roumen, Camilla Longden, Sebastian Tschiatschek, Rita Faia Marques, and Abigail Sellen. Social sensemaking with ai: Designing an open-ended ai experience with a blind child. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. URL https://doi.org/10.1145/3411764.3445290.

[106] Aunnoy K Mutasim, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. Pinch, click, or dwell: Comparing different selection techniques for eye-gaze-based pointing in virtual reality. In *Acm symposium on eye tracking research and applications*, pages 1–7, 2021.

[107] Brad A. Myers. User interface software tools. *ACM Trans. Comput.-Hum. Interact.*, 2(1):64–103, mar 1995. ISSN 1073-0516. doi: 10.1145/200968.200971. URL https://doi.org/10.1145/200968.200971.

[108] Jason Orlosky, Kiyoshi Kiyokawa, Takumi Toyama, and Daniel Sonntag. Halo content: Context-aware viewspace management for non-invasive augmented reality. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 369–373, 2015.

[109] Hyung Min Park, Seok Han Lee, and Jong Soo Choi. Wearable augmented reality system using gaze interaction. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 175–176. IEEE, 2008.

[110] J. Pascoe. Adding generic contextual capabilities to wearable computers. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*, pages 92–99, 1998. doi: 10.1109/ISWC.1998.729534.

[111] Charith Perera, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos.

Context aware computing for the internet of things: A survey. *IEEE communications surveys & tutorials*, 16(1):414–454, 2013.

[112] Thies Pfeiffer, Marc Erich Latoschik, and Ipke Wachsmuth. Evaluation of binocular eye trackers and algorithms for 3d gaze interaction in virtual reality environments. *JVRB-Journal of Virtual Reality and Broadcasting*, 5(16), 2008.

[113] Ken Pfeuffer, Yasmeen Abdrabou, Augusto Esteves, Radiah Rivu, Yomna Abdelrahman, Stefanie Meitner, Amr Saadi, and Florian Alt. Artention: A design space for gaze-adaptive user interfaces in augmented reality. *Computers & Graphics*, 95:1–12, 2021.

[114] Robin Piening, Ken Pfeuffer, Augusto Esteves, Tim Mittermeier, Sarah Prange, Philippe Schröder, and Florian Alt. Looking for info: Evaluation of gaze based information retrieval in augmented reality. In *IFIP Conference on Human-Computer Interaction*, pages 544–565. Springer, 2021.

[115] Stephen M Reder. On-line monitoring of eye-position signals in contingent and non-contingent paradigms. *Behavior Research Methods & Instrumentation*, 5(2):218–228, 1973.

[116] Bradley Rhodes. Wimp interface considered fatal. In *IEEE VRAIS'98: Workshop on Interfaces for Wearable Computers*, 1998.

[117] Julie Rico and Stephen Brewster. Gesture and voice prototyping for early evaluations of social acceptability in multimodal interfaces. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, page 16. ACM, 2010.

[118] Radiah Rivu, Yasmeen Abdrabou, Ken Pfeuffer, Augusto Esteves, Stefanie Meitner,

and Florian Alt. Stare: Gaze-assisted face-to-face communication in augmented reality. In *ACM Symposium on Eye Tracking Research and Applications*, pages 1–5, 2020.

[119] Quentin Roy, Futian Zhang, and Daniel Vogel. Automation accuracy is good, but high controllability may be better. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2019.

[120] Enrico Rukzio, Massimo Paolucci, Tim Finin, Paul Wisner, and Terry Payne. Mobile interaction with the real world. In *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '06, page 295–296, New York, NY, USA, 2006. Association for Computing Machinery. ISBN 1595933905. doi: 10.1145/1152215.1152296. URL https://doi.org/10.1145/1152215.1152296.

[121] B. Schilit, N. Adams, and R. Want. Context-aware computing applications. In *1994 First Workshop on Mobile Computing Systems and Applications*, pages 85–90, 1994. doi: 10.1109/WMCSA.1994.16.

[122] Dieter Schmalstieg and Tobias Hollerer. Augmented reality: Principles and practice. In *2017 IEEE Virtual Reality (VR)*, pages 425–426. IEEE Computer Society, 2017.

[123] Manisah Mohd Shah, Haslina Arshad, and Riza Sulaiman. Occlusion in augmented reality. In *2012 8th International Conference on Information Science and Digital Content Technology (ICIDT2012)*, volume 2, pages 372–378. IEEE, 2012.

[124] Rongkai Shi, Nan Zhu, Hai-Ning Liang, and Shengdong Zhao. Exploring head-based mode-switching in virtual reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 118–127, 2021. doi: 10.1109/ISMAR52148.2021.00026.

[125] Richard Skarbez, Missie Smith, and Mary C Whitton. Revisiting milgram and kishino's reality-virtuality continuum. *Frontiers in Virtual Reality*, 2:647997, 2021.

[126] Timothy Sohn, Kevin A Li, William G Griswold, and James D Hollan. A diary study of mobile information needs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 433–442. ACM, 2008.

[127] Marie-Hélène Stoltz, Vaggelis Giannikas, Duncan McFarlane, James Strachan, Jumyung Um, and Rengarajan Srinivasan. Augmented reality in warehouse operations: opportunities and barriers. *IFAC-PapersOnLine*, 50(1):12979–12984, 2017.

[128] Vildan Tanriverdi and Robert JK Jacob. Interacting with eye movements in virtual environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 265–272, 2000.

[129] Ying-Chao Tung, Chun-Yen Hsu, Han-Yu Wang, Silvia Chyou, Jhe-Wei Lin, Pei-Jung Wu, Andries Valstar, and Mike Y Chen. User-defined game input for smart glasses in public space. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3327–3336, 2015.

[130] R Brian Valimont, Dennis A Vincenzi, Sathya N Gangadharan, and AE Majoros. The effectiveness of augmented reality as a facilitator of information acquisition. In *Proceedings. The 21st Digital Avionics Systems Conference*, volume 2, pages 7C5–7C5. IEEE, 2002.

[131] Maarten Van Dantzich, Daniel Robbins, Eric Horvitz, and Mary Czerwinski. Scope: Providing awareness of multiple notifications at a glance. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 267–281. ACM, 2002.

