BIRMINGHAM CITY
University

Exploring Distance and Movement as Contextual Factors for Augmented Reality Interaction Design

By

Becky Spittle

A thesis submitted to Birmingham City University
for the degree of DOCTOR OF PHILOSOPHY

Supervised by:
Maite Frutos Pascual
Ian Williams
Chris Creed

Human-Computer Interaction Research Group
Faculty of Computing, Engineering and the Built Environment
School of Computing and Digital Technology
Birmingham City University

May 2025

ABSTRACT

As Augmented Reality (AR) technologies edge closer to ubiquity, context-aware interaction becomes essential for creating more seamless and practical interactions across a range of applications and use cases. This prompts research to understand how contextual factors could be effectively defined and/or inferred by AR devices, and how such information could be harnessed to provide the most suitable interactions in real-time. With this as the core motivation of the thesis, this research explores how we can work towards providing adaptive interactions for AR. By synthesising and building on prior work, the thesis proposes how two proxemic dimensions, Distance and Movement, could be leveraged as contextual factors to facilitate more flexible and personalised AR experiences.

As Proxemics was first proposed as a a theory rooted in social science, and Proxemic Interaction for context awareness in ubiquitous computing applications, where users interact with people, objects and devices in the real world, existing definitions for the spatial considerations they introduce fail to fully encompass the potential of referencing Distance and Movement for interactions with virtual content. To address this gap, and demonstrate the value of leveraging Distance and Movement as contextual factors in AR, two literature reviews and three empirical studies are conducted, highlighting how spatial considerations influence the performance and usability of commonplace AR interaction techniques.

The first literature review (Peripheral-free Interaction in Augmented Reality Environments) presents research on spatial interaction, as well as previous classifications for capturing context in AR, before introducing and exploring commonplace, peripheral-free

input techniques offered by the built-in hardware within AR devices. This review provides a foundation for understanding the range of contextual factors that influence AR interactions. The second literature review (Interaction Techniques for Immersive Environments) narrows the scope to explore explicit, peripheral-free interaction. By quantifying and evaluating the application of input modalities across a range of immersive technologies, the review considers the types of displays, input methods, use cases, and tasks addressed in current research, highlighting areas that require further investigation.

Following this, three empirical user studies were conducted. Studies 1 (Impact of Technique on Augmented Reality Interaction) and 2 (Impact of Distance on Augmented Reality Interaction) highlight the appropriateness of freehand and gaze-based interaction methods for completing near-field and far-field selection tasks in seated environments, with Study 3 (User-Defined Locomotion) exploring the spatial positioning and movement approaches employed by users with different techniques in a room-scale scenario. These studies provide recommendations for adapting AR interaction techniques based on Distance and Movement.

The findings reveal key insights into user performance, experience, preferences and behaviours, for interactions with virtual content placed across all four proxemic zones (Intimate, Personal, Social and Public). Although prior work has explored proxemic factors for AR interaction, no work has been found to directly consider how Distance and Movement could be referenced as contextual factors for adapting explicit AR input methods, and there are currently no recommendations for designing adaptive AR systems based on this. Results suggest that, by referencing Distance and Movement, AR interfaces could dynamically respond to improve usability in changing interaction scenarios. This thesis therefore begins to explore how proxemic dimensions can provide an approach to adaptive interaction design in AR, guiding future research and development towards more flexible, context-aware AR experiences.

ACKNOWLEDGMENTS

First, I owe a *HUGE* thank you to my supervisory team — Maite, Ian and Chris — who taught me even before I began this PhD journey, and saw the potential for Doctor Becky before I did. Maite, thank you for all your efforts in keeping me going from start to finish. I really appreciate your patience, insight, feedback and advice, both personally and academically, and the commitment it must have taken to navigate my Beckyisms (I don't think I'll hear the word "pragmatic" again without thinking of you!). Thank you all for your guidance, support (and, when needed, your brutal honesty), which made all the difference in getting me to the end.

I couldn't have submitted this thesis without the constant support of my Mom and Nan. Mom, thank you for talking me through the near-breakdowns and mundane tasks, and for always reminding me of the bigger picture when things felt overwhelming. Nan, thank you for always pushing me to do my best and being my biggest source of encouragement. I honestly doubt I'd have made it through undergrad without your words of love and wisdom, let alone completed a doctorate. I hope I've made you both proud.

To my fellow PhD-ers — especially Christina, Mattia, Haitham and Craig — thank you for always being there to offer your advice and support, listen when I needed to offload, and for sharing the occasional crisis. It was comforting to know that we all had our own struggles, and you made the PhD journey far more bearable just by being great people to complain to (and with). I appreciate you all and wish you every success!

I'm massively grateful to Microsoft for recognising the value in both me and my research, and for endorsing my work through the MSR PhD Fellowship. To the team at MSR Cambridge, some of whom are now very close friends, thank you for your guidance, insightful conversations, and ongoing support. I'll forever have memories of our chats

and get togethers, and here's to many more!

Many thanks also go to the students and staff at BCU who took part in my studies. Your time and participation turned questions into data and data into (what I like to think are) some interesting findings. Lots of appreciation also goes to my annual panel reviewers — Carlo, Waldo and Tycho — who year after year have offered constructive advice and boosted my confidence. Also, thank you to the reviewers of my papers and thesis. Your feedback has been invaluable, and continues to help shape and improve my research.

Finally, thank you to the (soon to be) Games Academy for taking me on as a lecturer mid write-up and showing saint-level patience throughout. Your support has meant a lot.

I can't thank you all enough for your help and guidance to get this PhD completed; now, on to the next!

Disclaimer: Apologies for any leftover Beckyisms[1]. Despite everyone's best efforts, I'm sure a few are still there somewhere.

[1] **Beck · y · ism** *noun.* An annoying thing Becky does—e.g. writing sentences backwards. — *term coined by Ian Williams*

"The inner machinations

of my mind are an enigma."

— Patrick Star, SpongeBob SquarePants

Contents

Page

1 Introduction **1**

1.1 Motivation . . . 3

1.2 Research Questions . . . 6

1.3 Thesis Structure . . . 9

1.4 Impact and Contribution . . . 11

1.5 Summary . . . 12

2 Peripheral-free Interaction in Augmented Reality Environments **14**

2.1 Introduction . . . 15

2.2 Spatial Interaction . . . 17

2.2.1 Proxemic Interaction . . . 18

2.3 Context Awareness in AR . . . 23

2.3.1 Pervasive AR . . . 23

2.3.2 Intelligent AR . . . 24

2.3.3 Human I/O . . . 25

2.3.4 Proxemic Dimensions for Context-Awareness in AR . . . 26

2.3.5 Transferable Interactions . . . 27

2.4 Interaction Techniques . . . 28

2.4.1 Speech Interaction . . . 29

2.4.2 Freehand Interaction . . . 31

2.4.3 Head and Eye Interaction . . . 36

2.5 Summary . . . 42

3 Interaction Techniques for Immersive Environments 44

3.1 Introduction . . . 45

3.2 Method . . . 46

3.2.1 Data Collection . . . 47

3.2.2 Data Analysis . . . 51

3.3 Analysis . . . 53

3.3.1 Top-Level Review . . . 53

3.3.2 Handheld Display . . . 57

3.3.3 Headworn Display . . . 61

3.3.4 Multiple Display Types . . . 66

3.4 Conclusions and Recommendations . . . 69

3.5 Limitations . . . 77

3.6 Summary . . . 78

3.7 Implications for Empirical Studies . . . 78

4 Impact of Technique on Augmented Reality Interaction (Study 1) 81

4.1 Introduction . . . 82

4.2 Background Research . . . 83

4.2.1 Extended Workspaces . . . 84

4.2.2 Interaction on the go . . . 86

4.3 Comparing Techniques for Near-Field Selection in Seated AR . . . 88

4.3.1 Method . . . 89

4.3.2 Results . . . 95

4.3.3 Discussion . . . 99

4.4 Summary . . . 102

5 Impact of Distance on Augmented Reality Interaction (Study 2) 103

5.1 Introduction . . . 104

5.2 Background Research . . . 105

5.2.1 Home Environments . . . 106

5.2.2 Interaction on the go . . . 107
5.2.3 Spectator Experiences . . . 109
5.3 Comparing Techniques for Far-Field Selection in Seated AR . . . 110
5.3.1 Method . . . 111
5.3.2 Results . . . 117
5.3.3 Discussion . . . 122
5.3.4 Summary . . . 127

6 User-Defined Locomotion - Distance and Movement (Study 3) 128
6.1 Introduction . . . 129
6.2 Background Research . . . 131
6.2.1 Implicit Interaction . . . 131
6.2.2 Explicit Interaction . . . 134
6.2.3 Summary . . . 137
6.3 User-Defined Distance and Movement Approaches . . . 137
6.3.1 Method . . . 138
6.3.2 Results . . . 145
6.4 Discussion . . . 164
6.5 Summary . . . 169

7 Discussion 171
7.1 Fulfillment of Research Questions . . . 172
7.1.1 RQ1: Gaps in Current Research . . . 172
7.1.2 RQ2: Seated Near-field Selection . . . 173
7.1.3 RQ3: Seated Far-field Selection . . . 174
7.1.4 RQ4: User-Defined Distance and Movement . . . 175
7.2 Distance and Movement . . . 176
7.2.1 Distance . . . 176
7.2.2 Movement . . . 179
7.3 Orientation, Identity and Location . . . 183

7.3.1 Orientation . . . 183
7.3.2 Identity . . . 186
7.3.3 Location . . . 188
7.4 Limitations . . . 189
7.4.1 Users . . . 190
7.4.2 Techniques . . . 190
7.4.3 Technology . . . 191
7.4.4 Task . . . 191
7.4.5 Environment . . . 192
7.4.6 Adaptation Possibilities . . . 192
7.4.7 Qualitative Insights . . . 193
7.5 Summary . . . 193

8 Conclusions and Future Work **195**
8.1 Findings and Contributions . . . 195
8.2 Future Work . . . 197
8.3 Closing Statement . . . 201

References **203**

List of Figures

1.1 Teaser figure for PhD motivation and potential proliferation of AR technologies for everyday interactions 3
1.2 Logical Progression of the thesis chapters 11

2.1 Context as defined by Pervasive AR (PAR) 23
2.2 Context as defined by Intelligent AR (iAR) 25
2.3 Detecting situational impairments by assessing the availability of human input/output channels (Human I/O) 25
2.4 Variants of hand poses observed in a pioneering gesture guessability study 31
2.5 Interaction steps to complete a docking task with the Airtap technique . 33
2.6 User interacting with a menu using hand pointing (Hover) 34
2.7 Worlds In Miniature: Actual sized objects in the room are interacted with by directly selecting and moving the miniature representations of the objects 35
2.8 ARtention: How gaze could provide a way to implicitly navigate between various information levels of a complex 3D visualisation 37
2.9 User interacting via head pointing and Airtap for making selections in an Internet of Things scenario 39
2.10 Consuming information and selecting information with gaze interaction . 41
2.11 Fundamental selection task exploring eye and head gaze 42

3.1 Distribution of data for the 15 papers considering solely handheld displays 58
3.2 Distribution of data for the 72 papers considering solely headworn displays 62
3.3 Distribution of data for the 15 papers considering multiple displays . . . 66

4.1 ARWin desktop showing a range of seated applications 85

4.2 InteractionAdapt: providing direct or indirect input techniques in different seated contexts . . . 85
4.3 Applications considered in an interactive Glanceable AR system . . . 86
4.4 Potential display layouts for interaction in an aeroplane environment . . 87
4.5 Participant interacting in a coffee shop . . . 88
4.6 Example of the environment during the observe stage and interact stage . 90
4.7 Interaction paradigms explored in the study . . . 91
4.8 Boxplot showing distribution of selection times . . . 96
4.9 Boxplots showing distribution of NASA-TLX subscale scores . . . 97
4.10 UEQ Subscale Scores . . . 98

5.1 Environment mock up for an AR application to control serveral smart home devices . . . 106
5.2 User interacting with a gaze and gesture interface that provides more discreet and socially acceptable interactions for public environments . . . 107
5.3 Examples of AR workspace layouts proposed by participants across four transport environments . . . 108
5.4 Examples of potential in-car AR games . . . 108
5.5 ARSpectator: example use case showing 3D structures of the stadium as well as line markings overlaid on the pitch . . . 109
5.6 Interaction paradigms explored in the study . . . 112
5.7 Boxplots showing distribution of selection times . . . 118
5.8 Boxplots showing distribution of NASA-TLX subscale scores . . . 120
5.9 UEQ Subscale Scores . . . 121

6.1 Smart factory maintenance scenario showing different levels of scaled information . . . 132
6.2 Participant respecting the personal space of an agent while asking for directions . . . 133

6.3 Adaptive Augmented Reality Workspace in use, where the virtual panels adjust when users walk and/or approach walls 134
6.4 Explicit activation of virtual content whilst walking 135
6.5 Participant performing target acquisition (grasping gesture) to perform object selections while walking 136
6.6 Walk the Line: application that leverages lateral shifts of the walking path as an input modality 136
6.7 Environment and Techniques 139
6.8 Boxplots showing distribution of task completion times 146
6.9 Boxplots showing distribution of participants' distance to target 148
6.10 Movement trajectories and end positions 150
6.11 Percentage of users stationary or walking during the entire trial 151
6.12 Walking patterns showcased by users across the entire trial 153
6.13 Walking patterns showcased by users in the last 5 seconds 154
6.14 Walking patterns showcased by users in the last 3 seconds 155
6.15 Walking patterns showcased by users in the last second 156
6.16 Locomotion speed showcased by users throughout the entire trial 157
6.17 Locomotion speed showcased by users in the last 5 seconds 158
6.18 Locomotion speed showcased by users in the last 3 seconds 159
6.19 Locomotion speed showcased by users in the last second 161
6.20 Boxplots showing distribution of NASA-TLX subscale scores 162
6.21 UEQ Subscale Scores 163

7.1 Appropriateness of interaction techniques in different proxemic zones . . 179

List of Tables

2.1 Five Dimensions of Proxemic Interaction: Distance, Orientation, Movement, Identity and Location 19

3.1 Search Terms: Query applied to the IEEE and ACM databases, where each row of the table represents 'AND' and each comma between search terms represents 'OR'. 47

3.2 Data categorisation approach: The factors assessed for the data analysis and their definitions. 49

3.3 Mapping the most appropriate inputs to distinct tasks on handheld displays: advantages and disadvantages. 71

3.4 Mapping the most appropriate inputs to distinct tasks on headworn displays: advantages and disadvantages. 72

4.1 Advantages and disadvantages of interaction techniques in the Intimate zone (<0.5 m). 101

5.1 Pairwise comparison results for selection times 118

5.2 Advantages and disadvantages of interaction techniques in the Personal (0.5m-1m), Social (1m-4m), and Public (>4m) zones. 126

6.1 Pairwise comparison results for selection times 147

6.2 Pairwise comparison results for distance from target 149

6.3 Percentage of data included in 'Last 5 Seconds' and 'Last 3 Seconds' time window analyses 152

6.4 Average walking speeds (Last 3 Seconds) 160

6.5 Average walking speeds (Last Second) 161

6.6 Advantages and disadvantages of interaction techniques in a room-scale environment. 169

Chapter One

Introduction

Contents

1.1 Motivation 3
1.2 Research Questions 6
1.3 Thesis Structure 9
1.4 Impact and Contribution 11
1.5 Summary 12

Immersive technologies encompass Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), and are collectively referred to as eXtended Reality (XR). Recent technical advances have driven unprecedented growth in both hardware and software capabilities, moving AR experiences from concept to a commercially viable medium (Gallardo et al., 2023; H. Bai, Sasikumar, et al., 2020; Uva et al., 2019). By merging physical and digital worlds, AR is redefining how users can engage with digital content in everyday contexts. It has emerged as a transformative technology poised to revolutionise daily interactions, and is projected to play a key role in shaping the future of human–computer interaction (HCI) (Xiaoan Liu et al., 2025; X. B. Liu et al., 2024; Davari, Stover, et al., 2024).

AR technologies introduce novel ways of communicating with computer-generated information (Aliprantis et al., 2019; Hertel et al., 2021). Unlike traditional 2D interfaces, AR enables tasks to be performed directly within real or virtual spatial contexts. Inter-

action extends beyond the sedentary nature of desktop environments to provide enriched, engaging 3D experiences that can support a wide range of activities and situations (Bach et al., 2018; Lages and Doug A. Bowman, 2019; Davari, Stover, et al., 2024).

Interaction is central to AR experiences, yet has proven more challenging to deliver effectively than in many other areas of HCI (Aliprantis et al., 2019). Because AR interfaces require new configurations of devices, techniques, and metaphors, they support a broader range of input and output modalities, creating a myriad of opportunities to design new interaction approaches (Laviola et al., 2017; Hertel et al., 2021; Gallardo et al., 2023).

These approaches often draw inspiration from human–human interaction, adopting modalities such as gaze, freehand, and speech (Muhammad Nizam et al., 2018; Davari, Stover, et al., 2024). This enables interactions that align with our natural senses and communication methods, and that can be adapted to different use cases based on human, environmental, and system factors (Grubert et al., 2017), thereby supporting the delivery of context-appropriate interactions in real time (X. B. Liu et al., 2024).

In HCI, context broadly encompasses any external factors that influence or relate to a user's interaction with an interface (Davari and Doug A Bowman, 2024). This includes characteristics that describe the user and their situation, such as behaviours, tasks, and preferences, the surrounding environment, including location and setting, and system properties such as sensing capabilities and device configuration (Seeliger, Weibel, and Feuerriegel, 2022; Grubert et al., 2017). By considering the range of factors encompassing context, and building on an extensive body of interaction research, designers and developers of AR technologies are empowered to create the most relevant interpretations of human-to-human interaction and apply this understanding to deliver more practical and intuitive interactions for AR environments (Jackson, 2020; X. B. Liu et al., 2024).

1.1 Motivation

In an era defined by rapid technological developments, we find ourselves on the verge of widespread AR adoption (Gallardo et al., 2023). Imagine a world where AR becomes an omnipresent assistant, used alongside or instead of current 2D display devices to empower users with an array of interaction possibilities. This could better support everyday tasks (X. B. Liu et al., 2024), and create more flexible, and immersive interactive experiences across a range of domains (Grubert et al., 2017; Seeliger, Weibel, and Feuerriegel, 2022). By harnessing peripheral-free interaction methods such as hand gestures, eye gaze, and head directionality, users will be capable of interfacing with technology in a wide range of situations and use cases (Hertel et al., 2021; X. B. Liu et al., 2024) (see Figure 1.1). This thesis invites you to envision such a future, where AR transcends the boundaries of ad-hoc developments, employed as and when they are needed, to provide an indispensable tool that revolutionises how we interact with technology and the world around us.

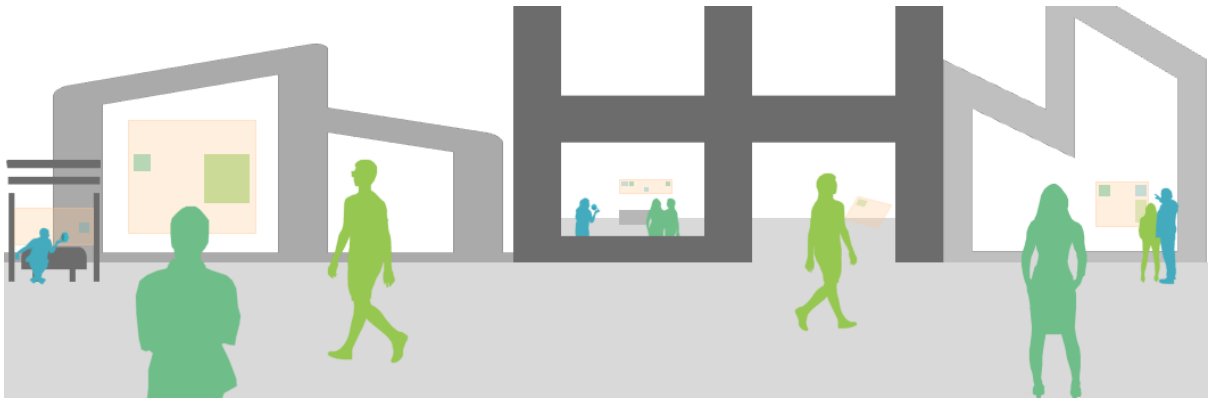

Figure 1.1: AR technologies are nearing widespread adoption, where they will soon be required to facilitate interactions in dynamic environments. Referencing spatial factors, such as distance and user movement, could help to provide adaptive interaction flows, where inputs such as hand gestures, gaze, and head directionality, could be employed interchangeably based on their affordances. By considering factors such as users distance from virtual content and their extent of motion, experiences could be better tailored to individual contexts.

The motivation behind this research stems from the need to establish a more intuitive approach to interaction design for AR technologies based on contextual factors. With the proliferation of AR-capable devices, such as Apple Vision pro and Meta Quest (Gallardo et al., 2023), research should be seeking to harmonise the interaction methods

used by individual users, across different platforms, applications and interaction scenarios. By doing so, we can work towards providing more seamless user experiences that fully harness the versatility and potential of AR technologies (X. B. Liu et al., 2024).

One of the most important contextual considerations when interfacing with AR technologies is how virtual elements are spatially distributed relative to the real world, which will impact how users 1) implicitly navigate an environment and 2) explicitly interact with interface components (Xiaoan Liu et al., 2025; Bach et al., 2018; Ballendat, Marquardt, and Greenberg, 2010). Due to the mobility provided by AR technologies, designing systems that understand and respond to user movements and the spatial relationships between people and interactive elements can be highly beneficial. For example, systems can offer the most appropriate interaction techniques and/or more relevant and timely content by understanding the context provided by distance, walking trajectories, walking speeds and spatial relationships (Davari and Doug A Bowman, 2024; Marquardt and Greenberg, 2015; Lages and Doug A. Bowman, 2019).

Further, by responding to the behaviours and preferences of individual users, systems could provide more accessible and intuitive communication methods, requiring less effort to interact with. This concept of adapting, activating or prompting interactions based on spatial factors has been considered by AR research (Grubert et al., 2017; Davari, Stover, et al., 2024), but is notably inspired by proxemics, a theory rooted in social science that studies how humans use space in communication and interaction. Integral to Proxemics is the definition of four proxemic zones, namely Intimate (less than 0.5 metres), Personal (0.5 to 1 metre), Social (1 to 4 metres), and Public (more than 4 metres) (Daza et al., 2021). Each zone characterises different interpretations of distance, where people adjust their position to match their social activities, and in response to the behaviours of other people and objects in their surroundings (Greenberg, Marquardt, et al., 2011).

Previous work in spatial computing suggests that by considering proxemic dimensions, Distance, Orientation, Movement, Identity and Location (Ballendat, Marquardt, and Greenberg, 2010; Greenberg, Marquardt, et al., 2011), systems can respond to user

behaviours and intentions (Marquardt and Greenberg, 2015). In AR environments, integrating proxemic factors can enhance the practicality of interactions by allowing systems to adapt to the user's activity and situation (X. B. Liu et al., 2024; Lages and Doug A. Bowman, 2019). The scope of this thesis therefore begins to explore the potential of referencing spatial factors, focusing on the influence of two proxemic dimensions on explicit AR interaction, considering the impact of Distance and Movement on the appropriateness of commonplace freehand and gaze-based techniques for fundamental selection tasks on head-worn displays (HWDs).

Exploring virtual content under immersive conditions often encourages users to stand and manipulate their bodies to navigate the real world and interact with virtual content (Bach et al., 2018). It is understood in human-human interaction that gaze and gestures are not only instinctive, intuitively employed as a fundamental component of communication, but that body orientation and movement is important to better understand and contextualise interactions (Scholl and McRoy, 2019). Although such insights have been applied to immersive environments, with interfaces now commonly using peripheral-free inputs (Hertel et al., 2021; Gallardo et al., 2023), AR researchers still face challenges in capturing the nuances of real-world communication when interacting with virtual content (X. B. Liu et al., 2024). Immersive technologies also offer opportunities that go beyond traditional human interaction by enabling advanced control and manipulation, e.g., employing gaze and freehand input for interacting with distant objects (Whitlock et al., 2018), allowing for interactions that are not possible in real-world scenarios (H. Bai, Sasikumar, et al., 2020; Muhammad Nizam et al., 2018).

When considering the input modalities that enable AR interactions, the inherent forms of communication that humans possess remain to be fully embraced (Munteanu et al., 2016; Davari, Stover, et al., 2024). As well as this, there is currently a lack of guidelines and support to help practitioners define the most appropriate interaction paradigms based on spatial factors (Frutos-Pascual, Creed, and I. Williams, 2019; X. B. Liu et al., 2024). As a result, input methods are often implemented subjectively, relying

on best judgement within the constraints of individual application scenarios, with little systematic consideration for how factors such as distance and movement could shape the appropriateness of interaction techniques (Seeliger, Weibel, and Feuerriegel, 2022). This risks constraining users to specific, fixed environments and fostering sub-optimal metaphors for fundamental tasks such as selection across changing interaction contexts (Piumsomboon, Clark, et al., 2013; Uva et al., 2019). The absence of shared guidelines also introduces broader complications and limitations. In particular, it limits interaction transferability, or how straightforward it is for users to learn input and interact across different applications and settings, making it difficult to effectively implement novel systems and provide practical solutions for long-term applications (Frutos-Pascual, Creed, and I. Williams, 2019; Grubert et al., 2017).

These challenges highlight the need for a deeper exploration into how AR interaction can respond to contextual factors. The thesis begins addressing this gap by considering how commonplace interaction techniques used for explicit input could be adapted based on factors surrounding 1) Distance: user distance from world-anchored virtual content, and 2) Movement: how users employ locomotion approaches relative to world-anchored virtual content (Grubert et al., 2017; Davari, Stover, et al., 2024).

1.2 Research Questions

This research proposes that the real potential posed by AR lies in its interaction possibilities: the fusion of modalities such as freehand, head pointing, and eye gaze, which could be employed interchangeably in different combinations as part of a continuous experience, to provide more practical, context-appropriate interactions (Grubert et al., 2017; X. B. Liu et al., 2024). To begin considering how key contextual factors can be understood by a system to adapt interaction techniques for explicit interaction, the below research questions (RQs) are addressed.

RQ1. What approaches exist for working towards context-aware AR, how have input methods previously been explored for AR interaction, and what are the gaps in the current research landscape?

A thorough exploration of existing AR research considering interaction is essential to identify prevailing gaps, trends and research trajectories. This is achieved by conducting two literature reviews in Chapters 2 and 3. The first is a narrative review which introduces the key concepts for the thesis, such as those surrounding peripheral-free input and the factors that support context-awareness in spatial interaction. Following this, a systematic review is presented, which focuses on quantifying and evaluating how peripheral-free input methods have been applied for explicit interactions across a spectrum of immersive technologies. Here, it is considered what display types, input methods, use cases, and tasks are being considered in current research. This analysis highlights which areas need further exploration and underscores the importance of spatial factors in defining the appropriateness of interaction techniques. This leads to informed recommendations for future work, providing a basis for the thesis and subsequent RQs.

RQ2. How effective are freehand and gaze-based techniques for near-field selection in seated AR environments?

Near-field interaction, where content is within arm's reach, is commonplace in AR interfaces (Hertel et al., 2021). This includes applications employed for extended workspaces (Cheng, Gebhardt, and Holz, 2023) and interaction on-the-go (Tung et al., 2015; Ng et al., 2021). RQ2 therefore aims to assess the suitability of fundamental freehand and gaze-based selection techniques in near-field seated scenarios. By conducting a user study with 32 participants, and evaluating performance metrics alongside qualitative feedback, the research contributes better understanding around the efficiency and user experience associated with each technique. The findings, which are presented in Chapter 4, provide valuable insights for designing interaction methods for near-field AR applications.

RQ3. How do freehand and gaze-based techniques perform for far-field selection across varying distances in seated AR environments?

Interacting with content beyond arm's reach introduces unique challenges that may affect user performance and preference (Xiaoan Liu et al., 2025). This includes reduced visual angle of targets, ambiguous pointing, and how to effectively implement paradigm adaptations (Whitlock et al., 2018; Cheng, Gebhardt, and Holz, 2023). Consequently, RQ3 explores how the appropriateness of freehand and gaze-based interaction techniques varies with distance, reporting on a user study with 32 participants, and evaluates the efficiency, usability and preference of each technique in a seated AR environment. The outcomes, which are presented in Chapter 5, offer recommendations for interaction design that work towards enhancing far-field AR interactions based on users' distance from virtual content.

RQ4. How do user-defined distances and locomotion approaches vary with different freehand and gaze techniques, and what influence does this have on their appropriateness in room-scale AR environments?

Understanding how users choose to approach and interact with objects of varying sizes from a standing position is crucial for designing effective interaction techniques. The final research question therefore investigates the impact of interaction technique and object size (small (5cm), medium (15cm), large (25cm)) on user locomotion behaviour, and is addressed through a user study with 40 participants. By analysing user-defined distance, position, locomotion approach and speed, selection time, task-load, user experience and preference, the advantages, limitations and affordances of different freehand and gaze-based techniques are evaluated. These findings, available in Chapter 6, inform recommendations for AR interaction design that accommodate user-object distance and users' locomotion approaches.

1.3 Thesis Structure

To provide clarity on how each component of this research contributes to addressing the research questions, this section outlines the structure of the thesis and maps chapters to their corresponding RQs.

Chapter 2: Peripheral-free Interaction in Augmented Reality Environments
This chapter lays the theoretical groundwork for the thesis by exploring existing literature on peripheral-free input and contextual factors in spatial interaction. This provides background for concepts surrounding interaction techniques, proxemic interaction, and how AR experiences can be adapted based on a range of human, environmental, and system factors. It addresses **RQ1** by highlighting how input methods have previously been explored for AR interaction, as well as comparing exisiting classifications and approaches proposed for inferring contextual information.

Chapter 3: Interaction Techniques for Immersive Environments
Building on the narrative review, this chapter provides a systematic evaluation of empirical research on explicit input methods used for immersive technologies. It further addresses **RQ1** by quantifying what display types, input methods, use cases, and tasks are being considered in current research, thereby identifying gaps that require further investigation and informing the empricial studies.

Chapter 4: Impact of Technique on Augmented Reality Interaction (Study 1)
This chapter reports on the first user study, which explores the appropriateness of freehand and gaze-based selection techniques for near-field interaction in seated AR environments. The study directly addresses **RQ2** by evaluating performance metrics and user experience in scenarios where content is within arm's reach.

Chapter 5: Impact of Distance on Augmented Reality Interaction (Study 2)
Extending the investigation to far-field interactions in Study 2, this chapter examines how the performance of freehand and gaze-based techniques varies across different distances

in seated AR setups. This work addresses **RQ3** by exploring the impact of increased distance (where virtual content is beyond arms reach) on interaction performance and user experience.

Chapter 6: User-Defined Distance and Locomotion (Study 3)

This chapter focuses on room-scale AR environments, investigating how users mediate their distances and employ locomotion when interacting with virtual objects from a standing position. It addresses **RQ4** by analysing user behaviour, considering their interaction strategies and movement trajectories, and how their approaches are impacted by the object size and interaction technique employed. This works towards uncovering how techniques could be adapted based on the affordances of freehand and gaze-based input when considering user-object distance and user movement.

Chapter 7: Discussion

In this chapter, the findings from the three empirical studies are synthesised to highlight the potential of harnessing user-object distance and user movement as contextual factors for adapting AR interactions. Recommendations are provided to aid practitioners in designing interactions that are responsive to distance and movement. It also reflects on the proxemic dimensions not directly examined in the empirical studies (Orientation, Identity, and Location), considering how these factors have been addressed in existing work and their value for advancing context-aware AR interaction. By envisioning how future systems might adapt to contextual factors, the discussion contributes towards enhancing user experience and interaction effectiveness in everyday, multi-purpose AR experiences (X. B. Liu et al., 2024; Grubert et al., 2017). Limitations are also highlighted to further contextualise the research conducted.

Chapter 8: Conclusions and Future Work

The final chapter summarises the contributions of the research and outlines directions for future work. It provides closing reflections on how the insights gained from addressing the four research questions can inform ongoing and future developments in context-aware AR

interaction. The structure of the thesis and connection between chapters is illustrated in Figure 1.2.

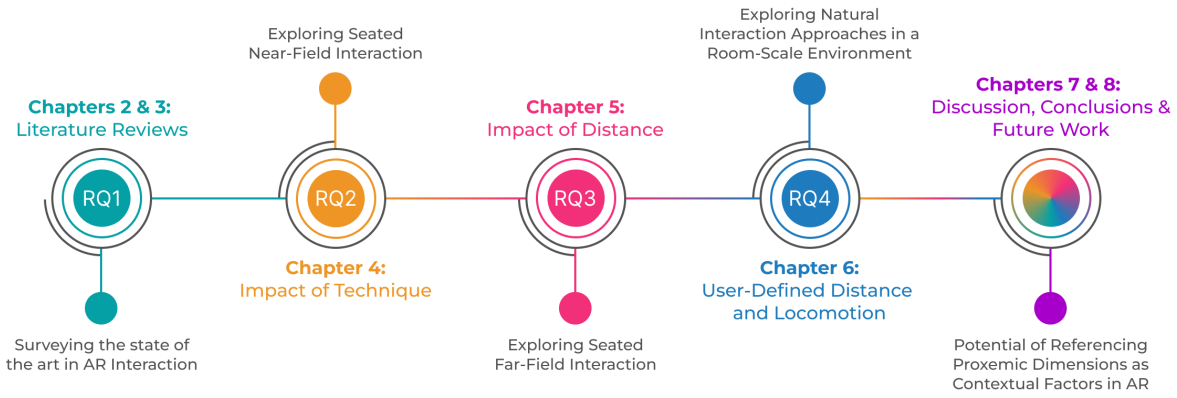

Figure 1.2: Logical Progression of the thesis chapters: Literature reviews (RQ1) inform the direction taken for user studies (RQ2-RQ4). Findings from these research questions are then synthesised, discussed and reflected upon to highlight the potential of referencing Proxemic Dimensions (notably Distance and Movement) for providing adaptive AR interactions.

1.4 Impact and Contribution

The work presented in this thesis has been, and continues to be, disseminated through high-impact, peer-reviewed venues:

- **Spittle, B.**, Frutos-Pascual, M., Creed, C. and Williams, I. (2022), "A Review of Interaction Techniques for Immersive Environments", in IEEE TVCG vol. 29, no. 9, pp. 3900-3921, 1 Sept. 2023. **[RQ1, Chapter 2]**
- **Spittle, B.**, Frutos-Pascual, M., Creed, C. and Williams, I. (2025). "Exploring the Impact of Distance on Extended Reality Selection Techniques", In Proceedings of the 27th International Conference on Multimodal Interaction (ICMI '25), October 13–17, 2025. **[RQ2 & RQ3, Chapters 4 & 5]**
- **Spittle, B.** (2021). "Maximising the Transferability of Interaction Techniques for Immersive Technologies". 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). Oct. 2021. [Doctoral Consortium]

Prospective Paper:

- **Spittle, B.**, Frutos-Pascual, M., Creed, C. and Williams, I. (2025). "Walk This Way: How Augmented Reality Interaction Techniques Influence User Movement and Spatial Positioning". Special Issue on Spatial Computing in the Journal of Behavior and Information Technology (under review) [**RQ4, Chapter 6**]

The work produced as part of this thesis not only advances theoretical understanding around how AR systems could adapt interaction methods based on contextual factors, but also provides practical insights that can help inform the next generation of AR interactions on consumer-level devices. Notably, the systematic literature review provides several recommendations for future work, providing an overview of the recent research landscape. Findings and design recommendations presented from the accepted/prospective papers, which report on the three user studies conducted, also offer valuable insights that can inform future research and applications - highlighting how commonplace AR techniques could be adapted based on distance and movement. Overall, the research works toward providing more flexible and personalised AR interactions.

1.5 Summary

This chapter introduced the scope and motivation of the thesis. It highlighted the potential for AR to become a pervasive technology, emphasising peripheral-free interaction as both a central opportunity and primary challenge for widespread adoption. The structure of the thesis was also outlined, showing how the literature reviews and empirical work interconnect, as well as their contributions and impact. Four research questions were defined, with RQ1 focused on identifying opportunities and gaps in the literature. Based on the literature reviews, three empirical studies were conducted to explore freehand and gaze-based input across near-field (RQ2), far-field (RQ3), and room-scale (RQ4) interaction. Together, these studies inform recommendations for adaptive AR interaction

design based on distance and movement. The next chapter begins by exploring the key concepts surrounding the thesis, focusing on peripheral-free input and factors relating to context-awareness in spatial interaction.