[132] Roel Vertegaal et al. Attentive user interfaces. *Communications of the ACM*, 46(3): 30–33, 2003.

[133] Roy Want. An introduction to ubiquitous computing. In *Ubiquitous computing fundamentals*, pages 1–36, 2010.

[134] Martin Weier, Thorsten Roth, André Hinkenjann, and Philipp Slusallek. Predicting the gaze depth in head-mounted displays using multiple feature regression. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, pages 1–9, 2018.

[135] Mark Weiser. Some computer science issues in ubiquitous computing. *Communications of the ACM*, 36(7):75–84, 1993.

[136] Mark Weiser. The computer for the 21st century. *ACM SIGMOBILE mobile computing and communications review*, 3(3):3–11, 1999.

[137] Mark Weiser and John Seely Brown. Designing calm technology. *PowerGrid Journal*, 1(1):75–85, 1996.

[138] Mark Weiser and John Seely Brown. The coming age of calm technology. In *Beyond calculation*, pages 75–85. Springer, 1997.

[139] Dirk Wenig, Johannes Schöning, Alex Olwal, Mathias Oben, and Rainer Malaka. Watchthru: Expanding smartwatch displays with mid-air visuals and wrist-worn augmented reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 716–721. ACM, 2017.

[140] Sean Michael White. *Interaction and presentation techniques for situated visualization*. Columbia University, 2009.

[141] Craig Wisneski, Hiroshi Ishii, Andrew Dahley, Matt Gorbet, Scott Brave, Brygg Ullmer, and Paul Yarin. Ambient displays: Turning architectural space into an interface between people and digital information. In *International Workshop on Cooperative Buildings*, pages 22–32. Springer, 1998.

[142] Matthias M Wloka and Brian G Anderson. Resolving occlusion in augmented reality. In *Proceedings of the 1995 symposium on Interactive 3D graphics*, I3D '95, pages 5–12, New York, NY, USA, 1995. Association for Computing Machinery. ISBN 0897917367. doi: 10.1145/199404.199405. URL https://doi.org/10.1145/199404.199405.

[143] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011. doi: 10.1145/1978942.1978963.

[144] Wenge Xu, Hai-Ning Liang, Anqi He, and Zifan Wang. Pointing and selection methods for text entry in augmented reality head mounted displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 279–288. IEEE, 2019.

[145] Difeng Yu, Hai-Ning Liang, Xueshi Lu, Tianyu Zhang, and Wenge Xu. Depthmove: Leveraging head motions in the depth dimension to interact with virtual reality head-worn displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 103–114. IEEE, 2019.

[146] Xianjun Sam Zheng, Cedric Foucault, Patrik Matos da Silva, Siddharth Dasari, Tao Yang, and Stuart Goose. Eye-wearable technology for machine maintenance: Effects of display position and hands-free operation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2125–2134, 2015.

Appendices

Appendix A

Experiment 1: AR Occlusion Management in a Real-World Monitoring Context

A.1 Questionnaires and Interview

A.1.1 Post-Condition: NASA TLX Questionnaire

NASA TLX Workload Questionnaire

Participant #:

Technique:

Please circle the appropriate response for each question:

1. **Mental Demand:**
   How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?

   1 (Very Low) – 7 (Very high)

2. **Physical Demand**
   How much physical activity was required? Was the task easy or demanding, slack or strenuous?

   1 (Very Low) – 7 (Very high)

3. **Temporal Demand**
   How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?

   1 (Very Low) – 7 (Very high)

4. **Overall Performance**
   How successful were you in performing the task? How satisfied were you with your performance?

   1 (Very Low) – 7 (Very high)

5. **Effort**
   How hard did you have to work (mentally and physically) to accomplish your level of performance?

   1 (Very Low) – 7 (Very high)

6. **Frustration Level**
   How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?

   1 (Very Low) – 7 (Very high)

A.1.2 Post-Condition Usability Questionnaire

Usability Questionnaire
Participant #:
Technique:
Please circle the appropriate response for each question:

1. I think that I would like to use this technique frequently.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

2. I found the technique unnecessarily complex.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

3. I thought the technique was easy to use.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

4. I think that I would need the support of a technical person to be able to use this technique.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

5. I thought there was too much inconsistency in this technique.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

6. I would imagine that most people would learn to use this technique very quickly.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

7. I found the technique very cumbersome to use.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

8. I felt very confident using the technique.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

9. I needed to learn a lot of things before I could get going with this technique.

   Strongly Disagree Disagree Neutral Agree Strongly Agree

10. I was physically comfortable using this technique.

    Strongly Disagree Disagree Neutral Agree Strongly Agree

A.1.3 Interview Questions

Post-Condition Interview Questions

1. What are your overall thoughts about the methods you tried?
2. What are the advantages of using this method to resolve occlusion and see the child?
3. What are the disadvantages of using this method to resolve occlusion and see the child?

Post-Study Interview Questions

Interview topics/questions:

1. In all the methods you tried, did you prefer making the content transparent or moving it? Why?
2. Did you prefer it when the system automatically moved the content or made it transparent, or when you did it manually? Why?
3. Did you prefer it when the system notified you about the position of the kid or when it did not? Why?
4. Do you believe that augmented reality will replace our mobile phones in the future?
5. What do you think about using these methods to access information in AR?
    a. Did AR feel intuitive?
    b. Would you use it? Why?
6. Are there any other comments you would like to share with us?

A.2 IRB Approved Informed Consent Form

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Consent to Take Part in a Research Study

Title of research study: IRB # 19-706 : Evaluating Methods for Dealing with Occlusion of Important Real World Objects by Virtual Glanceable Content in Always-on AR Interfaces

Principal Investigator: Dr. Doug A. Bowman Contact: (540) 231-2058 **or** dbowman@vt.edu

Other study contact(s): Shakiba Davari Contact: (650) 223-4063 or sdavari@vt.edu

Key Information: The following is a short summary of this study to help you decide whether or not you want to be part of this study. More detailed information is listed later on in this form.