Chapter Two

Peripheral-free Interaction in Augmented Reality Environments

Contents

2.1 Introduction **15**

2.2 Spatial Interaction **17**

2.2.1 Proxemic Interaction 18

2.3 Context Awareness in AR **23**

2.3.1 Pervasive AR 23

2.3.2 Intelligent AR 24

2.3.3 Human I/O 25

2.3.4 Proxemic Dimensions for Context-Awareness in AR 26

2.3.5 Transferable Interactions 27

2.4 Interaction Techniques **28**

2.4.1 Speech Interaction 29

2.4.2 Freehand Interaction 31

2.4.3 Head and Eye Interaction 36

2.5 Summary **42**

2.1 Introduction

Peripheral-free interaction refers to interaction where all input is captured through the computing device itself, with no need for additional external hardware such as keyboards, mice, and controllers (Dritsas et al., 2025). At present, the position of peripheral-free interaction utilised for AR interfaces is comparable to that of graphical user interfaces (GUIs) in the early 1980s. Increasingly, technologies are being developed to further reduce the barriers of computing, empowering users by increasing functionality and interaction capabilities (X. B. Liu et al., 2024; Aliprantis et al., 2019).

GUIs have now primarily replaced text-based inputs, making computing more accessible and providing users with greater control. In a similar way, AR technologies offer enhanced interaction benefits (e.g., freehand, gaze, speech) that create even more flexible and intuitive methods for communicating with technology and the environment (Gallardo et al., 2023; Bach et al., 2018). Peripheral-free interaction employs inherent forms of human communication, providing opportunities for users to utilise applications with little or no training, with some technologies allowing for uninterrupted immersion and straightforward interface customisation to better suit individual and environmental needs (Lages and Doug A. Bowman, 2019; Davari, Stover, et al., 2024).

Peripheral-free AR input has aimed to expand on traditional interfaces, such as those provided on desktop and smartphone, to provide more seamless interactions where the technology becomes invisible to the user. Contemporary AR HWDs integrate sensing, display and computation, making them capable of leveraging real-time processing of the user's behaviours and surroundings. This reduces reliance on separate peripherals and external screens while enabling integrated interaction and feedback (Aliprantis et al., 2019; Gallardo et al., 2023). Such ambitions align with Weiser (1999) and his vision of ubiquitous computing, in which computation fades from focal attention, allowing users

to focus on their task and environment while computational processes and most interface elements recede to the periphery, surfacing only when needed (Weiser and Brown, 1997). Here, movement and positioning can also be treated as a means of peripheral-free communication, conveying activity, intent, and social meaning (E. T. Hall, 1966; Greenberg, Boring, et al., 2014).

Realising this concept of "invisible" computing depends on accurate, moment-to-moment inference of user state and situational context (Davari and Doug A Bowman, 2024). However, while AR technologies can recognise images, process language, and sense real-world stimuli, they still lack the capacity for continuous, reflective thought that distinguishes humans from machines. Unlike most current computing systems, humans employ efficient, context-guided thought processes to achieve goals (Jackson, 2020). Consequently, developing systems that correctly interpret and adapt to user behaviour and intent, particularly when considering the use of peripheral-free input, is challenging. This issue arises as humans can generate mental images and derive language representations from both concrete and abstract scenarios, whereas formal logic (as still employed by most computer systems) has largely been confined to probability estimation (Qi and Wu, 2019), and often struggles to capture the ambiguities and contradictions inherent in the broader spectrum of human actions, thoughts and intentions (Jackson, 2020; Davari, Stover, et al., 2024). This prompts research to understand how contextual factors could be effectively defined and/or inferred by AR systems, and how such information can be harnessed to deliver the most suitable interaction methods in real time.

Before quantifying research trends in Chapter 3, which investigates how factors such as input methods, tasks, and interaction scenarios have been addressed, this chapter begins to consider the key considerations of peripheral-free interaction and context-awareness in AR. First, research on spatial interaction is presented, as well as previous classifications for undertanding and capturing context in AR. Then, commonplace, peripheral-free input techniques; namely freehand, gaze-based and speech, are introduced and explored.

2.2 Spatial Interaction

When interacting with AR technologies, a key factor is the spatial positioning of virtual and physical elements relative to the user and their environment. This arrangement has been shown to influence how users choose to navigate the environment (via implicit input), as well as their deliberate (explicit) interactions with interface components and surrounding objects (Xiaoan Liu et al., 2025; Bach et al., 2018; Ballendat, Marquardt, and Greenberg, 2010). By gauging information such as user–object distance, user movement (or lack thereof), and their attention relative to entities in the interaction space, spatial context can be interpreted by a system to adapt AR interactions in real time (Grubert et al., 2017; Greenberg, Boring, et al., 2014; Cheng, Gebhardt, and Holz, 2023).

This thesis begins to consider how contextual information in an AR environment can be captured through *the Distance, Orientation, and Movement of Identities within a Location*. More broadly, these five dimensions have been shown to form the foundation of spatial interaction in ubiquitous computing, and could offer a structured approach for interpreting interaction context in AR (Ballendat, Marquardt, and Greenberg, 2010; Greenberg, Marquardt, et al., 2011). Together, they reflect how systems can sense, infer, and respond to the various factors that influence or relate to a user's interaction, which includes the people, objects and features in the real world (Xiaoan Liu et al., 2025; Davari and Doug A Bowman, 2024; Grubert et al., 2017).

Contextual factors encompass the characteristics of the user and their situation (e.g., behaviours, tasks, preferences), the surrounding environment (e.g., layout, setting, presence of others), and the properties of the device/system itself (e.g., sensing configuration, hardware constraints) (Seeliger, Weibel, and Feuerriegel, 2022; Grubert et al., 2017). Together, the five dimensions of proxemic interaction can support a rich, layered understanding of context, drawing from the multimodal input sources built in to AR devices, such as cameras, microphones, eye trackers, accelerometers, gyroscopes and personal/private databases, to interpret the spatial and semantic relationships between

entities across physical and virtual landscapes (Davari, Stover, et al., 2024).

While the empirical studies presented in the thesis specifically focus on single user interaction, considering 1) their distance from world-anchored virtual content and 2) their movement approaches relative to world-anchored virtual content, and how this shapes fundamental interaction on AR HWDs, the review that follows introduces all five dimensions to situate this focus within the broader literature, and highlight the relevance of proxemic dimensions to context-aware interaction design in AR.

2.2.1 Proxemic Interaction

The term "proxemics" was coined by anthropologist Hall (E. T. Hall, 1966) and identifies how people perceive, interpret, and utilise distance, posture, and orientation to mediate relations to other people, as well as fixed and moving/movable features in their environment (Greenberg, Marquardt, et al., 2011). Integral to Hall's theory are his definitions of four proxemic zones, namely Intimate (less than 0.5 metres), Personal (0.5 to 1 metre), Social (1 to 4 metres), and Public (more than 4 metres) (Daza et al., 2021). Each zone characterises different interpretations of distance, where people adjust their position to match their culture, social activities, and in response to the behaviours of other people and objects in their surroundings (Greenberg, Marquardt, et al., 2011).

Building on Proxemics as rooted in Human-Human Interaction, Proxemic Interaction was initially developed for ubiquitous computing applications, based on interactions with several physical devices. Ballendat, Marquardt, and Greenberg (2010) illustrated how proxemics can afford context-awareness in a system, which can be used to regulate implicit and explicit interactions, adapt system output and interactions when people and devices move through proxemic zones, capture users' directed attention to other people, devices and objects, and mediate collaborative interactions. They defined four dimensions for applying proxemics to ubiquitous computing; distance, orientation, movement and identity, stating that by capturing these four factors, a system would be able to

Table 2.1: Five Dimensions of Proxemic Interaction: Distance, Orientation, Movement, Identity and Location (Ballendat, Marquardt, and Greenberg, 2010; Greenberg, Marquardt, et al., 2011)

Dimension	**Definition**
Distance	The continuous or discrete (i.e. based on proxemic zones) measure of the position of an entity with respect to another entity.
Orientation	The absolute or relative measure that determines the direction an entity is facing with respect to another entity.
Movement	The changes in an entity's distance and orientation over time.
Identity	The identifiers that describe the entities in the space. This can be (1) detailed individual identities for each element in an environment or (2) categories such as real-world objects, virtual objects, and people.
Location	The physical context surrounding the entities that define a particular interaction space and its characteristics.

understand basic proxemic relationships. An additional dimension, location, was later defined by Greenberg, Marquardt, et al. (2011), to highlight the importance of the physical environment (i.e. the size of the interaction space and layout of entities within it) and the social context of the interaction (i.e. private or public, relaxed or formal) on interaction approaches. Based on this previous work (Ballendat, Marquardt, and Greenberg, 2010; Greenberg, Marquardt, et al., 2011), these five proxemic dimensions are defined in Table 2.1.

As Proxemic Interaction was originally proposed as a framework for context awareness in ubiquitous computing, its dimensions were framed around interactions between users and physical displays and devices situated in the real world (Ballendat, Marquardt, and Greenberg, 2010; Greenberg, Marquardt, et al., 2011). In such systems, distance thresholds, orientation cues, and identities could be detected, but typically through instrumented environments, external sensors and markers, and pre-assigned metadata (Marquardt, Diaz-Marino, et al., 2011). While effective for demonstrating the value of proxemic theories in computing, these approaches have rarely scaled beyond controlled, pre-defined settings (X. B. Liu et al., 2024). Conversely, AR devices consolidate many of these sensing capabilities into a single wearable platform. HWDs are capable of na-

tively tracking peripheral-free inputs, continuously mapping physical environments, and seamlessly integrating digital content (Dritsas et al., 2025). As a result, AR extends the capabilities of proxemic interaction, allowing dimensions to be applied not only to physical entities but also to virtual interface components and agents.

In AR, proxemic dimensions apply to a wide range of entities. These notably include *user–object* relations (e.g., walking around or manipulating a virtual model (Huang et al., 2022)), *user–user* relations (e.g., proxemic spacing and gaze during collaboration (Norouzi et al., 2019; Sloth et al., 2023)), and *object–object* relations (e.g., positioning digital content relative to physical surfaces (Cheng, Gebhardt, and Holz, 2023; McGill, Williamson, et al., 2019)). Each relation may involve physical entities, virtual entities, or a combination of both, creating complex interaction ecologies. AR therefore requires proxemic dimensions to be reinterpreted to account for these cross-reality relationships, with each dimension encapsulating one or more layers of context in AR:

Distance reflects the spatial proximity between entities, which will impact factors such as ergonomics, affordances, and the appropriateness of interaction modalities (Norouzi et al., 2019; Whitlock et al., 2018). For example, when considering user-object distance, direct freehand interaction may be more suitable for manipulating near-field virtual content (Xiaoan Liu et al., 2025), whereas pointing or raycasting techniques have commonly permitted interactions with distant content beyond arm's reach (Caputo, Bartolomioli, and Giachetti, 2023). Minimum and maximum interaction ranges with different input methods are also key considerations, defining what interactions are possible or practical at different distances with different modalities (W. Kim and Xiong, 2024; Whitlock et al., 2018).

Orientation refers to the absolute or relative facing direction of users, devices, physical objects, or virtual elements. AR HWDs are capable of natively tracking the front facing direction of objects as well as users' head directionality, eye gaze, and hand orientation, turning orientation into a fine-grained proxy for attention and intent (Pfeuffer, Abdrabou, et al., 2021). For instance, systems can display virtual information related

to the object currently in view (Davari, Stover, et al., 2024), or align shared content so that it is visible to multiple collaborators despite different viewpoints (Sloth et al., 2023). Orientation has also been shown to impact how users interpret the behaviour of avatars (Genay, Lecuyer, and Hachet, 2022), or the accessibility and affordances of objects in the scene (Symes, Ellis, and Tucker, 2007; Lin et al., 2023).

Movement describes how distance and orientation of entities change over time, capturing the dynamics of spatial interaction. When considering user movement, traditional ubicomp systems can notably detect entry, exit, or approach behaviours relative to features within an interaction space (Marquardt, Diaz-Marino, et al., 2011), but AR devices are capable of continuously tracking both macro-movements (e.g., walking toward a virtual object or navigating a physical room (Huang et al., 2022; Norouzi et al., 2019)) and micro-movements (e.g., subtle head tilts, body leans, and hand adjustments (Yu et al., 2019; Pham et al., 2018)) across different interaction spaces. These movement cues provide added insight into not only the users state but also their level of engagement and intent (Pfeuffer, Abdrabou, et al., 2021). For example, walking quickly towards or away from an entity may suggest approach or departure, while leaning towards it may suggest heightened interest. Movement has also been shown to support safety-critical adaptations, such as predicting walking paths to prevent collisions with furniture or other users (Hwang, Peli, and Jung, 2023).

Identity concerns the recognition of who or what is present in a scene. Whereas distance, orientation, and movement describe how entities relate and behave in space, identity helps describe what those entities are. For example, a "whiteboard" identity remains the same whether anchored on a wall or floating in mid-air, but its meaning in practice depends on its distance from the user and objects in the environment, how it is oriented, and how users move around it.

Additional metadata can also be defined or inferred to enrich individual identities with supplementary contextual information. For example, when considering users, this could include roles and/or permissions (e.g., presenter vs. participant (Sloth et al., 2023)),

behavioural histories (Davari, Stover, et al., 2024), or inference of biometric/physiological signals (e.g., gaze patterns, heart rate, facial expression (Bhalla et al., 2021)). For virtual objects, identity may encompass attributes such as size, colour, shape, or level of interactivity (Piumsomboon, Clark, et al., 2013). The richer the metadata, the deeper the system's contextual understanding, allowing experiences to be tailored to the properties of identites and the relationships between them (Davari and Doug A Bowman, 2024).

Location grounds identities within a meaningful "where", extending beyond simply understanding physical location (e.g., via gps co-ordinates) to also include semantic and social dimensions (Zhang et al., 2022; Singh et al., 2020). In HCI, location often encompasses absolute positioning (e.g., GPS, Wi-Fi triangulation) or relative zones (Singh et al., 2020; Marquardt, Diaz-Marino, et al., 2011). However, AR allows for more nuanced situating of identities based on what is detected throughout the environment (Xiaoan Liu et al., 2025). For exmaple, an AR system can automatically anchor a "virtual note" to a "desk" within a workspace (location) (Cheng, Gebhardt, and Holz, 2023; McGill, Kehoe, et al., 2020), or adapt system behaviour according to social etiquette (e.g., suppressing notifications while in a library or during a meeting (Davari, Stover, et al., 2024)).

Location also incorporates ambient factors such as time of day, lighting, or weather, linking spatial awareness and semantic inference with additional environmental measures (X. B. Liu et al., 2024; Dritsas et al., 2025). In this way, location describes a collection of identities based on their physical and social meanings, which could be referenced to deliver personalised interactions that are appropriate to the entire setting.

Together, the five proxemic dimensions demonstrate how spatial, behavioural, and semantic cues could provide a layered understanding of interaction context in AR. They define not only the inherent qualities and grounding of physical and virtual entities (identity, location) but also the dynamic and relational considerations as users and objects move through an interaction space (distance, orientation, movement). In this way, proxemic interaction offers a vocabulary for describing what contextual information can be captured and why it matters for adaptation (Marquardt, Diaz-Marino, et al., 2011). How-

ever, realising these possibilities in practice requires detailed frameworks that specify how such contextual data should be sensed, structured, and acted on. Recent approaches to context-awareness in AR have therefore sought to propose system architectures and design taxonomies that support adaptive system behaviours.

2.3 Context Awareness in AR

Several approaches have been proposed to integrate contextual understanding into AR interaction, each offering unique perspectives on how AR systems can adapt to users and their environments.

2.3.1 Pervasive AR

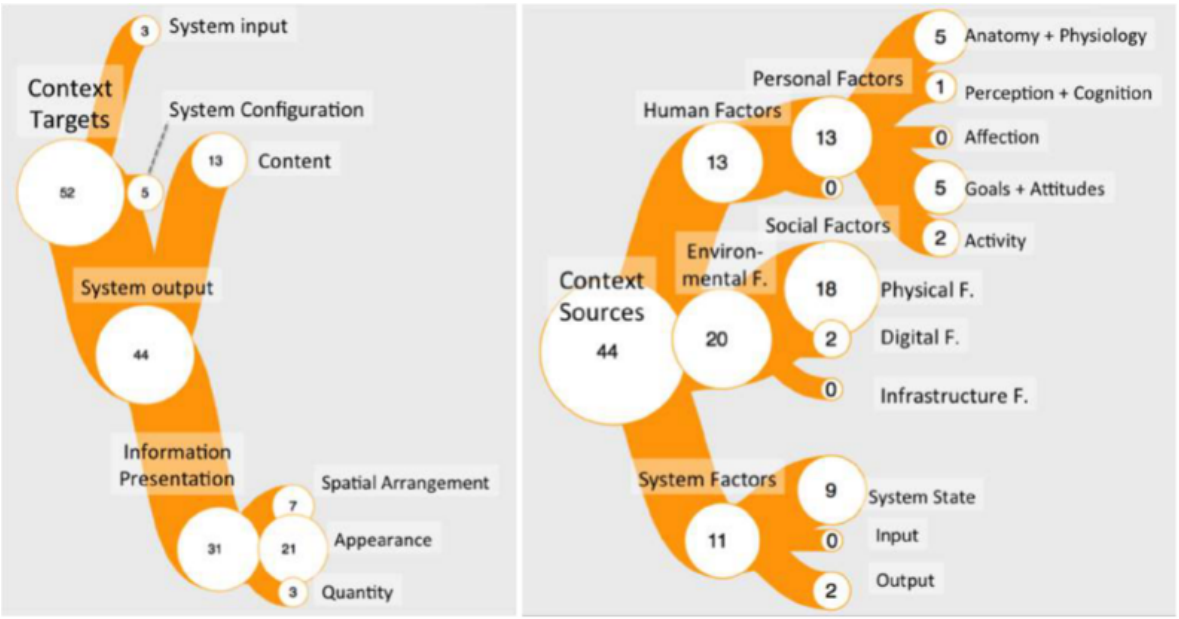

Figure 2.1: Pervasive AR (PAR) categorises context into three facets: context sources, targets, and controllers. Sources are split into human, environmental and system factors, with sources defined by input, output and system configuration (Grubert et al., 2017).

Grubert et al. (2017) introduce Pervasive AR (PAR), a taxonomy that re-uses the high-level categorisation of context sources, targets, and controllers (Lacoche et al., 2014). Context sources include human, environmental, and system factors that influence

interactions, whereas context targets represent the aspects of the system that can be adapted. This encompasses inputs, outputs and system configuration, which includes factors like the techniques used (Cheng, Gebhardt, and Holz, 2023; X. B. Liu et al., 2024), interface layout and content presentation (Davari, Stover, et al., 2024; Ng et al., 2021) or the methods employed by a system to sense input and deliver output (e.g. adapting tracking algorithms employed (X. B. Liu et al., 2024) or display brightness (Schuchhardt et al., 2015)). Lastly, context controllers define the mechanisms for adaptation, which can be either automated or user-directed (Strauss et al., 2024). This structured approach provides a foundation for developing adaptive systems by defining what data to capture and how it should be considered (see Figure 2.1).

2.3.2 Intelligent AR

More recently, Davari, Stover, et al. (2024) presented Intelligent AR (iAR), a framework which structures context sources into three key categories: specified data, sensed data, and extracted data (see Figure 2.2).

In their categorisation, they define specified data as user-provided input, such as display information (e.g. usernames) and personal preferences/settings. Sensed data, akin to context sources as introduced by Grubert et al. (2017), is collected in real time from sensors, cameras, and other input devices, reflecting the user's current state and surroundings. Finally, extracted data is accessed from personal or public databases. This is stored information that has previously been specified or sensed, enabling systems to anticipate user needs based on factors such as their preferences and past behaviours. They focus on making adaptations to output, such as the visibility, spatial layout and appearance of different types of information. This is achieved based on capturing dynamic or persistent contextual information, considering the impact of mobility, environment, and real-world objectives on adaptations. iAR leverages specified, sensed and extracted data to create adaptive AR experiences that minimise reliance on explicit input and

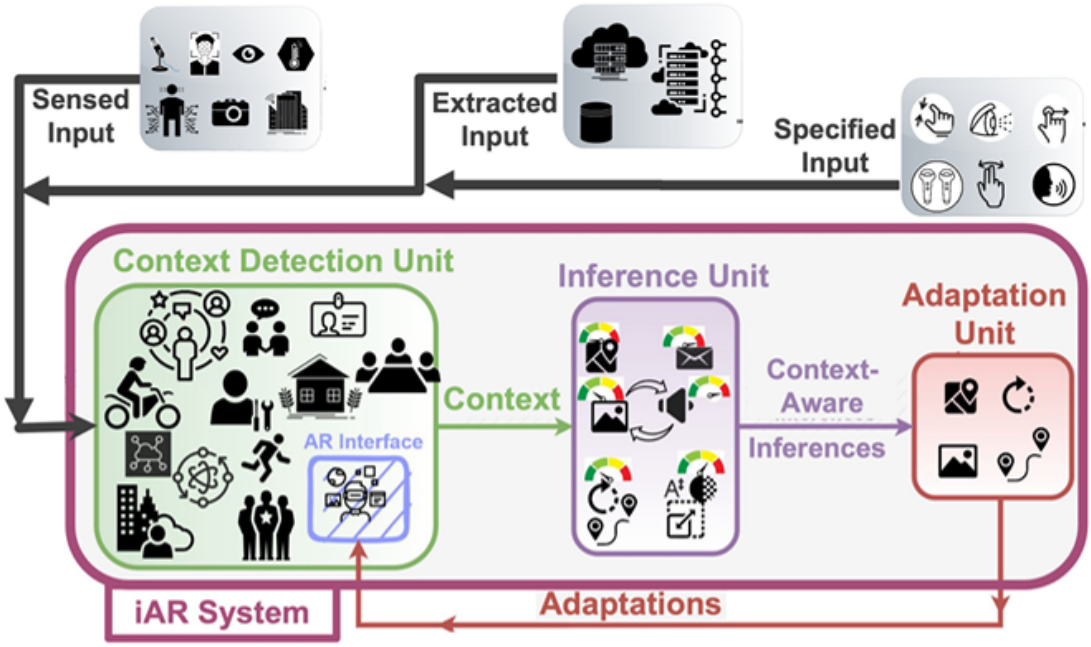

Figure 2.2: Intelligent AR (iAR) structures context sources into three main categories: specified data, sensed data, and extracted data which a system can use for detection, inference and adaptation (Davari, Stover, et al., 2024)

user-specified data, which Davari, Stover, et al. (2024) argue is more prone to error than autonomous adaptations.

2.3.3 Human I/O

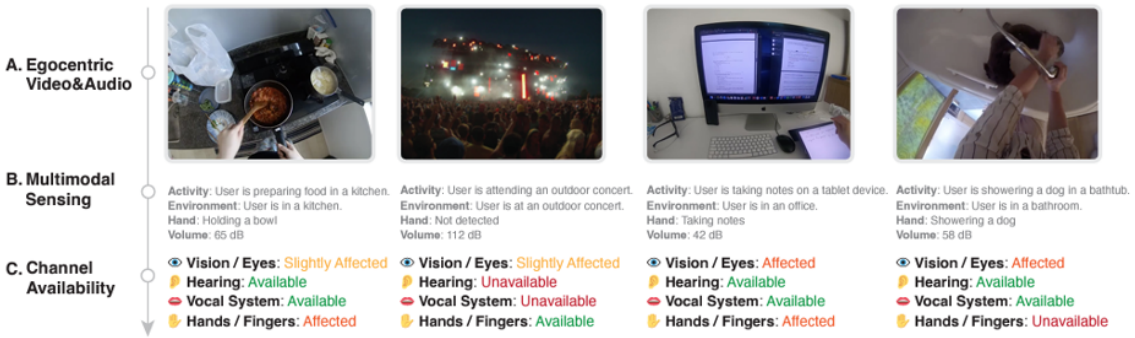

Figure 2.3: Human I/O detects situational impairments by assessing the availability of human input/output channels: vision, hearing, vocal system, and hands. This is achieved by capturing data via egocentric video and audio streams, which is processed through multimodal AI and large language models to generate real-time predictions (X. B. Liu et al., 2024)

Human I/O (X. B. Liu et al., 2024) primarily focuses on adapting explicit in-

put methods, detecting situational impairments by assessing the availability of human input/output channels: vision, hearing, vocal system, and hands. Their system harnesses sensed data (Davari, Stover, et al., 2024), captured via egocentric video and audio streams, which is processed through multimodal channels and large language models to generate real-time predictions. This is achieved via a four-level classification of channel availability (ranging from Available to Unavailable) as opposed to a binary scale as employed in most systems, showing high accuracy in diverse interaction scenarios. Unlike the taxonomies proposed by Grubert et al. (2017) and Davari, Stover, et al. (2024), Human I/O offers a practical, unified solution aimed at improving accessibility and explicit interaction.

2.3.4 Proxemic Dimensions for Context-Awareness in AR

While the above approaches differ in their underlying strategies and implementations, they all encompass dimensions of proxemic interaction (Distance, Orientation, Movement, Identity, and Location) in several ways. For example, Grubert et al. (2017) reference every dimension, e.g. user-object distance and movements of people around the user. Although they explicitly include spatial arrangement as a consideration in context targets, this is not highlighted in the same way for context sources, even though they note the impact of environment crowdedness on interaction. Higher emphasis is also placed on output as a context target, with Grubert et al. (2017) highlighting less research has explored adaptations to input. Similarly, iAR discuss content layout but not the distance/orientation of objects explicitly, referring to 'nearby objects'. They also explore using orientation for fixing objects to a users head/torso (Davari, Stover, et al., 2024), however little attention is paid to adapting input.

Human I/O (X. B. Liu et al., 2024) infers semantic information by considering who and what is present in a scene, allowing for understanding of a users activity (including walking scenarios) and environment. This research places a stronger emphasis on under-

standing contextual factors to adapt input methods, but does not study AR interaction explicitly.

Overall, these approaches underscore the evolving landscape of context-awareness in AR, demonstrating that multi-source data integration and real-time adaptive capabilities is essential for delivering intelligent and user-centred AR experiences (Grubert et al., 2017; Davari, Stover, et al., 2024; X. B. Liu et al., 2024). Despite this, there is still a lack of grounding to understand how to capture and reference contextual factors in AR, and how to adapt explicit AR interaction techniques based on what is captured and inferred.

2.3.5 Transferable Interactions

As we move towards ubiquitous applications of immersive technologies, and to avoid the ad-hoc development of bespoke AR solutions (Frutos-Pascual, Creed, and I. Williams, 2019), it is essential to understand how inputs can be best mapped to different contextual factors (Davari and Doug A Bowman, 2024). Peripheral-free interaction in immersive environments can be divided into explicit and implicit inputs. Explicit interactions are defined as any intentional input provided to execute distinct tasks and manipulate the scene, notably to interact with virtual content within the 3D environment, whereas implicit interactions are a combination of inherent motion and location awareness within the interactive space, which triggers an inherent interaction (i.e. walking around a spatially registered object) (Maximilian Speicher, B. D. Hall, and Nebeling, 2019).

Recent advancements in multimodal AI and real-time data processing have demonstrated the potential for adaptive AR systems that respond to user, system and environmental factors (Xiaoan Liu et al., 2025; X. B. Liu et al., 2024; Grubert et al., 2017). By leveraging a combination of specified, sensed, and extracted data, AR interfaces can dynamically adapt and/or refine interaction techniques, accommodating diverse user needs and environmental constraints (Davari, Stover, et al., 2024). Capturing information such as how far a user is from physical or virtual entities, if/how they move, what they are

focused on, who or what is present in their interaction space (as well as any associated metadata and characteristics), preferences, and how the environment is arranged, allow contextual information to be interpreted by a system to tailor AR interactions in real time (Grubert et al., 2017; Greenberg, Boring, et al., 2014; Cheng, Gebhardt, and Holz, 2023).

Technologies could facilitate real-time adaptations by processing relevant data, and inferring the most appropriate adjustments to input and output (X. B. Liu et al., 2024; Davari, Stover, et al., 2024). This, in turn, would make it possible to capitalise on the advantages and disadvantages of different interaction techniques (X. B. Liu et al., 2024), adapt input and output to factors like the physical affordances of the environment and a user's cognitive demands (Cheng, Gebhardt, and Holz, 2023; Lindlbauer, Feit, and Hilliges, 2019), and ultimately better align AR interactions with user activities and expectations in a range of contexts (Grubert et al., 2017; X. B. Liu et al., 2024; Davari, Stover, et al., 2024).

2.4 Interaction Techniques

AR research has seen a recent surge in the development of novel interaction techniques that aim to enhance the value and scope of AR experiences. To increase their portability and practicality, most AR-capable devices now provide explicit interaction methods achievable using their built-in components, notably peripheral-free, touchless interactions (McGill, Williamson, et al., 2019). This has introduced a myriad of controllerless interaction capabilities on a range of AR displays, which include speech (Adam S. Williams, Garcia, and F. Ortega, 2020), freehand gesture (Piumsomboon, Clark, et al., 2013), head and eye gaze (Hertel et al., 2021). This section introduces these commonplace interaction techniques considered in AR research and throughout the thesis.

2.4.1 Speech Interaction

As language is essential for supporting elements of human understanding, needed to define, communicate and discuss the nature, limitations, and scope of explanations (Jackson, 2020), speech is employed as a common input method for digital interfaces (Hertel et al., 2021). Speech has been implemented into a range of ubiquitous devices and technologies, including modern consumer AR-capable headsets, where many AR platforms provide voice input as an intuitive, hands-free and eyes-free communication method (Spittle, Frutos-Pascual, et al., 2022; Augstein, Neumayr, and Pimminger, 2019; Muhammad Nizam et al., 2018).

For some time, it has been expected that communication with computers through natural language processing (NLP) represents the next major development in digital technology (Hirschberg and Manning, 2015), which we are now seeing with developments in AI tools, large language models and assistants like ChatGPT, Microsoft Copilot and Google Gemini (Hochmair, Juhász, and Kemp, 2024). Using these tools, systems are capable of supporting conversational flows, which are especially useful for AR applications involving tasks like generative content creation (T. Chen et al., 2023) and remote collaboration (Jing, G. Lee, and Billinghurst, 2022).

Speech can act as a straighforward way to interface with technologies, however, due to the high number of possible aliases for speech proposals, speech interfaces have often lacked flexibility to adapt to the nuances of language and natural communication (Monteiro et al., 2023). This includes difficulties adapting to individual users, where factors such as dialects, accents, and ambiguities cause issues with system interpretation (Jackson, 2020). AR studies have shown that some constructs of speech, such as simple verb-direction pairs (for example “move-up”, “move-down”) achieve strong consensus. However, when users invoke synonyms for tasks like scaling (“grow”, “expand”, “zoom”) or rotation (“spin”, “tilt”, “rotate”), agreement degrades and recognition errors increase (Adam S. Williams and Francisco R. Ortega, 2020). Likewise, abstract object references

("cube", "object", "that") vary unpredictably across users, often forcing designers to implement finite vocabularies using best judgment (X. Zhou, Adam Sinclair Williams, and Francisco Raul Ortega, 2022). This means that many previous speech interfaces have employed brittle, rigid mappings that are not always intuitive to employ (Jaewook Lee et al., 2023; Jackson, 2020).

Although language processing methods are still often limited in representing ambiguities and contradictions for the broad range of human intentions, thoughts, and beliefs (Jackson, 2020), speech remains robust at any distance and range, even beyond human visual limits (Monteiro et al., 2023). This is a benefit in AR environments, as unlike most gaze and freehand interactions, voice commands can be employed to interact with virtual content that sits outside the user's field of view or beyond comfortable interaction zones. Despite this advantage, this strength rarely translates well to unimodal AR input. Instead, speech has been found most effective when treated as a complementary input method, where voice can be used to provide high-level, descriptive tasks in a communicative way (e.g., to change parameters like colour and shape (M. Lee et al., 2013; Jaewook Lee et al., 2023)), while hand gestures, gaze tracking or controllers are often more appropriate for interactions requiring precision, and where spatial accuracy is paramount (Morotti et al., 2021; Buchta et al., 2022; Whitlock et al., 2018).

Even with the recent proliferation of powerful NLP pipelines and large language models, which are capable of parsing a wider range of open-ended utterances, speech interaction in AR remains inherently ambiguous. Formal-logic based recognisers (as still employed in many existing applications) excel at a fixed set of concise commands but tend to struggle when users describe elements of scenes, where targets are difficult to reference explicity and undefined pronouns are employed. This, in turn, creates semantic gaps that speech parsers are unable to reliably bridge in real time without added contextual information (Divekar et al., 2019; Lamberti et al., 2017; Jaewook Lee et al., 2023). In practice, combining modalities mitigates these shortcomings, where speech can offer an intuitive way to explicitly communicate descriptive requests to system (Adam S. Williams,

Garcia, and F. Ortega, 2020; T. Chen et al., 2023) or other users (Jing, G. Lee, and Billinghurst, 2022), while freehand and gaze-based inputs are more likely to deliver the speed and control expected from traditional, desktop-style workflows (Morotti et al., 2021; Monteiro et al., 2023; Spittle, Frutos-Pascual, et al., 2022).

2.4.2 Freehand Interaction

In recent years, freehand interaction has been made more feasible with the release of affordable and more reliiable sensors, which has offered richer and more intuitive communication possibilities (Xiaoan Liu et al., 2025; Rutten and Geerts, 2020). There is a repertoire of movements exploited for freehand interaction. As depicted in Figure 2.4, these include finger movements and gestures, palm movements, such as palm shapes and orientation, as well as hand movements and gestures, which can be interpreted in relation to other body parts (Koutsabasis and Vogiatzidakis, 2019).

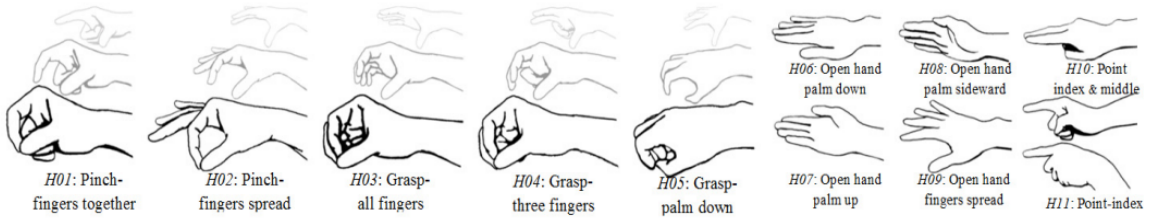

Figure 2.4: Variants of hand poses observed in a pioneering gesture guessability study for freehand interaction in AR (Piumsomboon, Clark, et al., 2013)

Interaction metaphors have also been developed exclusively for smartphones and HWDs. This has maximised the ubiquity and mobility of applications by implementing hand and finger recognition via built-in cameras (Sharma et al., 2024; Chan et al., 2016). Siddhpuria et al. (2017) argue that freehand input is suitable for the mobile and discrete nature of interactions for AR and ubiquitous computing and identify how hand gestures can effectively be sensed by built-in hardware components for a range of applications, for example, to control home appliances and distant displays.

Chan et al. (2016) also explored how subtle and minimum-effort gestures can be

effectively applied for portable ubiquitous computing commands. They highlight how single hand gestures, such as 'pinch' gestures, can be performed regularly in public contexts, whereas large gestures may be perceived as more physically demanding and socially awkward. When interacting with freehand gesture-based interfaces, physically and conceptually simple gestures have been favoured over more complex gestures. Rutten and Geerts (2020) highlight how static gestures, such as opening or closing the hand, have been regarded as more efficient, precise, enjoyable and intutive to employ when compared to dynamic gestures, where users are required to conduct arm and hand movements, as seen with 'push' or 'circling' gestures.

As selection is a familiar interaction that is used for a variety of existing interfaces, for example, when pressing physical buttons or using a touch-screen display, selection gestures have been heavily explored (Hertel et al., 2021). Such interactions include pointing with the index finger or tapping with the index finger (Xinyi Liu et al., 2022; Rutten and Geerts, 2020). Selection has been employed for a range of tasks in AR environments, for example, for text entry (W. Xu, Liang, He, et al., 2019), and to simulate clicks in desktop environments for UI interaction or 3D object selection/manipulation (Frutos-Pascual, Creed, and I. Williams, 2019).