Why am I being invited to take part in a research study?

We invite you to take part in this research study because you are an English speaker of age 18 years or older.

What should I know about being in a research study?

- Someone will explain this research study to you
- Decision to participate in the study is solely yours to make
- You can choose not to take part
- You can agree to take part and later change your mind
- Your decision will not be held against you
- You can ask all the questions you want before you decide

What should I know about being in <u>this</u> research study?

- This research is intended to evaluate a variety of methods for interaction with digital content in an augmented reality (AR) system. The results of this research helps us understand the best ways to interact with digital data and information in AR while maintaining awareness of the real world around us and produce guidelines for AR user interface design.
- You will be asked to sit on a chair, wear a VR headset and interact with the application we designed. Afterwards, you will fill out questionnaires about your experience. More detailed information about the study procedures can be found under "What happens if I say yes, I want to be in this research?"
- There is a risk that you might become nauseous or dizzy. More detailed information about the risks of this study can be found under "Is there any way being in this study could be bad for me? (Detailed Risks)"
- Your participation in this study will not guarantee any benefit for you or others. However, you will get the chance to interact with state-of-the-art virtual reality equipment and experience futuristic methods of interacting with digital content. Also, possible benefits to others include produce guidelines for AR user interface design.
- Decision for participation is this study is solely yours. If you are a student, the decision whether to participate or not participate will have no effect on your grades or relationship with Virginia Tech.

Consent to Take Part in a Research Study

Detailed Information: The following is more detailed information about this study in addition to the information listed above.

Who can I talk to?

If you have questions, concerns, complaints, or think the research has hurt you, talk to the research team at *540-231-2058 or dbowman@vt.edu*

This research has been reviewed and approved by the Virginia Tech Institutional Review Board (IRB). You may communicate with them at 540-231-3732 or irb@vt.edu if:

- You have questions about your rights as a research subject
- Your questions, concerns, or complaints are not being answered by the research team
- You cannot reach the research team
- You want to talk to someone besides the research team to provide feedback about this research

How many people will be studied?

We plan to include about 64 people in this research study.

What happens if I say yes, I want to be in this research?

You will become a participant in this research. You will be invited to try out the AR/VR system we developed, where your data will be collected to help us understand the best ways to interact with digital data and information in AR while maintaining awareness of the real world. The study will take place in the Sandbox Studio (Room 160, Moss Arts Center, 190 Alumni Mall, Blacksburg, VA 24060), and will take approximately 75 minutes for each participant.

What are the experimental procedures?

Upon signing up, the participants will be asked if they have participated in previous related studies. When participants arrive, they will be greeted and asked to read and sign the informed consent form after they feel their questions (if any) are thoroughly answered. Next, they will answer to our pre-questionnaire. Then the participants will be provided with written or verbal instructions for the experiment and familiarized with the lab and the equipment they will be using. Participants will wear a Lenovo Mirage Solo Virtual reality (VR) headset. Using these devices, participants will then go through eight sets of trials (one for each method we designed) to complete a series of tasks in VR, each taking about 7 minutes. The user will be sitting on a chair during the experiment and the tasks will involve physical movements like looking around the environment and pointing at virtual objects. Before starting each set of trials, the user will be trained on how the method works and learn how to interact with it. After each set, participants will fill out a questionnaire about the method they just used and answer some brief interview questions. Participants can stop to take a rest at any point between trials. After all tasks are completed, participants will be interviewed one final time. The experiment will be screen recorded for further interpretations and analysis of user behavioral patterns.

Consent to Take Part in a Research Study

What happens if I say yes, but I change my mind later?

You can leave the research at any time, for any reason, and it will not be held against you. All the data collected to the point of withdrawal will be properly discarded and destroyed immediately.

Is there any way being in this study could be bad for me? (Detailed Risks)

Using VR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache, or disorientation.

What happens to the information collected for the research?

We will make every effort to limit the use and disclosure of your personal information, including research study data and screen recordings, only to people who have a need to review this information. We cannot promise complete confidentiality. Organizations that may inspect and copy your information include the IRB, Human Research Protection Program, and other authorized representatives of Virginia Tech.

If identifiers are removed from your private information or samples that are collected during this research, that information or those samples could be used for future research studies or distributed to another investigator for future research studies without your additional informed consent.

The results of this research study may be presented in summary form at conferences, in presentations, reports to the sponsor, academic papers, and as part of a thesis/dissertation.

Can I be removed from the research without my OK?

Unless if a participant feels discomfort and cyber sickness, they will not be removed from the study without their consent.

What else do I need to know?

This research is being funded by *Google Inc.*

We will not offer to share your individual test results with you.

Your information and samples (both identifiable and de-identified) might be used to create products or to deliver services, including some that may be sold and/or make money for others. If this happens, there are no plans to tell you, or to pay you, or to give any compensation to you or your family.

There are no significant benefits that individual subjects might experience from participating in this research.

Consent to Take Part in a Research Study

Title of research study: IRB # 19-706 : Evaluating Methods for Dealing with Occlusion of Important Real World Objects by Virtual Glanceable Content in Always-on AR Interfaces

Principal Investigator: Dr. Doug A. Bowman Contact: (540) 231-2058 **or** dbowman@vt.edu

Other study contact(s): Shakiba Davari Contact: (650) 223-4063 or sdavari@vt.edu

Signature Block for Capable Adult

Your signature documents your permission to take part in this research and your confirmation that you can speak in English and are at least 18 years old. We will provide you with a signed copy of this form for your records.

Signature of subject	Date
Printed name of subject	
Signature of person obtaining consent	Date
Printed name of person obtaining consent	

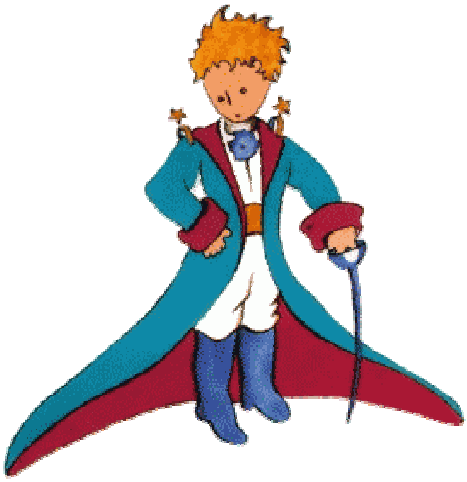
Appendix B

Experiment 2: AR Information Acquisition in Social and Solo Context

B.1 Experimental Design

B.1.1 Reading Material for the Solo Context

In the solo context, a printed copy of the first 10 pages of the book "The Little Prince" was provided for reading.