Pham et al. (2018) highlight that, as interaction is mediated by how digital content is perceived, the scale and distance of interactive elements, the size of environments, and the approach employed to present digital information to the user will affect freehand gesture approaches. This has resulted in different design approaches for freehand interactions across distances.

For near-field interaction (where content is within arms reach), isomorphic gesture paradigms are primarily used, which mirror the physical motions we employ when interacting with objects in the real world (Piumsomboon, Altimira, et al., 2014). This means the gestures are designed to feel intuitive, as if users are manipulating tangible objects. In contrast, when it comes to freehand interaction at a distance, metaphoric gestures are often adopted (Caputo, Bartolomioli, and Giachetti, 2023). These gestures

are based on familiar digital interactions used with everyday technologies, like desktop computers. For example, the pinch gesture, or "Airtap" gesture as defined by HoloLens 2 device manufacturers (Microsoft, 2021b), combines pointing and pinching motions, which act as a metaphor for a mouse click (Frutos-Pascual, Creed, and I. Williams, 2019). Research has shown that this gesture is as effective as pressing a button on a handheld controller (Mutasim, Batmaz, and Stuerzlinger, 2021), and can be used for a wide range of tasks, including selections, object manipulations, and scrolling on 2D panels placed beyond arm's reach (Hertel et al., 2021).

Caputo, Bartolomioli, and Giachetti (2023) focused on using different freehand techniques, including Airtap (see Figure 2.5), for room-scale selection and manipulation, where user-object distances were 1.6m and 3.2m. Results suggested that the efficiency of the method was not significantly affected by the target distance, however, they note that larger variability in distances could have determined increased effects. Lilligreen, Henkel, and Wiebel (2022) also reported no significant difference in usability between near (directly with the hands) and far (with a ray) interaction paradigms in an outdoor context, but highlight that some participants deemed the airtap technique to be atypical and unintuitive.

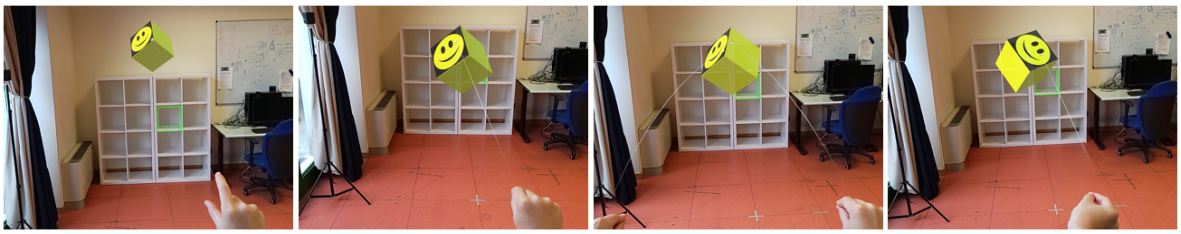

Figure 2.5: Interaction steps to complete a docking task with the Airtap technique. The cube is selected, rotated and resized to be inserted in the highlighted shelf compartment (Caputo, Bartolomioli, and Giachetti, 2023)

Even though Airtap gestures offer such high flexibility (Kang, J.-h. Shin, and Ponto, 2020; T. Wang et al., 2021), and have been found to have a relatively short learning curve (Pourmemar and Poullis, 2019), users commonly leave their hand posture in a "comfort grip" between interactions, which often causes a system to infer selections that are not intended (Pfeuffer, B. Mayer, et al., 2017). An alternative approach is to re-

move the "pinch" selection mechanism and reference deictic pointing gestures (Bernardos, Gómez, and Casar, 2016; Xinyi Liu et al., 2022; Pourmemar and Poullis, 2019).

Deictic pointing is part of instinctive social communication, where humans have the natural tendency to use their index finger or hand to indicate objects of interest. In AR environments, users are able to "point out" a target just as they would in face-to-face interaction. As exemplified in Figure 2.6, interactions can be triggered using a "Hover" technique, which functions similarly to a dwell mechanism. Interaction is initiated by holding the cursor over a target for a predetermined duration (Pourmemar and Poullis, 2019), effectively combining the intuitive act of pointing with a timed selection process.

Figure 2.6: User interacting with a menu using hand pointing (Hover), where selections are made after a dwell-time of 1.5s (Pourmemar and Poullis, 2019)

Although the reduced complexity of Hover could make freehand interaction easier to employ, time-based selection has been found erroneous due to the Midas touch problem. Midas touch occurs when systems lack a clear way to distinguish between intended, deliberate inputs and incidental movements. As a result, actions like hovering over an item can inadvertently trigger commands, thereby reducing the effectiveness and usability of the interaction technique (Bhowmick, Kalita, and Sorathia, 2020).

Users have also still reported high levels of physical demand and fatigue with hand pointing, which can be attributed to the technique requiring users to keep their hands raised when interacting with virtual content (Pourmemar and Poullis, 2019). Even though the technique has received poor ratings across all performance and usability measures for

precise selection tasks (Xinyi Liu et al., 2022), Hover has been perceived to be a mature technique, providing good performance in terms of accuracy, efficiency, and organisation when selecting relatively large objects (Bernardos, Gómez, and Casar, 2016).

Refinement Techniques for Freehand Interaction

Research demonstrates how system adaptations could be achieved through refinement methods, to reduce the impact of angular size and distance on freehand techniques. For example, depth-manipulation techniques such as Go-Go have been developed, which allow users to extend their reach beyond physical arm's length to interact with distant content. This is achieved by utilising a non-linear mapping system, enabling users to manipulate virtual hands within a larger area around them and employ far-field interactions more easily (Ugarte et al., 2022). Further, G. Wang et al. (2022) suggest presenting AR content at the ideal position and scale. To achieve this, systems could employ linear scaling, to resize content based on user-object distances (C. Liu, Plopski, and Orlosky, 2020).

Another technique, Worlds in Miniature (see Figure 2.7), provides users with miniature copies of objects, combining the advantages of an input space, a cartographic

Figure 2.7: Worlds In Miniature: Actual sized objects in the room are interacted with by directly selecting and moving the miniature representations of the objects (Kang, J.-h. Shin, and Ponto, 2020)

map, and an overview/detail interface. This offers a tool for navigation and object manipulation, simplifying the process of interaction with distant content (Danyluk et al., 2021; Kang, J.-h. Shin, and Ponto, 2020). Although the world in miniature technique is relatively easy to employ in VR, in many instances, such as in large-scale AR environments (Grubert et al., 2017; Bhowmick, Kalita, and Sorathia, 2020), it may soon become impractical. This is because AR environments involve a real-world scene, which would need to be scanned and accurately duplicated in real-time (Caputo, Bartolomioli, and Giachetti, 2023).

While refinement techniques, such as those discussed above, have proven effective in enhancing far-field freehand techniques by mitigating the physical constraints inherent in direct object interaction (Xiaoan Liu et al., 2025), they still lack the flexibility needed for designing techniques that work across varied scenarios and environments. Instead, it would be ideal for system inputs to adapt to users, their context, and preferences. This adaptability could be achieved by leveraging the strengths of different input modalities, each with unique affordances, rather than trying to make a single technique fit every interaction (Grubert et al., 2017; X. B. Liu et al., 2024).

2.4.3 Head and Eye Interaction

Although primarily used for implicit interactions (i.e. to provide a system with an understanding of user intent, or to search for content in the environment (Piening et al., 2021)), head and eye inputs have frequently been employed for explicit interactions (Hertel et al., 2021). Ng et al. (2021) note how hands-free techniques allow users to interact in a wide range of use cases that are not attainable with freehand techniques. Both head and eye techniques can be used instead of, or alongside (Kang, J.-h. Shin, and Ponto, 2020), freehand techniques and have notably been considered useful in cases where the user's hands are occupied with another task (Sadri et al., 2019; Pourmemar and Poullis, 2019).

Despite gaze-based techniques being more discreet and permitting interaction when encumbered (Heo et al., 2020; Ng et al., 2021), they are more difficult to design and employ than freehand for more complex tasks in AR (Sadri et al., 2019). This means they are typically reserved for simpler interactions, such as button presses, text input, object selections and menu navigation (Xinyi Liu et al., 2022; Pourmemar and Poullis, 2019; Pfeuffer, Abdrabou, et al., 2021). Figure 2.8 presents an example of how gaze has been used to navigate between various information levels of an educational 3D visualisation.

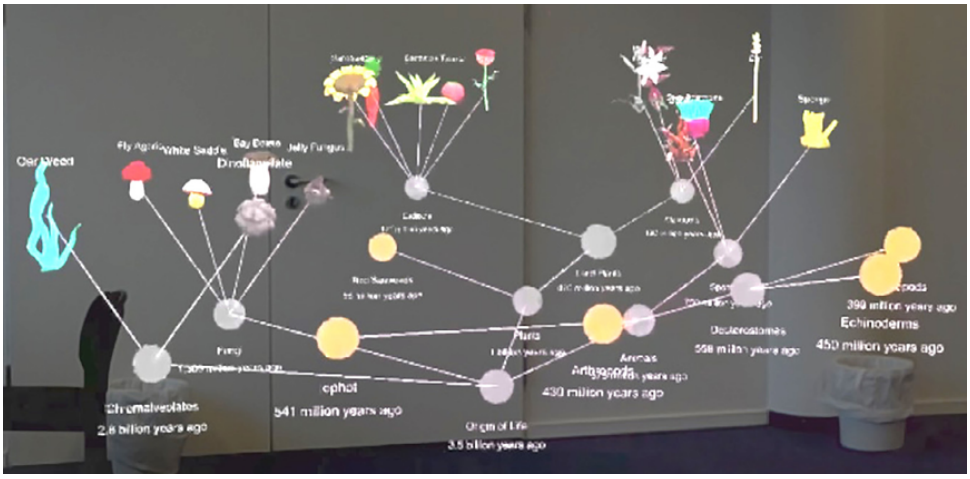

Figure 2.8: ARtention: How gaze could provide a way to implicitly navigate between various information levels of a complex 3D visualisation, in this case to learn about evolutionary biology (Pfeuffer, Abdrabou, et al., 2021).

Head Input

Head-based interaction not only serves as a valuable supplementary modality, integrating seamlessly with speech and hand gestures to enrich interactivity and clarify user intent (Yu et al., 2019; S. Mayer, Laput, and Harrison, 2020), but it has also been successfully implemented as a primary, standalone input method. For example, W. Xu, Liang, He, et al. (2019) discuss how head gesture has been implemented for text entry, a widespread and essential activity in immersive environments. They highlight how head motions provide both device-free and hands-free input techniques, the user able to position the cursor via their head orientation and make letter selections by employing 'nod' actions.

Piumsomboon, G. Lee, et al. (2017) also explore using head gestures as discrete inputs for tasks such as browsing picture galleries or controlling media players, where they were able to map head rolls to binary input commands like 'next' or 'previous'. Similarly, Špakov and Majaranta (2012) determined head gestures to be promising methods for navigation (turning) and functional mode switching (tilting).

Methods to permit the manipulation of content in immersive environments have also been explored. The majority utilise flexible and intuitive head movements, which are based on yaw, pitch and roll orientations (Yu et al., 2019). Sadri et al. (2019) demonstrate how controls can be mapped to head motion inputs, defining parameters to update interaction functions. For example, they define thresholds that determine whether head roll, or head yaw and pitch are referenced by the system, to control the speed and axis of rotation. Yu et al. (2019) highlight head orientation is commonly utilised for HMD devices, to assist with tasks such as content relocation, switching between states, panning and zooming, as well as for scaling and rotation. For example, the approach presented by Sadri et al. (2019) permits hands-free geometric transformation of 3D virtual models. Here, users can employ subtle and relatively undemanding head motions to support first-order control (rate), based on a models isotropic scale and its rotation around arbitrary axes, as well as zero-order control for translation.

Although Head input has been explored for conducting a range of tasks, gaze techniques have notably been found straightforward to learn and employ for selection, as they simply require users to direct their focus towards interactive components (Pfeuffer, Abdrabou, et al., 2021; Xinyi Liu et al., 2022). The process of employing head pointing for selection and manipulation is comparable to operating a handheld controller, however, only the data from the HWD is required, making it more straightforward for a user to implement (W. Xu, Liang, He, et al., 2019; Whitlock et al., 2018).

Yu et al. (2019) highlight the potential of employing head motion, arguing that the mode of input is an accurate and relatively fast interaction approach, which commonly offers a more time-efficient refinement technique than hand gestures or handheld devices.

Head techniques benefit from using built-in sensors of HWDs that are less affected by environmental factors. Thus, head tracking is more likely to receive consistent and accurate data than freehand and eye techniques (Uzor and Kristensson, 2021). This generally makes head pointing straightforward and reliable to employ, even when interacting with small distant targets (Xinyi Liu et al., 2022).

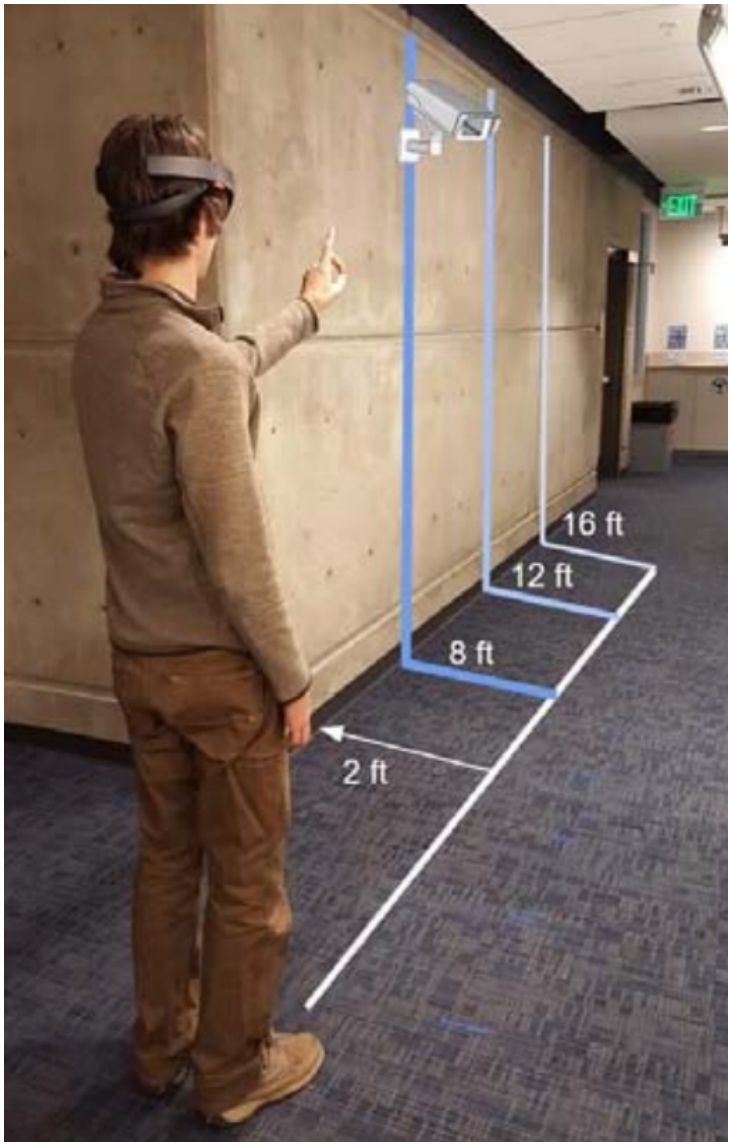

Figure 2.9: User interacting via head pointing and Airtap for making selections in an Internet of Things scenario. Techniques are explored at different distances across Social and Public zones (Whitlock et al., 2018).

For example, Whitlock et al. (2018) explored the impact of distance on techniques, comparing head-based pointing (with Airtap or Voice as a selection mechanism) and a hardware controller, for objects positioned between 2.4m and 4.9m from the user (see

Figure 2.9). Although handheld remotes are commonplace for real-world interactions, they found that participants performed comparably when using head-gaze as a pointing method, and it was perceived as significantly more efficient and usable than device-mediated interactions. They note that this was primarily due to the process of locating and maintaining awareness of the cursor, which was significantly easier with head due to the cursor appearing central to the field of view. Similarly, when comparing head with hand pointing, Pourmemar and Poullis (2019) highlighted that some users reported difficulty controlling the ray with hand techniques, an issue not observed with head pointing.

Eye Input

Although head-gaze is still currently the more affordable option (Hertel et al., 2021), eye tracking is becoming a more standard feature of HWDs; as provided by the HoloLens 2, Magic Leap 2, Meta Quest Pro and Apple Vision Pro (Gallardo et al., 2023). Where dwell time is used as a selection mechanism, especially with eye gaze, it can become difficult for a system to differentiate between user attention (i.e. where they are implicitly looking) and user intention (i.e. if they wish to trigger an interaction). Pfeuffer, Abdrabou, et al. (2021) highlight that this is because there are two fundamental tasks involved in gaze interaction: information consumption and interface selections, where there is strong consideration needed for the transitions that occur from reality to the virtual interface, and from single to multi-layer content (see Figure 2.10).

Whereas consumption is about revealing and seeing lightweight information, with the system analysing gaze implicitly (e.g. to display or adapt output), during selections, users decide to explicitly select a target, normally via dwell with unimodal gaze techniques. However, due to the difficulties presented in reliably distinguishing attention from intent, some research has found that users prefer a separate selection mechanism, such as button presses on a controller or blink (F. Lu, Pavanatto, and Doug A Bowman, 2023).

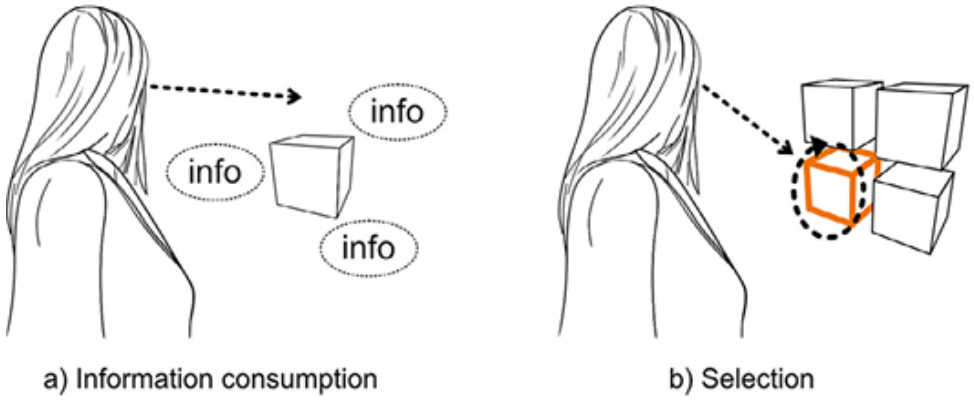

Figure 2.10: Gaze interaction involves a) Consuming information, where the system analyses gaze in the background, and 2) selecting information, where users decide to voluntarily interact with content (Pfeuffer, Abdrabou, et al., 2021).

Furthermore, the accuracy of eye techniques has been reported to be temperamental, decreasing considerably when targets are smaller and as the distance of objects increases (Mirbagheri and Chau, 2024; S. Wei, Bloemers, and Rovira, 2023; Kytö et al., 2018). This means that head gaze has often been found more reliable for indicating targets to select than eye (C. Liu, Plopski, and Orlosky, 2020). Eye-tracking systems have also been found less consistent, notably because they tend to be highly sensitive to noise (e.g. changing light adjusting pupil dilation, and rapid eye movements causing offsets and jitter) which increases the likeliness of inaccuracies and, thus, erroneous system behaviour (Uzor and Kristensson, 2021; Pfeuffer, Abdrabou, et al., 2021). Unlike head gaze, eye technologies also require manual calibration for each user to ensure accurate tracking (Mirbagheri and Chau, 2024), with a reported accuracy tolerance of around 2° being provided with the HoloLens 2 (Microsoft, 2023a). This often makes selecting small, distant content more challenging.

For example, Y. Y. Qian and Teather (2017) found that although participants could comfortably and reliably select larger targets with eye techniques, when content was smaller and/or farther away in a VR scene, interactions became frustrating and uncomfortable. They compared eye, head, and a multimodal technique using both eye and head for fundamental selection across different distances in the Public zone: 5m, 7m, 9m and mixed depths (see Figure 2.11). Results indicated that eye-only selection

Figure 2.11: Fundamental selection task involving moving the cursor (controlled by either the eye tracker, head orientation or both) to a highlighted sphere and pressing the "z" key on a hardware keyboard (Y. Y. Qian and Teather, 2017)

provided the worst performance in terms of error rate, selection times, and throughput, which was primarily attributed to issues with eye tracking and calibration. The hybrid approach, combining eye and head, failed to leverage the benefits of both methods, with head offering the fastest selection times, the best accuracy, and being strongly preferred.

Even though head and eye techniques can afford an effective pointing mechanism, research is yet to fully understand the appropriateness of these interaction methods at different distances (Hussain, Park, and H. K. Kim, 2023; Whitlock et al., 2018) and in different interaction contexts (Spittle, Frutos-Pascual, et al., 2022; X. B. Liu et al., 2024). This prompts more research to understand the advantages and limitations of these techniques in different use cases.

2.5 Summary

Chapter 2 has presented a narrative review of peripheral-free interaction in AR. It introduced proxemic dimensions (distance, orientation, movement, identity, and location) as a lens for framing context in AR, and examined how commonplace interaction techniques such as freehand, gaze, and speech have been supported on AR devices. Although this review has introduced the key concepts considered in the thesis, the literature should also

be considered systematically to better understand how these techniques are applied across different applications and settings. To address this, Chapter 3 presents a systematic literature review of peripheral-free interaction techniques used for explicit input across a spectrum of immersive technologies. VR is included alongside AR because it often employs peripheral-free inputs, offering study designs, evaluation methods, and interaction techniques that are directly transferable to AR. By quantifying how display types, input methods, use cases, and tasks have been studied, the next chapter identifies research gaps that help define the empirical investigations reported in chapters 4, 5 and 6.

Chapter Three

Interaction Techniques for Immersive Environments

Contents

3.1 Introduction **45**
3.2 Method **46**
3.2.1 Data Collection 47
3.2.2 Data Analysis 51
3.3 Analysis **53**
3.3.1 Top-Level Review 53
3.3.2 Handheld Display 57
3.3.3 Headworn Display 61
3.3.4 Multiple Display Types 66
3.4 Conclusions and Recommendations **69**
3.5 Limitations **77**
3.6 Summary **78**
3.7 Implications for Empirical Studies **78**

Note: This chapter is adapted from previously published work

Spittle, B., Frutos-Pascual, M., Creed, C. and Williams, I. (2022), "A Review of Interaction Techniques for Immersive Environments", in IEEE TVCG vol. 29, no. 9, pp. 3900-3921, 1 Sept. 2023.

3.1 Introduction

The recent proliferation of immersive technology has led to the rapid adoption of consumer-ready hardware for AR and VR. While this increase has resulted in a variety of platforms that can offer a richer interactive experience, the advances in technology bring more variability in display types, interaction sensors and use cases. This provides a spectrum of device-specific interaction possibilities, with each offering a tailor-made solution for delivering immersive experiences to users, but often with a lack of standardisation and consistency across devices and applications (Grubert et al., 2017).

To build on the literature review in Chapter 2, and better understand the landscape of XR interaction research, a systematic review and an evaluation of explicit, task-based interaction methods in immersive environments are presented in this chapter. To form a grounding for the positioning of the thesis, and inform the scope of the empirical studies presented in Chapters 4, 5, and 6, a corpus of 102 papers published between 2013 and 2024 is reviewed. This is to thoroughly explore state-of-the-art user studies, which investigate input methods and their implementation for immersive interaction tasks (pointing, selection, translation, rotation, scale, viewport, menu-based and abstract).

Focus is given to how peripheral-free input methods have been applied within the spectrum of immersive technology (AR, MR, VR; XR), which is achieved by categorising findings based on display type, input method, study type, use case and task. Results illustrate key trends surrounding the benefits and limitations of each interaction technique and highlight the gaps in current research. The review provides a foundation for understanding the current and future directions for interaction studies in immersive environments, which, at this pivotal point in immersive technology adoption, provides routes forward for achieving more valuable interactive experiences based on contextual factors. By exploring how content producers can fully reap the benefits of peripheral-free interaction capabilities, we can work towards making interaction more standardised and transferable across the range of XR tasks, devices and use cases.

This chapter is structured as follows. Section 3.2 outlines the methodology used to capture and review the data, including the inclusion and exclusion criteria, the applied categorisations, and the factors considered for analysis. Section 3.3 presents the analysis of the literature, providing quantitative results and insights related to the reviewed factors. These findings are organised according to the primary categories of XR devices: Handheld, Headworn, and Multiple-Displays. Finally, Section 3.4 offers recommendations, conclusions, and directions for future work. This informs the basis for the empirical studies conducted in Chapters 4, 5 and 6.

3.2 Method

The focus of this review surrounds immersive technologies and provides a true representation of the input techniques explored for XR interaction. Searches were not restricted to specific publishing venues, meaning a range of papers were considered. This included full and short papers sourced from journals and conference proceedings.

Paper quality was assessed based on how thoroughly the research addressed the factors defined as key areas for exploration (in table 3.2). Although affiliations were taken into account to prevent bias, and several highly cited papers were included in the review, publication impact was not a primary concern. Population size was also noted, to help determine the impact of surveyed works, yet the number of participants did not influence whether a paper was included.

The methods that were applied to filter, collect and prepare the data for analysis are further defined in the following subsections.

3.2.1 Data Collection

A sample of 294 eligible papers was collated from ACM Digital Library, IEEE Explore and other databases prevalent in the fields of HCI and computer science, such as Springer, Elsevier, IFIP and Oxford Press. To define the corpus of papers for consideration, information was required for factors surrounding the type of study, display used, testing conditions, experimental set-up/design and the data collected.

Table 3.1: Search Terms: Query applied to the IEEE and ACM databases, where each row of the table represents 'AND' and each comma between search terms represents 'OR'.

Topic	Search Terms	Location
Study Type/Technology	elicit*, compar*, virtual, augmented, mixed, VR, AR, MR, immersive	Title
Display/Input	mobile, HMD, HWD, head mounted, head worn, tablet, smart phone, interact*, input, technique*	Abstract
Interaction	method, intuitive, natural, modality, multimodal, ambigu*	Abstract
Modality	speech, voice, head, eye, gaze, hand, gesture*	Abstract
Tasks	point*, select*, manipulat*, mov*, translat*, position*, rotat*, scal*, menu	Abstract
Use case	environment*, context*, scenario*, condition*, adapt*, hands free, eyes free	Abstract
General	participant*, subject*, user*, study	Full text

Search terms were derived through pilot searches in ACM DL and IEEE Xplore. These pilots involved testing draft queries based on keywords that were common in sample papers, and reviewing the number and relevance of search results. Synonymns and variants (e.g., "head mounted" and "head worn"; "speech" and "voice") were included to improve coverage. Focus was given to studies that directly evaluated explicit input techniques while filtering out broader work (e.g. that considering rendering, hardware, or output). This refinement ensured that the final query captured a consistent and comprehensive set of papers.

Search terms were applied to the advanced search engines of the chosen databases, as categorised in table 3.1. Search terms concerning 'Study Type' or 'Technology' were

referenced in the Title. The remaining terms were searched within the Abstract, apart from those classified as 'General', where they were applied to the body of text.

Inclusion/Exclusion Criteria

To inform inclusion/exclusion criteria, the review conducted by Z. Bai and Blackwell (2012) was considered. This work provides a reference point for evaluation techniques, trends and challenges, which are provided to benefit XR researchers intending to design, conduct and interpret usability evaluations. Consequently, their considerations were deemed transferable for this review. Building on this, the criteria were defined to focus specifically on peripheral-free input methods available on current consumer-level AR/VR displays (Hertel et al., 2021), as these are the most prevelant techniques capable of supporting continuous, multi-purpose interactions (X. B. Liu et al., 2024; Grubert et al., 2017). Studies were only included if they evaluated explicit input techniques with participants, ensuring an empirical basis for comparison. By excluding domain-specific systems (e.g., CAVE, projection), hardware prototypes, or research focused on output as opposed to input, the scope was kept consistent with the thesis research questions and reflective of interaction techniques currently available on consumer devices (Gallardo et al., 2023).

To ensure papers were relevant and comparable, they had to meet the following criteria:

Display: They should consider a) Headworn (HMDs or smart glasses), b) Handheld (wireless smart devices), or c) Static (monitor) display types.

These display types were targeted as they are ubiquitous, consumer-level devices that are also widely employed for XR research. Where output was delivered to the participant via a monitor, studies were also required to consider either a headworn or handheld display.

Table 3.2: Data categorisation approach: The factors assessed for the data analysis and their definitions.

Factor	**Categorisation**	**Definition**
Display Type	Headworn Display	Head Mounted/Headworn Displays (HMDs/HWDs)/smart-glasses
	Handheld Display	Smartphones/tablets
	Multiple Displays	A combination of Headworn with Handheld, or one of these displays alongside a static display (i.e. desktop monitors/TV screens)
Input	Freehand	Using predefined gestures or unconstrained hand input with no wearable devices
	Speech-based	Using specific voice commands or natural language
	Head-based	Gaze interaction, orientations, rotations and head gestures
	Eye-based	Gaze interaction and Eye gestures
	Hardware-Based	Where a handheld display or external controller is employed; such as a touchscreen/touch-pad, button/switch, or 6-DoF manipulation of a handheld device
Type of Study	Elicitation	Where the users were asked to define their own interaction methods
	Assessment	Where users were asked to use a specific input/task combination and researchers assessed usability and feasibility for a given application/parameter
	Comparison	Where parameters (i.e. interaction methods or input/task combinations) were evaluated against a baseline or each other
Use case	Test Environment	–
	Lab	Constrained research setting
	Wild	Realistic use setting
	Scenario	–
	Static	Where interactions are conducted from a single position
	In motion	Where participants are free to move, or where interactions are performed whilst in motion
Tasks	Pointing	Searching for interactive elements (e.g. via a cursor or ray casting)
	Selection	Initiating/confirming an interaction
	Translation	Moving or relocating an interactive element
	Rotation	Changing the orientation of an interactive element
	Scaling	Reducing or enlarging the size of an interactive element
	Viewport control	Zooming and panning within an environment via a specific function (as opposed to implicitly moving around a scene)
	Menu-Based	Displaying a structured set of tabs, commands and/or utilities for the user to interact with
	Abstract	Non-spatial interactions such as editing (delete, undo, redo, insert, group, among others)

Even though the display conditions included are heterogeneous (papers reporting on multiple combinations of hardware setups), those only considering less accessible displays, such as CAVE, smart mirrors, and projection environments were excluded. This is because these display types are more restricted to specific domains (i.e. applicable for ad-hoc, research and business applications, as opposed to more generalisable consumer interactions).

Input: Studies had to concern one or more of the following peripheral-free input methods: a) Speech, b) Head, c) Eye, d) Freehand, and/or e) Hardware-Based interaction with handheld smart devices (i.e. touchscreen or 6-DoF motion gestures).

These inputs were defined as they are the most widespread and applicable to interaction with the targeted display devices. They are also generally straightforward to implement using the built-in components of XR devices.

Although some studies used hardware switches/controllers or marker-based interaction, they were only included for review if they considered at least one of the targeted inputs (as defined in Table 3.2). For example, if a study used head input for pointing but used a physical button/switch to initiate a selection, or if an external input type was included in comparison to a target input, then the paper was deemed to provide value to the review. If the paper only examined external inputs across all conditions (i.e. a dedicated controller for pointing and selecting), then it was not included. Studies that were deemed to predominantly consider the impact of output, as opposed to input, were also discounted.

Study type: Studies were required to explore interaction for AR/VR applications. They also had to consider user accomplishments of application tasks or interactions, based on the defined input methods, or low-level tasks which assess human perception or cognition. However, this had to be strongly related to input approaches, implementing at least one form of explicit interaction.

Papers that were found to consider interaction outside of XR technologies were classed as false positives. Studies that focused on novel hardware technologies were also excluded, as well as those primarily considering implicit interaction and output effects (i.e. to guide users to the correct interaction).

Participants: Papers should clearly state the number of participants, the purpose of the study, and its contribution.

Publication date: Studies should have been published between 2013 and 2024. 2013 was defined as the cut-off date due to the impactful work presented by Piumsomboon et al. (Piumsomboon, Clark, et al., 2013). For their research, the surface taxonomy provided by Wobbrock et al. (Wobbrock, Morris, and Wilson, 2009) was adapted to be better suited to AR gesture design. This resulted in the first user-defined taxonomy for intuitive hand interaction with holograms.

Of the 294 papers initially deemed to fulfil inclusion/exclusion criteria, 49 papers were selected for full review from ACM DL and 42 from IEEE Xplore. These publications were complemented by 11 papers from Springer, Elsevier, IFIP, or Oxford Press. This resulted in a corpus of 102 papers, representing roughly a third of the eligible publications. More recent and relevant studies were prioritised to provide an in-depth, state-of-the-art representation of current technologies and input capabilities.

3.2.2 Data Analysis

When conducting the review, there were five predominant areas of interest that embodied the factors considered. Table 3.2 provides the categorisations and definitions that were applied to analyse the sample of papers.

The first primary research area is *Display Type*, which is defined as the hardware employed for visualising virtual content.

Input concerned the interaction methods observed as part of the user studies, which were used to interface with the display. *Type of Study* refers to the type of user evaluation conducted, with *Use Case* exploring the conditions that studies are conducted under. This involved reporting on the testing environment and users' scenario, particularly their pose (i.e. whether they were instructed to remain seated or if they were free to move), and highlighting to what extent interaction approaches were pre-defined and restricted for the research.

The final consideration was *Tasks*, which defined the interactions that the research reported on. The task categorisations were informed by the work of Piumsomboon, Clark, et al. (2013) and represent distinct functions, which are often combined to complete more complex activities in immersive environments.

These five factors are primary considerations for interaction and are commonly explored in reviews. For example, Hertel et al. (2021) extract prevalent characteristics of interaction techniques based on input method and task and develop a taxonomy that sorts and groups them accordingly. Dey et al. (2016) also discuss these factors in their review to identify primary application areas. They describe the methods and environments that are used for user studies, to propose guidelines and future research opportunities. Furthermore, the factors represent themes considered by Laviola et al. (2017), where theoretical foundations, devices, techniques and design guidelines are explored in detail.

To clearly dissect information and highlight patterns and trends, data was extracted from each paper and coded within a matrix (based on the factors in table 3.2). There were three matrices, separated by display type (Headworn, Handheld and Multiple displays). The range of categories defined were not strictly binary, with papers being codified into more than one category where applicable. For example, a significant number of papers examined more than one input method in comparison, or combination when multimodal approaches were explored.

Data extraction and categorisation were conducted by a single researcher to ensure

consistency in coding. The same researcher applied the definitions in Table 3.2 across all papers, reducing variability in how categories were assigned. While this approach did not involve inter-rater validation, it provided a coherent basis for identifying patterns and trends within the sample.

3.3 Analysis

This section provides a summary of the data captured and highlights identified trends. Initially, a top-level analysis is conducted to encapsulate the data, reporting on the factors that were defined as key areas for exploration in section 3.2.2. Following this, the data was analysed by display type. This was to provide a breakdown of the inputs employed, testing conditions implemented and tasks observed for different immersive platforms.

3.3.1 Top-Level Review

This subsection summarises the data captured from the 102 papers included for review. Of these papers, 84 were sourced from conferences and 18 from journals. The data discussed is presented for handheld, headworn and multiple displays in Figures 3.1, 3.2 and 3.3, respectively.

Display Type

Roughly two-thirds of studies employed only headworn displays. There were an equal number of papers that implemented either solely handheld displays or multiple displays. Overall, 70 papers were found to target AR technologies and 38 were classified as VR. 6 papers reported to provide insight into both AR and VR.

Input

Most papers investigated either hardware-based input (41 of which considered interaction with external hardware controllers (D. L. Chen, Balakrishnan, and Grossman, 2020; Becker, Rauchenstein, and Sörös, 2019)) or freehand gesture. Head was explored slightly less, closely followed by speech. A total of 56 papers were found to include multimodal input techniques.

Type of Study

All studies were identified as assessments, the majority of which also included a comparison. There were considerably fewer papers reporting on elicitation studies. Although it was not a focus of the review, information was also captured surrounding the factors that were assessed and/or compared.

As input is strongly related to how users respond to output, papers notably included visual parameters as variables (such as distance (Whitlock et al., 2018) and scale (Pham et al., 2018)) to test input approaches. Comparison studies generally analysed more than one input technique or display/interaction device, either under AR/VR conditions, or sometimes considering an immersive application against a standard, non-immersive baseline (Bothén, Font, and Nilsson, 2018).