The Little Prince

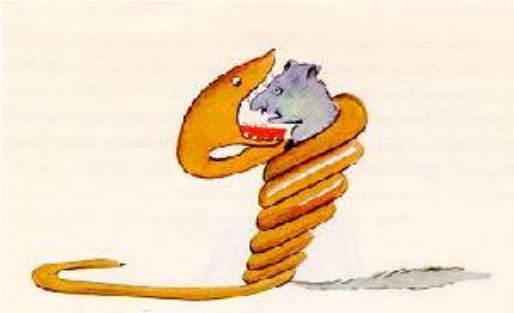
written and illustrated by
Antoine de Saint Exupéry

translated from the French by Katherine Woods

1

1

Once when I was six years old I saw a magnificent picture in a book, called *True Stories from Nature*, about the primeval forest. It was a picture of a boa constrictor in the act of swallowing an animal. Here is a copy of the drawing.



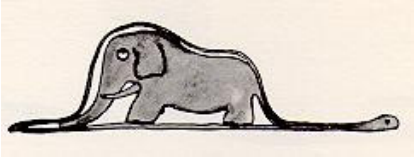
In the book it said: "Boa constrictors swallow their prey whole, without chewing it. After that they are not able to move, and they sleep through the six months that they need for digestion."

I pondered deeply, then, over the adventures of the jungle. And after some work with a colored pencil I succeeded in making my first drawing. My Drawing Number One. It looked something like this:

I showed my masterpiece to the grown-ups, and asked them whether the drawing frightened them.

But they answered: "Frighten? Why should any one be frightened by a hat?"

My drawing was not a picture of a hat. It was a picture of a boa constrictor digesting an elephant. But since the grown-ups were not able to understand it, I made another drawing: I drew the inside of a boa constrictor, so that the grown-ups could see it clearly. They always need to have things explained. My Drawing Number Two looked like this:



The grown-ups' response, this time, was to advise me to lay aside my drawings of boa constrictors, whether from the inside or the outside, and devote myself instead to geography, history, arithmetic, and grammar. That is why, at the age of six, I gave up what might have been a magnificent career as a painter. I had been disheartened by the failure of my Drawing Number One and my Drawing Number Two. Grown-ups never

B.1.2 Conversation Material for the Social Context

Verbal Conversations

Once the session starts, we will continue our conversation about a specific topic. To keep you engaged in the conversation, I will ask you some questions throughout. I'd appreciate it if you avoided short answers and expanded on them as much as possible. Also, please keep in mind that you do not have to answer these questions truthfully, so please feel free to fabricate information during these conversations.

Topic 1: Hike:

What do you usually do in your free time? Do you like hiking as well?

There are a lot of good hikes around here. I have been to McAfee, tinker cliff, cascades fall, dragon tooth, …. Have you been to any one of them? Which ones?

Have you been to cascades falls? I went there a few times. The first time it was early morning in the summer. And since it is a short hike it was perfect for that start-of-the-day hike. I would definitely recommend it. Have you ever touched the water closer to the Falls?

If you have not already, I'd suggest next time you go there touch the water. Last time I was there the water level was low and it was summer, so I took off my shoes and walked in the water. It was so cold but also refreshing. You should try it once if you like these kinds of experiences. Word of advice though, do not stay in that water for long. It is usually very cold. Last time I did it, I stayed in water for 5 minutes and my feet started to hurt. Also, be extremely cautious if you do it, since it is slippery.

Do you have any interesting stories or memories from a hike? I have been to Tinker cliffs, only once. And that once time I was with a group of friends who don't hike that frequently. They got tired really quickly and started asking: "when are we gonna arrive? Are we there yet?" That did not help the group. Just made the hike feel way longer than it actually was. But still in the end, the view was amazing. It was definitely worth it.

Topic 2: Restaurants around Blacksburg:

How long have you been in Blacksburg? Have you had some time to explore different restaurants in town?

I have been here for 3 years now and have been experimenting a little bit with the different restaurants myself. What kind of food do you like?

I personally love a little bit of everything. When it comes to food it is difficult to choose one. I like Japanese, Mexican…. Speaking of, actually, a new Japanese restaurant has opened recently in

downtown: Tokyo Grill. I'm really curious to know how their food is, specially that it is close to campus and my lab. If the food is good, I can go there for lunch sometimes. Have you ever been there?

What are your favorite restaurants around Blacksburg?

Yeah those are all good ones. I personally love the wine lab, and Palisades. I mean, Palisades is a little outside of town, but still, I'd say close enough. Around campus, Maroon door is pretty good too. They opened a couple of weeks ago. Have you been to any of them?

Topic 3: Weather

Where are you from? How is the weather there? Does it snow frequently there?

Yeah I was actually in Canada for a couple of years and it was too cold. I have a lot of stories about the cold weather there . This one time, my friend was going to Tim Hortons (I mean it's Canada, it must be Tim Horton's). Anyways I asked him to get me a coffee too, and this particular store is like 3 minutes away. Long story short, he came back 5 minutes late with a cold coffee and his beard was frozen. Have you ever been to Canada?

Do you have any fun stories about extreme weather conditions? How long have you been here in Blacksburg? And how do you like the weather here?

Yeah, the first time I came here it was mid-March, And I was coming here from California, So I thought it would be warm and did not bring any warm clothes. It was too cold, and it snowed! Well good news is, I did not get a cold then.

B.2 Questionnaires and Interview

B.2.1 Post-Condition Questionnaire

Interface-Social

Projects Actions Contacts Library Help

Survey Actions Distributions Data & Analysis Reports

Interface-Social

iQ Score: Fair

Information Access

Block Options

Q1

When using the current interface in this conversation, on scale of 1 to 7:
How easy was it to access information from the applications?