Relating to study type, an overview of participant sample and study protocol conditions is also provided, based on the parameters listed below:

Participant Sample: The average number of participants was 22.38 (SD = 11.12), with the largest sample being 73 (Bhowmick, Kalita, and Sorathia, 2020) and the smallest 7 (Sidorakis, Koulieris, and Mania, 2015).

Participant age: 9 out of 102 studies did not report on average sample age. 11 papers provided vague demographics, either stating their participants were above 18 (J. Zhao

et al., 2020), or briefly referring to the ages of participants without explicitly stating their range (Bothén, Font, and Nilsson, 2018). For the remaining 82 studies, the average age was 27.82 (SD = 5.33).

Participant experience: 82 studies reported on participants' relevant background experience (i.e. with the technologies, devices and interaction paradigms involved). 19 of these studies involved participants with previous basic or intermediate experience using relevant technologies, while 9 involved a sample with no previous experience. 51 studies included participants with different levels of experience, with 3 papers explicity reporting to include experts in their recruitment.

Study duration: Overall, 61 papers reported average completion times per participant, with sessions ranging from 20 to 120 minutes. 92 papers reported on studies conducted during a single iteration, while 10 reported on longitudinal studies, capturing data from the same participants on multiple occasions to assess learnability over time.

Another aspect addressed as part of study type was the kind of contribution. 70 papers were proposed to address or understand fundamental problems associated with explicit interaction in immersive environments, with 48 papers being considered relevant to a specific implementation. 16 explored fundamental findings and went on to apply them to a final application.

Notable areas of contribution surround selection (Franco and Cabral, 2019; Esteves, Y. Shin, and Oakley, 2020; Bhowmick, Kalita, and Sorathia, 2020), object manipulation (D. L. Chen, Balakrishnan, and Grossman, 2020; Adam S. Williams, Garcia, and F. Ortega, 2020; Piumsomboon, Altimira, et al., 2014), text entry (W. Xu, Liang, He, et al., 2019; X. Lu et al., 2019), game interaction(Tung et al., 2015; Bothén, Font, and Nilsson, 2018), character control and animation (Ye et al., 2020; Arora et al., 2019), human-human (Väyrynen et al., 2018) and human-robot collaboration (Krupke et al., 2018; Frank, Moorhead, and Kapila, 2016), map exploration (Satriadi et al., 2019), UI (user interface) and menu-based interaction (Pourmemar and Poullis, 2019; Bailly, Leit-

ner, and Nigay, 2019), Medical/Healthcare (Prilla, Janßen, and Kunzendorff, 2019; Sadri et al., 2019), interactive learning (Munsinger, White, and Quarles, 2019; Bazzaza et al., 2014) and AR assistants (J. Zhao et al., 2020; S. Mayer, Laput, and Harrison, 2020). Some studies could be classified into more than one of these categories, such as the work of Sadri et al. (2019), which focuses on anatomic model manipulation for medical applications.

Use Case

The majority of studies were conducted under constrained, predetermined conditions in a laboratory environment. Only a small number of studies (n = 13) were delivered outside of the research lab (in the wild). Even though the majority of studies used mobile technologies (untethered headworn and handheld devices), most papers reported on studies conducted from a single position in the testing space. Around a fifth of studies (n = 22) considered employing the freedom of movement offered by such devices.

Tasks

96 papers discussed a combination of tasks for their evaluations. Selection tasks were by far the most prevalent, followed by pointing and translation. Although reported slightly less than translation, transformation tasks were also broadly included (rotation slightly more than scale), as well as UI/menu-based interaction. Viewport control, such as zooming and panning, was explored considerably less.

Studies often assessed more complex interactions by adopting different combinations of explicit tasks. The majority of combinations included 2 tasks, which were noted by 31 papers, followed by 3 tasks (included in 29 papers). In 19 papers, 5 or more tasks were considered, and 4 tasks were featured in 17 papers.

Data was captured from participants based on a range of objective and subjective

factors. 100 papers reported on quantitative metrics and 86 presented qualitative feedback. In 84 of the papers, both quantitative and qualitative measures were considered. This is likely the case as a mixed-methods approach is held as the most valid and reliable (Schoonenboom and Johnson, 2017). 16 papers included solely quantitative data and 2 papers considered only qualitative data.

Data captured was namely error/accuracy and completion times (as objective metrics for assessments/comparisons). Subjective responses were usually collected via custom or industry-standard questionnaires (such as NASA-TLX (Index, 2020), System Usability Scale (SUS) (Brooke, 1995) and User Experience Questionnaire (UEQ) (Schrepp, 2015)). These were generally quantified for analysis alongside objective measures. Many studies also captured more in-depth subjective feedback in the form of interviews, recorded observations and think-aloud protocols. Elicitation studies primarily quantified subjective agreement rates to define a consensus of user-defined gestures.

3.3.2 Handheld Display

Data captured for studies that considered solely handheld display devices, namely smartphones and tablets, is detailed in the following subsections. An overview of the data can be found in Figure 3.1.

Study Type

All studies employing solely a handheld device addressed a specific parameter as a factor for assessment, to explore the influence of output or approach on interaction performance. This included how a pointer or cursor is indicated or behaves (Yin et al., 2019; Perea, Morand, and Nigay, 2020), where the user performs the gesture (front or back of the device) (M. Kim and J. Y. Lee, 2016), the impact of task on interaction (Yin et al., 2019; M. Kim and J. Y. Lee, 2016; Ye et al., 2020; Samini and Palmerius, 2016; H. Bai, G. A.

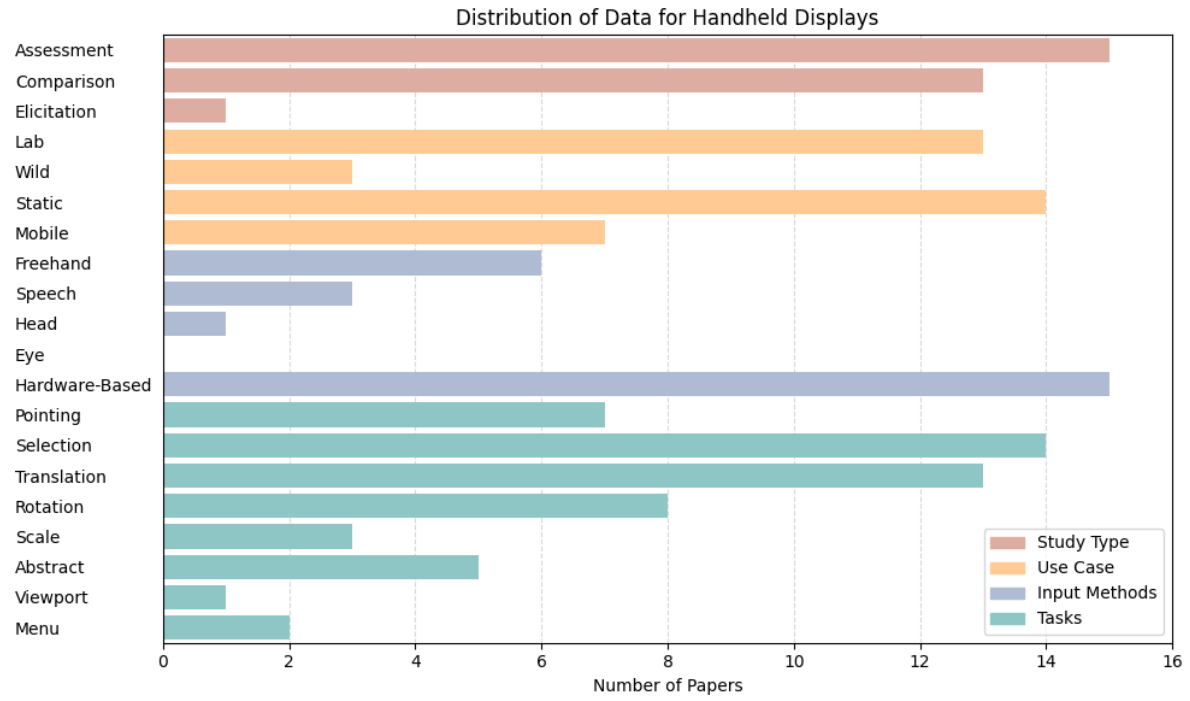

Figure 3.1: Distribution of data for the 15 papers considering solely handheld displays (focusing on study type, use case, input methods and tasks).

Lee, et al., 2014; Su, Sunar, and Ismail, 2020; Mossel, Venditti, and Kaufmann, 2013), or the size/distance of an interactive element (Yin et al., 2019; S. Mayer, Laput, and Harrison, 2020; J. Qian et al., 2020). 8 of the papers also explored the benefits of a novel technique or interface.

All comparison studies used the input method as a variable, however, Tanikawa et al. (Tanikawa et al., 2015) also considered the effect of display devices, by comparing a smartphone with a tablet. Furthermore, Kim and Lee (M. Kim and J. Y. Lee, 2016) explored to what extent a wide-angle lens improved usability (enhancing FOV). Although tasks were performed under AR conditions using handheld for all of the comparisons, in some instances (S. Mayer, Laput, and Harrison, 2020; Tanikawa et al., 2015; H. Bai, G. A. Lee, et al., 2014; Nazri and Rambli, 2015; Perea, Morand, and Nigay, 2020), touchscreen input was also used as a baseline to observe the effectiveness of other inputs (such as freehand gesture or multimodal approaches).

The elicitation study employed motion gestures in 6-DoF, where participants were

asked to define motions to control an augmented character (first by manipulating a human-like doll and then a mobile device (Ye et al., 2020)). The gestures were later implemented within a novel interface for assessment, using the smartphone display. 6 papers offered solely fundamental contributions, whereas 5 papers considered their contribution on a general scale, as well as applying it to a specific application. There were 4 instances where the research was exclusively application based (Ye et al., 2020; Frank, Moorhead, and Kapila, 2016).

As highlighted in figure 3.1, studies were predominantly conducted in a controlled environment, under laboratory conditions. However, 3 studies explored interaction in the wild. Here, S. Mayer, Laput, and Harrison (2020) conducted their research in an outdoor environment for part of the experiment, Y. Wei, Orlosky, and Mashita (2021) in a large office building with emloyees, and Raeburn and Tokarchuk (2021) expored immersive storyelling in home environments. Participants were also asked to remain static for the majority of studies. Where free motion was permitted during testing, 4 studies considered device motion and trajectories. No papers were found to report on human-motion data.

Input Methods

In line with the review by E. S. Goh, Sunar, and Ismail (2019), the majority of studies employed hardware-based input via the handheld device itself. The touchscreen display was used in all studies for at least one condition (i.e. for interaction with GUI elements and for intuitive object manipulation via touchscreen legacy gestures (Frank, Moorhead, and Kapila, 2016)). 7 papers also considered manipulation of handheld displays for explicit interactions, with almost half of the studies implementing freehand interaction. As illustrated in figure 3.1, speech and head-based inputs were explored least.

Some studies reported on novel interaction approaches that discussed at least two types of input. For example, touch and hand were compared (H. Bai, G. A. Lee, et al., 2014; M. Kim and J. Y. Lee, 2016; Nazri and Rambli, 2015) and combined (M. Kim and

J. Y. Lee, 2016; Nazri and Rambli, 2015; Su, Sunar, and Ismail, 2020) in several papers. Hand, touch and device manipulation were also evaluated in the work of Su, Sunar, and Ismail (2020). Furthermore, J. Qian et al. (2020) compared hand gesture with dwell-based selection via device manipulation, whilst S. Mayer, Laput, and Harrison (2020) considered head-based interaction with speech and implicit hardware-based input. A single study also noted the impacts of different combinations of multimodal interaction (touch, hand, speech), with alternative output conditions (Nazri and Rambli, 2015). In total, 9 papers investigated multimodal methods, however, 7 employed solely hardware-based input (a mixture of touchscreen interaction and device movement).

Tasks

As presented in figure 3.1, the task most often observed with handheld displays was selection, closely followed by translation. Approximately half of the studies explored rotation and pointing tasks. Abstract and scaling tasks were considered by close to a third of studies, whilst menu and viewport manipulation via a specific function (manipulating displayed content based on users' POV (Samini and Palmerius, 2016)), were examined least.

In terms of the input methods used to complete the different tasks, 12 papers implemented touch interaction for selection. Physical movement of the device with 6-DoF was generally employed for explicit pointing and manipulation tasks, using some kind of visual indicator (i.e. a rod, cursor or raycast (Yin et al., 2019; Samini and Palmerius, 2016)), however, Tanikawa et al. (2015) only considered movements with up to 3-DoF. Gestures with the physical device were also compared with standard touch gestures for object manipulation, through techniques such as multi-touch interaction (M. Kim and J. Y. Lee, 2016).

Object manipulation tasks were achieved by combining touch with physical device movement in 6 papers (where touch triggered the interaction and movement defined the

translation/rotation/scaling axis and behaviour). S. Mayer, Laput, and Harrison (2020) went beyond hand and hardware-based interaction by implementing speech for abstract commands and head gaze for pointing. As well as this, Nazri and Rambli (2015) assessed how users freely employ different forms and combinations of input (speech and hand) alongside standard touch interaction, to complete a gamified task, and Raeburn and Tokarchuk (2021) considered imagined interactions, asking participants to verbalise their desired inputs through a think-aloud protocol.

The tasks were delivered differently depending on the study design. Assessments predominantly investigated predefined tasks and interaction methods, which were most often taught to participants through a training stage. Comparisons primarily observed the impact of different interaction methods on task execution, whereas the elicitation study explored user gestures based on a defined list of actions, to understand user approaches to different types of tasks.

3.3.3 Headworn Display

The following subsections elaborate on the data captured for studies considering headworn display devices in standalone. An overview of the data for headworn displays is provided in Figure 3.2.

Study Type

Of the papers represented in figure 3.2, 36 papers measured the impacts of output and 27 assessed how interaction is affected by different tasks. Changes in output were notably related to the size or distance of virtual content, which was explored in 21 of the publications. 30 assessments also concerned novel applications or techniques. 8 papers reported on the number of fingers/hands employed for mid-air interactions.

Few papers recognised factors surrounding environmental conditions. Only 2 pa-

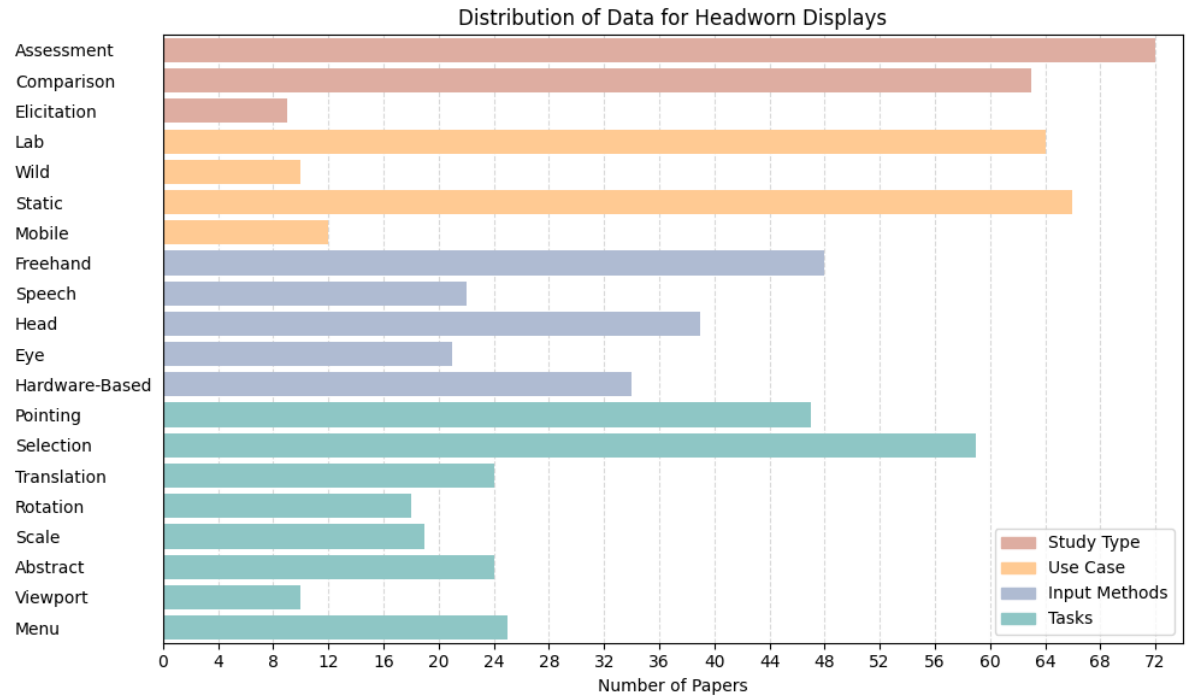

Figure 3.2: Distribution of data for the 72 papers considering solely headworn displays (focusing on study type, use case, input methods and tasks).

pers were found to report on the influence of lighting when interacting indoors and outdoors (Manuri and Piumatti, 2015; Brancati et al., 2018), one of which also discussed the impact of ambient noise levels (Manuri and Piumatti, 2015) when employing speech input. 5 papers were found to report on longitudinal studies to assess learning curves (Pourmemar and Poullis, 2019; X. Lu et al., 2019).

Where factors were also compared, 56 studies discussed different interaction methods or techniques. 2 of these studies examined the device type, where different interaction form factors were explored (Frutos-Pascual, Creed, and I. Williams, 2019; Tung et al., 2015). Alallah et al. (2018) also compared the affects of input and from which point of view (performer vs observer).

Elicitations were again considered least. These studies were related to small target selection (Bhowmick, Kalita, and Sorathia, 2020), multimodal interaction (limited to speech and gesture (X. Zhou, Adam Sinclair Williams, and Francisco Raul Ortega, 2022; Adam S. Williams, Garcia, and F. Ortega, 2020; Adam S. Williams and Francisco R. Or-

tega, 2020)) and gesture interaction (Blaga et al., 2021; Pham et al., 2018; Piumsomboon, Clark, et al., 2013), for manipulation tasks or animation in VR (Arora et al., 2019). Tung et al. (2015) considered more general approaches to input, exploring multiple modalities for using smart glasses for game interactions in public spaces (Tung et al., 2015).