1 (not easy)	2	3	4	5	6	7 (very easy)
○	○	○	○	○	○	○

Add Block

Social Intrusiveness

Block Options

Q2

When using the current interface in this conversation, on scale of 1 to 7:
How much was the interface intrusive to your awareness of the social cues happening around you?

1 (not intrusive)	2	3	4	5	6	7 (very intrusive)
○	○	○	○	○	○	○

Add Block

Usability

Block Options

Q3

This part consists of pairs of contrasting attributes that may apply to the interface.
The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by moving the slider to where it most closely reflects your impression.

1 (annoying)	2	3	4	5	6	7 (enjoyable)
○	○	○	○	○	○	○

Q4

1 (not understandable)	2	3	4	5	6	7 (very understandable)
○	○	○	○	○	○	○

Q5

1 (creative)	2	3	4	5	6	7 (dull)
○	○	○	○	○	○	○

Q6

1 (easy to learn)	2	3	4	5	6	7 (difficult to learn)
○	○	○	○	○	○	○

Q7

1 (valuable)	2	3	4	5	6	7 (inferior)
○	○	○	○	○	○	○

Q8

1 (boring)	2	3	4	5	6	7 (exciting)
○	○	○	○	○	○	○

Q9	1 (not interesting)	2	3	4	5	6	7 (very interesting)
Q10	1 (unpredictable)	2	3	4	5	6	7 (predictable)
Q11	1 (fast)	2	3	4	5	6	7 (slow)
Q12	1 (inventive)	2	3	4	5	6	7 (conventional)
Q13	1 (obstructive)	2	3	4	5	6	7 (supportive)

Q14

1 (good)	2	3	4	5	6	7 (bad)
○	○	○	○	○	○	○

Q15

1 (complicated)	2	3	4	5	6	7 (easy)
○	○	○	○	○	○	○

Q16

1 (unlikable)	2	3	4	5	6	7 (pleasing)
○	○	○	○	○	○	○

Q17

1 (usual)	2	3	4	5	6	7 (leading edge)
○	○	○	○	○	○	○

Q18

1 (unpleasant)	2	3	4	5	6	7 (pleasant)
○	○	○	○	○	○	○

Q19

1 (secure)	2	3	4	5	6	7 (not secure)
○	○	○	○	○	○	○

Q20

1 (motivating)	2	3	4	5	6	7 (demotivating)
○	○	○	○	○	○	○

Q21

1 (meet expectations)	2	3	4	5	6	7 (does not meet expectations)
○	○	○	○	○	○	○

Q22

1 (inefficient)	2	3	4	5	6	7 (efficient)
○	○	○	○	○	○	○

Q23

1 (clear)	2	3	4	5	6	7 (confusing)
○	○	○	○	○	○	○

Q24

1 (impractical)	2	3	4	5	6	7 (practical)
○	○	○	○	○	○	○

Q25

1 (organized)	2	3	4	5	6	7 (cluttered)
○	○	○	○	○	○	○

Q26

1 (attractive)	2	3	4	5	6	7 (unattractive)
○	○	○	○	○	○	○

Q27

1 (friendly)	2	3	4	5	6	7 (unfriendly)
○	○	○	○	○	○	○

Q28

1 (conservative)	2	3	4	5	6	7 (innovative)
○	○	○	○	○	○	○

Add Block

B.2.2 Interview Questions

A Study on the Effect of the Information Access Interface in Different Contexts

Interview topics/questions:

For the Solo Scenario:

1. Which one of the 3 interfaces did you prefer? Why?
   - ❑ Mobile Interface
   - ❑ Glanceable AR

   Regardless of which one you preferred:
2. What, if any, were the benefits of the AR compared to the smartphone?
3. What, if any, were the disadvantages of the AR compared to the smartphone?

For the Solo Scenario:

4. Which one of the 3 interfaces did you prefer? Why?
   - ❑ Mobile Interface
   - ❑ Glanceable AR
   - ❑ Intelligent AR
5. What, if any, were the benefits of the Intelligent AR compared to Glanceable AR?
6. What, if any, were the disadvantages of the Intelligent AR compared to Glanceable AR?
7. How do you feel the Intelligent AR affects your social interaction? And why? Please compare both the smartphone and glanceable AR.

8. What do you like about using AR for information access, compared to using a smartphone?
9. What do you like about using a smartphone for information access, compared to using AR?

B.3 IRB Approved Informed Consent Form

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Consent to Take Part in a Research Study

Title of research study: IRB #21-258: A Study on the Effect of the Information Access Interface in Different Contexts

Principal Investigator: Dr. Doug A. Bowman Contact: (540) 231-2058 **or** dbowman@vt.edu

Other study contact(s): Shakiba Davari Contact: (650) 223-4063 or sdavari@vt.edu

Key Information: The following is a short summary of this study to help you decide whether or not you want to be part of this study. More detailed information is listed later on in this form.

Why am I being invited to take part in a research study?

We invite you to take part in this research study because you are an English speaker of age 18 years or older.

What should I know about being in a research study?

- Someone will explain this research study to you.
- Decision to participate in the study is solely yours to make.
- You can choose not to take part.
- You can agree to take part and later change your mind.
- Your decision will not be held against you.
- You can ask all the questions you want before you decide.

What should I know about being in <u>this</u> research study?

- The study takes up to 75 minutes and aims to learn about the effect of the information access interface (mobile phone vs AR) in different contexts. The results of this research help us understand the best ways to interact with digital data and information in AR in different contexts and produce guidelines for AR user interface design.
- You will be asked to sit on a chair and go through 5 different sessions where you either read an article or participate in a conversation with the investigator. During each session, you will use a mobile phone or wear an AR headset to interact with the application we designed and access information from them. After each session, you will fill out questionnaires about your experience. You will also answer a few interview questions once the whole experiment is done. This study will be recorded for further analysis. More detailed information about the study procedures can be found under "What happens if I say yes, I want to be in this research?"
- There is a risk that you might become nauseous or dizzy. More detailed information about the risks of this study can be found under "Is there any way being in this study could be bad for me? (Detailed Risks)"
- Should you decide to participate in the study, you will be compensated $15 upon completion of the session. Your participation in this study will not guarantee any other benefit for you or others. However, you will get the chance to interact with state-of-the-art augmented reality equipment and experience futuristic methods of interacting with digital

Consent to Take Part in a Research Study

content. Also, possible benefits to others include producing guidelines for AR user interface design.

- Decision for participation in this study is solely yours. If you are a student, the decision whether to participate or not participate will have no effect on your grades or relationship with Virginia Tech.

Detailed Information: The following is more detailed information about this study in addition to the information listed above.

Inclusion Criteria:

The participant population for this study includes people who: 1) are at least 18 years old, 2) are English speakers, 3) have perfect or corrected normal vision (contact lenses are supported). Participants should not have any symptom or known recent exposure to COVID-19.

Who can I talk to?

If you have questions, concerns, complaints, or think the research has hurt you, talk to the research team at *540-231-2058 or dbowman@vt.edu*

This research has been reviewed and approved by the Virginia Tech Institutional Review Board (IRB). You may communicate with them at 540-231-3732 or irb@vt.edu if:

- You have questions about your rights as a research subject.
- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team to provide feedback about this research.

How many people will be studied?

We plan to include about 48 people in this research study.

What happens if I say yes, I want to be in this research?

You will become a participant in this research. You will be invited to try out the AR system we developed, where your data will be collected to help us understand the best ways to interact with digital data and information in AR in different contexts. The study will take place in the Sandbox Studio (Room 160, Moss Arts Center, 190 Alumni Mall, Blacksburg, VA 24060), and will take approximately 75 minutes for each participant.

What are the experimental procedures?

Upon signing up, the participants will be greeted and asked to read and sign the informed consent form after they feel their questions (if any) are thoroughly answered. For parts of the experiment, participants will wear a Microsoft HoloLens 1 Augmented Reality (AR) device. After that, participants will be asked to fill out a pre-study background questionnaire. The investigator then gives an introduction to the study and its objectives and introduces the lab and the equipment they will be using to the user. During the Study, the user will go through 5

Consent to Take Part in a Research Study

different phases. For each phase the user will be in one of the two solitary and social contexts where they either read an article or participate in a conversation with the investigator. During these phases they use a mobile phone or one of our two proposed AR interfaces to access information. Before each phase, the user will go through a 2-minute training session. The training session will be similar to but shorter than the actual session, to make sure that the participant knows their task and how to use the interface and interact with it. After each phase, the participant will be asked to answer a questionnaire. You will also answer a few interview questions once the whole experiment is done. The study will be recorded for further analysis.

What is the total duration of this study?

The duration of an individual subject's participation will be 75 minutes.

What happens if I say yes, but I change my mind later?

You can leave the research at any time, for any reason, and it will not be held against you. All the data collected to the point of withdrawal will be properly discarded and destroyed immediately. Please keep in mind that only the participants who complete the session will be compensated.

Is there any way being in this study could be bad for me? (Detailed Risks)

Using AR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache, or disorientation.

What happens to the information collected for the research?

We will make every effort to limit the use and disclosure of your personal information, including research study data and recordings, only to people who have a need to review this information. We cannot promise complete confidentiality. Organizations that may inspect and copy your information include the IRB, Human Research Protection Program, and other authorized representatives of Virginia Tech.

If identifiers are removed from your private information or samples that are collected during this research, that information or those samples could be used for future research studies or distributed to another investigator for future research studies without your additional informed consent.

The results of this research study may be presented in aggregation at conferences, in presentations, reports to the sponsor, academic papers, and as part of a thesis/dissertation.

Can I be removed from the research without my OK?

If the participants show COVID-19 symptoms or refuse to follow the COVID-19 safety protocols (e.g., wearing a mask) their participation will end, they will be thanked for their time, and any data will be removed from the study.

What else do I need to know?

Consent to Take Part in a Research Study

Participants who complete the session will be compensated $15 for their time.

We will not offer to share your individual test results with you.

Your information and samples (both identifiable and de-identified) might be used to create products or to deliver services, including some that may be sold and/or make money for others. If this happens, there are no plans to tell you, or to pay you, or to give any compensation to you or your family.

Title of research study: IRB #21-258: A Study on the Effect of the Information Access Interface in Different Contexts

Principal Investigator: Dr. Doug A. Bowman Contact: (540) 231-2058 **or** dbowman@vt.edu

Other study contact(s): Shakiba Davari Contact: (650) 223-4063 or sdavari@vt.edu

Signature Block for Capable Adult

Your signature documents your permission to take part in this research. We will provide you with a signed copy of this form for your records.

Signature of subject	Date
Printed name of subject	
Signature of person obtaining consent	Date
Printed name of person obtaining consent	

Appendix C

Experiment 3: AR Content Placement Strategy in Four Contexts

C.1 IRB Approved Informed Consent Form

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Consent to Participate in a Research Study

Title of research study: IRB #23-879: Exploring the Design of Context-Aware Augmented Reality Interfaces for Document Navigation

Principal Investigator: Dr. Doug A. Bowman Contact: (540) 231-2058 or dbowman@vt.edu

Other study contact(s): Shakiba Davari Contact: (65) 223-4063, mailto:sdavari@vt.edu

Key Information: The following is a short summary of this study to help you decide whether to be a part of this user study. More detailed information is listed in this form.

Why am I being invited to take part in a research study?
We invite you to take part in a research study about Augmented Reality (AR) user interfaces. The whole experiment will take roughly 120 minutes. Participation is entirely voluntary.

What should I know about being in a research study?

- The co-investigators will explain the research study and the procedure for you and provide answers to any questions or concerns you have.
- Whether or not you take part is completely up to you.
- You can ask all the questions you want before you decide.
- You can choose not to participate in the study.
- You can agree to take part and later change your mind.
- Your decision will not be held against you.

What should I know about this research study?

- The research study will take approximately 120 minutes to complete.
- This study explores various AR interfaces, and their effect on the user experience and performance, to access information from virtual documents in four different conditions.
- You will be asked to wear a Microsoft HoloLens 2 AR device and experience four AR user interfaces, in four different conditions, totaling to 16 parts. After each study part, you will be asked to fill out questionnaires about your experience (**More detailed information can be found under "What are the experimental procedures?"**).
- The study tasks involve using hand gestures to interact with virtual objects that are placed in the real world around you and have conversations with virtual avatars while seated or walking inside a room.
- Some people may experience dizziness or nausea as a side effect of AR headsets (**More detailed information can be found under "Is there any way being in this study could be bad for me? (Detailed Risks)").**
- We offer a $15 compensation to the participants who do not earn any Virginia Tech Ut-Prosim service credits for their participation in the study.
- If you are a student at Virginia Tech, the decision whether to participate or not will have no effect on your academic performance or relationship with Virginia Tech.

Consent to Participate in a Research Study

Detailed Information:

The following is more detailed information about this study in addition to the information listed above.

Inclusion Criteria:

Participation in this study is open to English speaker adults who are: 1) aged at least 18 years, 2) able to understand spoken and written English, and can carry on a conversation in English, 3) capable to stand up, walk and perform hand gestures independently, and 4) have (perfect or corrected) normal vision (eyeglasses are supported).

Who can I talk to?

If you have questions, or concerns regarding the research, please send an email to Shakiba Davari at mailto:sdavari@vt.edu or talk to the research team directly at Gilbert Place (Room 3212, 220 Gilbert Street, Blacksburg, VA 24060)

This research has been reviewed and approved by the Virginia Tech Institutional Review Board (IRB). You may communicate with them at 540-231-3732 or irb@vt.edu if:

- You have questions about your rights as a research subject.
- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team to provide feedback about this research.

How many people will be studied?

We plan to include about 70 people in this research study.

What happens if I say yes, I want to be in this research?

You will become a participant in this research. As a participant in this study, you will be invited to Gilbert Place (Room 3212, 220 Gilbert Street, Blacksburg, VA 24060) where you can try out and provide your feedback about various AR interfaces that we developed.

What are the experimental procedures?

- The user study session will start by welcoming the participant and an introduction to the study and its motivations.
- We will then provide you with enough time to read/review the consent form and ensure all your questions and concerns are addressed.
- Next, through a pre-study questionnaire, we will collect demographic information and prior experience with Augmented/Virtual Reality.
- We will, then, introduce the AR headset, our designed AR interfaces, the details of the study conditions and study tasks.
- This study explores four AR interfaces to access information from virtual documents in four conditions in an indoor area. Participants will be asked to wear a Microsoft HoloLens 2 device and repeat the study tasks in 16 parts:
- **Study Conditions:**
    - *Solo:* In two of the conditions, participants will be performing the study task alone, once standing (*Stationary*) and once moving around (*Mobile*).

Consent to Participate in a Research Study

- *Social:* In the other two, they will speak with virtual avatars and answer their questions, once standing (*Stationary*) and once moving around (*Mobile*).

- **Study Task:** the participant will hear a question (from the headset in the solo conditions and from the virtual avatars in the social conditions). They should then interact with the virtual content through hand gestures to open the related document and read the answer to the question from it.
- **AR Interfaces:** The four interfaces are different in how they represent the virtual documents to the user. More specifically, each interface displays documents at a different level of detail or places them differently in the real-world.
- Each part will start with a training phase and will be followed by a survey and interview questions about their experience.
- During the study, the participants' head position, orientation, gaze direction, and manual interactions with the interfaces and their respective timestamps will be recorded on the headset.
- You are allowed to remove the headset after the completion of the tasks and during the survey and interview phases.
- The whole experiment will take roughly 90 to 120 minutes (approximately 5 minutes per part).
- Participants who complete the session will receive $15 compensation for their time in cash at the end of the session.

What is the total duration of this study?

The duration of an individual subject's participation will be 90-120 minutes.

What happens if I say yes, but I change my mind later?

You can leave the research at any time, for any reason, and it will not be held against you.
All the data collected to the point of withdrawal will be properly discarded and destroyed immediately.

Is there any way being in this study could be bad for me? (Detailed Risks)

Using AR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache, or disorientation. During tasks involving physical movement, there is also some risk that you could collide with obstacles in the physical environment. Each participant will be accompanied by an experimenter alongside as a safety guarantee. Before the participant's arrival we will clean and sanitize the headset.

What happens to the information collected for the research?

Each participant's data will be linked by a randomly assigned participant ID. This ID will not link to any private or identifiable information. The data for this study will not be released outside the research study group. The results of this research study may be presented in aggregation at conferences, presentations, reports to the sponsor, academic papers, and as part of a thesis/dissertation.

Consent to Participate in a Research Study

We will make every effort to limit the use and disclosure of your personal information only to people who have a need to review this information. We cannot promise complete confidentiality. Organizations that may inspect and copy your information include the IRB, Human Research Protection Program, and other authorized representatives of Virginia Tech.
If identifiers are removed from your private information or samples that are collected during this research, that information or those samples could be used for future research studies or distributed to another investigator for future research studies without your additional informed consent.

What else do I need to know?

- Participants who complete the session will receive a $15 compensation for their time.
- We will not offer to share your individual test results with you.
- Your information and samples (both identifiable and de-identified) might be used to create products or to deliver services, including some that may be sold and/or make money for others. If this happens, there are no plans to tell you, or to pay you, or to give any compensation to you or your family.
- There are no significant benefits that individual subjects might experience from participating in this research. However, the subjects will get to experience novel Augmented Reality technology.

Consent to Participate in a Research Study

Title of research study: IRB #23-879: Exploring the Design of Context-Aware Augmented Reality Interfaces for Document Navigation

Principal Investigator: Dr. Doug A. Bowman Contact: (540) 231-2058 or dbowman@vt.edu

Other study contact(s): Shakiba Davari Contact: (65) 223-4063, mailto:sdavari@vt.edu

Signature Block for Capable Adult

Your signature documents your permission to take part in this research. We will provide you with a signed copy of this form for your records.

Signature of subject | Date

Printed name of subject

Signature of person obtaining consent | Date

Printed name of person obtaining consent

Appendix D

Experiment 4: User-Specified AR Adaptations in Eight Contexts

D.1 IRB Approved Informed Consent Form

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Consent to Take Part in a Research Study

Title of research study: ***IRB #23-409: Determining User Preferences for Augmented Reality Apps Based on Context***

Principal Investigator: *Dr. Doug A. Bowman - Contact: (540) 231-2058* ***or*** *dbowman@vt.edu*

Other study contact(s): *Daniel Stover - Contact: (703) 431-8190* ***or*** *danielstover@vt.edu*

Key Information:

The following is a short summary of this study to help you decide whether or not to be a part of this study. More detailed information is listed later on in this form.

We invite you to take part in a research study about Augmented/Virtual Reality (AR/VR) user interfaces. The research will take around 45-60 minutes to complete. The participation is entirely voluntary.

What should I know about being in a research study?

- Someone will explain this research study to you
- Whether or not you take part is up to you
- You can agree to take part now and later change your mind
- Your decision will not be held against you
- You can ask all the questions you want before you decide.

What should I know about this research study?

This research project is intended to compare a variety of methods for interacting with augmented reality (AR) systems. This research will help us understand user preferences for AR app placement depending on the user's context. You will be asked to wear a Microsoft HoloLens 2 AR headset and interact with several pieces of virtual content displayed in the headset while performing a real-world task. Afterwards, you will be asked to fill out questionnaires about your experience (More detailed information can be found under **"What are the experimental procedures?"**). The research will take around 45-60 minutes to complete. There are risks that you might feel dizziness or nausea (More detailed information can be found under **"Is there any way being in this study could be bad for me? (Detailed Risks)"**). There are no significant benefits that individual subjects might experience from participating in this research. We cannot promise any benefits to others from your taking part in this research. However, the subjects will get to experience novel Augmented/Virtual Reality technology. Potential benefits to others also include the AR/VR industry gaining a better understanding of usability issues of AR interfaces to support everyday uses. If you are a student at Virginia Tech, the decision

Consent to Take Part in a Research Study

whether to participate or not will have no effect on your academic performance or relationship with Virginia Tech.

Detailed Information:

The following is more detailed information about this study in addition to the information listed above.

Inclusion Criteria: The study population for this study includes people who: 1) are at least 18 years old, 2) are English speakers, 3) have perfect or corrected normal vision (contact lenses and glasses are supported).

Who can I talk to?

If you have questions, concerns, or complaints, or think the research has hurt you, please send an email to Daniel Stover at *danielstover@vt.edu* or talk to the research team directly at *Room 3215, Gilbert Place, 400 Old Turner St Ste 100, Blacksburg, VA 24060*

This research has been reviewed and approved by the Virginia Tech Institutional Review Board (IRB). You may communicate with them at 540-231-3732 or irb@vt.edu if:

- You have questions about your rights as a research subject
- Your questions, concerns, or complaints are not being answered by the research team
- You cannot reach the research team
- You want to talk to someone besides the research team to provide feedback about this research

How many people will be studied?

We plan to include about 50 people in this research study.

What happens if I say yes, I want to be in this research?

You will become a participant in this research and will sign up for a time slot. You will be invited to try out the AR system we developed, and give us your feedback about ways to interact with data and information in AR. The study will take place on the 3rd floor of Newman Library, 560 Drillfield Drive Blacksburg, VA 24061. The meeting location is in the lounge.

What are the experimental procedures?

The experiment can be divided into five steps. In the first steps, participants will be welcomed upon arrival. We will introduce briefly what our study is about, the motivations behind it to the participants, and explain to participants any question they might have. Second, participants will be asked to read and sign the consent form. Third, they will be asked to fill out a pre-study questionnaire to collect demographic information and prior experience with Augmented/Virtual Reality. Fourth, participants will be asked to use and configure AR apps while conducting both real and virtual tasks. During the study session, the AR system will record the participants' point of view, audio, task performance, as

Consent to Take Part in a Research Study

well as the placement of the AR apps. Fifth, participants will be verbally asked to answer post-study questionnaires, which will be written down. The whole experiment will take roughly 45 to 60 minutes.

What is the total duration of this study?

The duration of an individual subject's participation will be 45-60 minutes.

What happens if I say yes, but I change my mind later?

You can leave the research at any time, for any reason, and it will not be held against you.
All the data collected to the point of withdrawal will be properly discarded and destroyed immediately.

Is there any way being in this study could be bad for me? (Detailed Risks)

Using AR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache, or disorientation. During tasks involving physical movement, there is also some risk that you could collide with obstacles in the physical environment. Each participant will be accompanied by an experimenter alongside as a safety guarantee. To increase participant safety, we will make sure to sanitize the AR headset before and after each participant's study.

What happens to the information collected for the research?

We will make every effort to limit the use and disclosure of your personal information, including survey responses, diaries, and interview notes, only to people who have a need to review this information. We cannot promise complete confidentiality. Organizations that may inspect and copy your information include the IRB, Human Research Protection Program, and other authorized representatives of Virginia Tech.

If identifiers are removed from your private information or samples that are collected during this research, that information or those samples could be used for future research studies or distributed to another investigator for future research studies without your additional informed consent.

The results of this research study may be presented in summary form at conferences, in presentations, reports to the sponsor, academic papers, and as part of a thesis/dissertation.

What else do I need to know?

Your information and samples (both identifiable and de-identified) might be used to create products or to deliver services, including some that may be sold and/or make money for others. If this happens, there are no plans to tell you, or to pay you, or to give any compensation to you or your family.

There are no significant benefits that individual subjects might experience from participating in this research. However, the subjects will get to experience novel Augmented/Virtual Reality technology.

Signature Block for Capable Adult

Your signature documents that you have read and understand the information outlined in this document and your permission to take part in this research. Please keep a digital signed copy of this form for your records.

Consent to Take Part in a Research Study

Signature of subject	Date
Printed name of subject	
Signature of person obtaining consent	Date
Printed name of person obtaining consent	