The majority of studies were conducted under controlled laboratory conditions, with few papers reporting on results gathered in more realistic environments. 2 papers addressed both controlled and uncontrolled conditions. Studies conducted 'in the wild' were primarily related to specific use cases (i.e. a cultural heritage site (Brancati et al., 2018), care home (Prilla, Janßen, and Kunzendorff, 2019) or in an industrial environment (Väyrynen et al., 2018; Pringle et al., 2019)), with Alallah et al. (2018) exploring fundamental interaction in public spaces. Participants were again asked to remain static for most studies. 6 studies were found to consider both static and mobile conditions.

Input Methods

As shown in figure 3.2, the input method explored most with headworn displays was freehand interaction. This was followed by head-based and hardware-based input, which were both included in more than half of the papers. Speech and eye interaction were explored least, but still relatively frequently.

Multiple input types were explored in most of the studies, with 14 papers reporting on a single input modality. The publications were mostly observing 2 input methods (in 33 papers), or 3 input methods (in 17 papers). These input methods represented different permutations of hand, head-based, eye-based, speech and hardware-based inputs. 41 papers applied at least one combination of multimodal input (i.e. to decouple inputs to complete distinct tasks (Munsinger, White, and Quarles, 2019) or to couple inputs to improve the accuracy of interactions (Henrikson et al., 2020)). 9 papers used multiple inputs solely in comparison as individual techniques.

The most frequent multimodal input combination was head with a hardware con-

troller, which was included in 17 papers. This was followed by hand with head, which was explored in 15 papers. Hand with speech, Hand with Eye and Eye with hardware controller were all considered in 11 papers. Head input with speech was also explored in 7 papers. Some studies considered multiple combinations of hand, head, speech and hardware-based inputs. For example, Pathmanathan et al. (2020) compared combinations of head and eye pointing and gesture and controller input for object manipualtion tasks. Tung et al. (2015) also investigated how users naturally choose to apply hand, gaze and speech inputs in public spaces.

Furthermore, 13 papers concerning head, eye or speech input discussed how systems could adapt for hands-free interaction approaches. This predominantly included applications for healthcare or maintenance (Sadri et al., 2019; Prilla, Janßen, and Kunzendorff, 2019; Bailly, Leitner, and Nigay, 2019; Lamberti et al., 2017; Väyrynen et al., 2018), where users are generally required to operate their hands to complete real-world tasks, and for text entry, where it may be inconvenient to use an external controller, or look down to type on a smartphone (X. Lu et al., 2019; W. Xu, Liang, He, et al., 2019).

Tasks

As depicted in figure 3.2, selection tasks were again the most widely reported. However, for headworn displays, this was followed by pointing. Menu-based interactions, abstract tasks and translation were addressed by a similar number of papers, followed by scaling and rotation, whereas viewport control was explored considerably less.

24 studies investigated a combination of 3 different tasks and 22 explored 2 tasks. 8 studies were found to employ 4 tasks, with those considering more than this predominantly being elicitation studies (Tung et al., 2015) or purpose-built applications (Marini et al., 2024). 6 studies reported on a single task, this included evaluating more abstract, indirect interactions; either assessing multimodal input, namely voice and hand gestures for interacting with a virtual character (Z. Chen et al., 2017), fundamental selection (Mif-

sud et al., 2022) or evaluating speech and conversational interfaces for indoor wayfinding and navigation in AR (J. Zhao et al., 2020).

In terms of multimodal input methods, where head or eye gaze was combined with a controller, it was generally to separate functions for pointing and selection mechanisms, i.e. using gaze to identify an object of interest, through a form of visual output (such as a raycast), and an external binary form of input (such as hand-gesture or controller (Özacar et al., 2016)) to confirm the interaction. As previously highlighted, speech was generally employed to accompany freehand or head-based interactions. Papers often explored the affects of unimodal and multimodal combinations on task performance and usability.

Freehand gesture was considered for selection in 39 papers and for direct controls for canonical tasks, such as translation, rotation and/or scaling, in 29 papers. In 2 papers, freehand interaction was also used for indirect gesture controls to provide instructions to virtual avatars (Z. Chen et al., 2017; Tung et al., 2015).

Speech was predominantly used for abstract tasks (applied in 13 papers to trigger discrete interactions). Head input was employed in 12 papers and eye input in 7 papers for menu-based interactions, with 13 papers found to apply dwell for selections in at least one condition. Head input was also used for abstract interactions in 10 papers and object manipulation in 9 papers. In some cases, head gestures such as nods, shakes and tilts were utilised to manipulate an interface or virtual object (Prilla, Janßen, and Kunzendorff, 2019; Pereira et al., 2017). Eye was explored for object manipulation in 6 papers (Pathmanathan et al., 2020; Pfeuffer, B. Mayer, et al., 2017).

Where viewport control via a specific function was employed, 7 papers were related to VR interaction (Bhowmick, Kalita, and Sorathia, 2020; Yu et al., 2019; H. S. Lee et al., 2024) and 3 to AR interaction. Applications of viewport control for AR covered map exploration (Satriadi et al., 2019) and game input (Tung et al., 2015). It was also employed for an elicitation study, where users chose to manipulate the scene to interact with distant objects (i.e. metaphorically zooming/pulling objects towards them)

as opposed to physically approaching interactive content (Pham et al., 2018).

3.3.4 Multiple Display Types

Finally, we highlight the data captured from studies that considered multiple display devices (which is presented in Figure 3.3). These are classified as handheld/headworn, handheld/monitor and headworn/monitor.

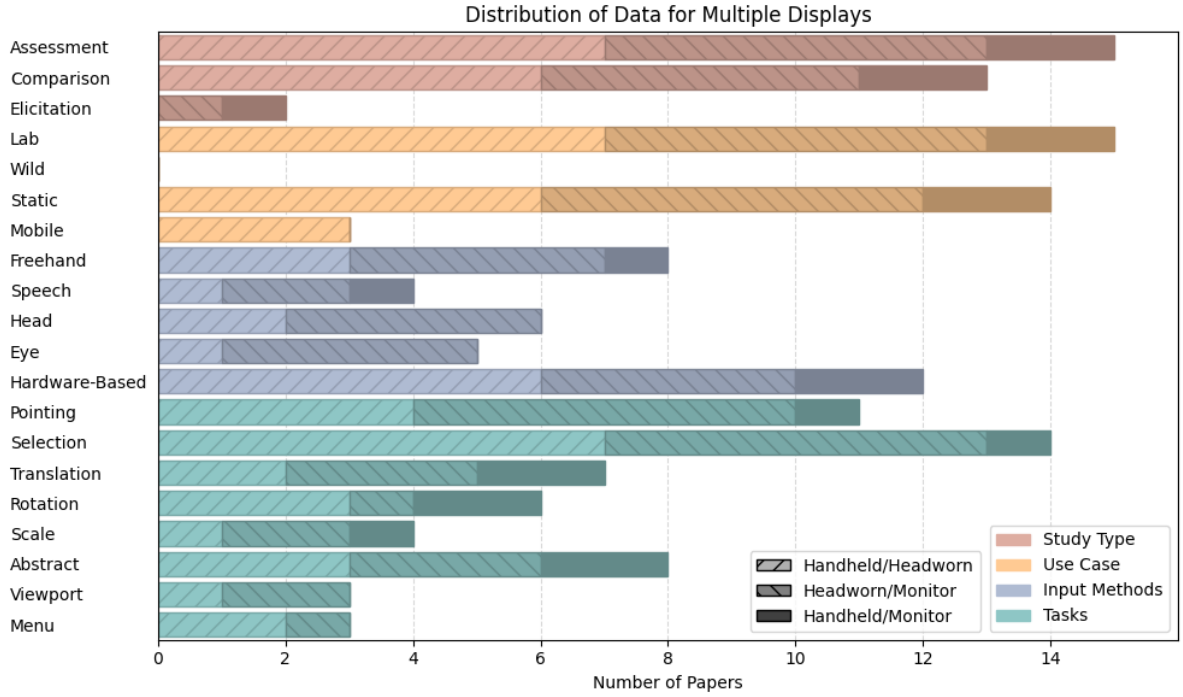

Figure 3.3: Distribution of data for the 15 papers considering multiple displays, broken down by handheld/headworn (7 papers), headworn/monitor (6 papers) and handheld/monitor (2 papers) - focusing on study type, use case, input methods and tasks.

Study Type

Where multiple display types were included in studies, they were explored in combination in 9 instances and comparison in 6.

The multi-platform combination that was most frequently employed was handheld/headworn. 3 studies combined devices in tandem for a seamless multi-platform experience (Henderson, Ceha, and Lank, 2020; Zhu and Grossman, 2020; Waldow et al.,

2018). The remaining papers compared headworn and handheld interactions (Marques et al., 2020). Assessments primarily concerned the influence of task, which was addressed in 6 papers, and output, which was explored in 4 papers. However, 1 study investigated in-pocket text input by the thigh (Henderson, Ceha, and Lank, 2020) and another considered how walking different path types affected interaction whilst in locomotion (Ghosh et al., 2020).

All headworn/monitor studies were assessments and comparisons, apart from N. Wang, Zielasko, and Maurer (2024) who conducted an elicitation to explore how users approach transfering 3D virtual objects between a standard monitor and AR HWD in a cross-reality session. Y. Y. Qian and Teather (2017) discussed the impact of target distance and input methods on interaction performance, employing a desktop set-up in combination with smart glasses to complete the assigned task. Z. Wang et al. (2020) also combined both types of display to compare unimodal and multimodal interaction. Although N. Wang, Zielasko, and Maurer (2024) permitted users to use any of the interaction methods considered, the work of Z. Wang et al. (2020) was the only study to explore the impacts of using 3 inputs simultaneously (eye gaze, gesture and speech).

Xuanhui Xu et al. (2022) compared the platform (desktop or VR) and level of interactivity for assessing veterinary students via multiple choice questionnaires. Here, participants' personal devices were used for the baseline and Oculus Quests (1 and 2) were distributed between participants for the VR condition. Bothén, Font, and Nilsson (2018) also compared a standard desktop-based gaming experience to a VR game, which employed head gaze for interaction. Similarly, Heydn et al. (2019) compared target aquisition, pointing and shooting techniques for games across desktop and VR environments. Although Xuanhui Xu et al. (2022) designed and modelled their virtual environment based on the structure of an actual veterinary anatomy lab, the study was still conducted in a lab environment, with no studies considering real-world interaction environments.

One of the papers considering monitor/handheld reported on an elicitation study, where a TV monitor was used to display referents and the handheld display to design

interactions (Dong et al., 2020). The other paper compared an immersive desk, monitor-based set-up, to interaction on a tablet or smartphone (Bazzaza et al., 2014), for interacting with an educational AR magazine. Both studies assessed how the task affects interaction methods and subjective preferences.

Input Methods

The general trends for input devices used with combined displays are shown in Figure 3.3. Where handheld/headworn was explored, hardware-based interaction was used most frequently. This was followed by hand, head and speech, respectively. 2 papers explored multimodal interaction (Waldow et al., 2018; Zhu and Grossman, 2020), both of which implemented hand and hardware-based input. Waldow et al. (2018) also investigated using head-based input alongside gesture for constrained object manipulation.

Where a monitor was used with a headworn device, 4 papers considered multimodal input. N. Wang, Zielasko, and Maurer (2024) explored all input types, uncovering how users employ different combinations of interactions that utilise multiple modalities. Another considered head input in comparison and conjunction with eye gaze (Y. Y. Qian and Teather, 2017). Z. Wang et al. (2020) explored how the number of inputs affects performance metrics, comparing different combinations of eye gaze, hand gesture and speech. Heydn et al. (2019) considered different combinations of inputs beside speech, whereas Xuanhui Xu et al. (2022) compared controller input with freehand interaction. Finally, Bothén, Font, and Nilsson (2018) examined head gaze and standard interaction with an Xbox controller, observing how the task and gaming experience of participants impacted objective and subjective results.

For handheld/monitor, Dong et al. (2020) employed both touch-based surface and hardware-based (6-DoF) gestures, whereas Bazzaza et al. (2014) considered multimodal interaction, as a combination of hand, speech and hardware-based input.

Tasks

In handheld/headworn conditions, selection was reported in all papers. This was followed by pointing in over half of the papers. Rotation and abstract tasks were explored slightly less, ahead of translation and menu-based interaction. Viewport and scale tasks were considered least. The 2 papers that investigated multimodal interaction, with handheld/headworn devices in combination, incorporated the most number of tasks. Zhu and Grossman (2020) employed all tasks except scale, Waldow et al. (2018) considered 4 tasks (pointing, selection, rotation and scale). In the only instance that head-only input was recorded (Esteves, Verweij, et al., 2017), pointing, selection and menu were considered. Similarly, the only paper involving speech input concerned abstract tasks (for text editing), which is in line with most speech-based, headworn conditions.

As shown in figure 3.3, all headworn/monitor conditions considered pointing and selection. The studies concerning speech input both included abstract interactions, with Xuanhui Xu et al. (2022) reporting the only instance of a rotation task in this category. Bothén, Font, and Nilsson (2018) explored translation, menu-based and viewport tasks, representing interactions such as aiming, shooting and walking within a VR game. Translation was also explored by Xuanhui Xu et al. (2022) and N. Wang, Zielasko, and Maurer (2024), with scale being considered in 2 papers (Z. Wang et al., 2020; Xuanhui Xu et al., 2022).

For handheld/monitor, both studies considered translation, rotation and abstract interactions, however, Dong et al. (2020) also examined scaling, and Bazzaza et al. (2014) pointing and selection.

3.4 Conclusions and Recommendations

As we move towards consumer-level immersive applications, XR technologies will become broader and more intertwined. Input designers will need to consider in what contexts

applications are employed, and provide input techniques that are capable of adapting to users' situations, activities and surroundings within both real and virtual environments. Despite this, the interaction methods commonly employed in XR applications are still often built around fixed input mappings and tasks. While effective for controlled studies or single-purpose applications, these rigid approaches restrict the potential of XR devices to adapt to different interaction contexts (X. B. Liu et al., 2024). As XR becomes a pervasive technology embedded in everyday consumer use, interaction design will require more flexible and adaptive techniques (Davari, Stover, et al., 2024). To address this, the two reviews presented in Chapters 2 and 3 have explored how different inputs have been applied and received, for a range of AR/VR applications in different domains. This has led to the identification of trends and the primary advantages and disadvantages of input techniques, which are employed for consumer-level immersive devices.

Advantages and disadvantages were extracted from the results and discussion sections of reviewed papers. Statements reporting positive sentiment (e.g., faster task performance, reduced workload, higher usability ratings) were coded as advantages, while statements highlighting negative sentiment (e.g., reduced accuracy, fatigue, occlusion issues) were coded as disadvantages. Similar points were grouped thematically across studies to avoid repetition, and the most frequently reported themes were collated into Tables 3.3 and 3.4.

Overall, results highlight the present absence of a single uniform solution to interaction. Furthermore, due to the range of users/use cases and devices, there is an ongoing challenge for researchers and developers to apply robust logic, to seamlessly adapt inputs to tasks and scenarios. Despite this, patterns have been highlighted in this review which confirm the appropriateness of certain input modalities for XR tasks. Findings, which are discussed in more detail in the published literature review (Spittle, Frutos-Pascual, et al., 2022), suggest that the most appropriate interaction approaches can be predicted, based on valuable trends attributed to the device, task and use case.

Notably, head and eye were found to be beneficial for pointing tasks (Esteves,

Table 3.3: Mapping the most appropriate inputs to distinct tasks on handheld displays: advantages and disadvantages. Eye is not included in the table as no papers were found for handheld displays.

Input Method	Advantages	Disadvantages
Hand	• Can be combined with hardware-based techniques to provide enhanced performance for object **manipulation tasks** (translation/rotation/scale) (M. Kim and J. Y. Lee, 2016) • Intuitive to employ (M. Kim and J. Y. Lee, 2016; H. Bai, G. A. Lee, et al., 2014; J. Qian et al., 2020; Su, Sunar, and Ismail, 2020) • Able to be performed either at front or back of device (M. Kim and J. Y. Lee, 2016) • More enjoyable and immersive for close range interaction (J. Qian et al., 2020; Su, Sunar, and Ismail, 2020)	• Direct manipulation affected by hand-occlusion (M. Kim and J. Y. Lee, 2016; Yin et al., 2019) • Significantly slower than screen dwell techniques for **selection** (J. Qian et al., 2020) • Not always practical as users generally require one hand to hold the device, which may induce fatigue (H. Bai, G. A. Lee, et al., 2014; M. Kim and J. Y. Lee, 2016)
Head	• Effective for **pointing**/identifying objects and regions of interest (S. Mayer, Laput, and Harrison, 2020) • Can be referenced to decrease completion time for **Abstract** speech commands (interaction requires shorter and less precise utterances) (S. Mayer, Laput, and Harrison, 2020)	• Affected by distance/location of targets (too close or too far) (S. Mayer, Laput, and Harrison, 2020) • Requires holding the phone in an unnatural position to capture head directionality information (S. Mayer, Laput, and Harrison, 2020) • Requires experiencing a learning curve (S. Mayer, Laput, and Harrison, 2020)
Speech	• Effective for **Abstract/menu-based interactions** and has a lower workload than hand/hardware-based input (S. Mayer, Laput, and Harrison, 2020) • Can be used to improve interaction experience/provide more interaction capabilities (Nazri and Rambli, 2015)	• Requires longer, more precise utterances when used in standalone mode (S. Mayer, Laput, and Harrison, 2020)
Hardware-based	• Raycasting techniques are fast and effective for **pointing/selecting** large, visible content (Mossel, Venditti, and Kaufmann, 2013; Yin et al., 2019) • Hardware-based gestures (with 6-dof) provide an easy, natural and intuitive method for object/character control **(translation/rotation/scaling)** and can produce higher agreement rates than hand gestures (Ye et al., 2020) • Touch and motion inputs can be separated into independent mechanisms to improve usability (Su, Sunar, and Ismail, 2020; Y. Y. Qian and Teather, 2017) • Touchscreen legacy gestures are generally easy and comfortable for simple object manipulation tasks (Frank, Moorhead, and Kapila, 2016)	• Multitouch/motion gesture interaction is often more cumbersome and prone to error due to finger occlusions/sensor tracking (M. Kim and J. Y. Lee, 2016; Tanikawa et al., 2015; Ye et al., 2020) • Raycasting less effective for occluded or small targets (Yin et al., 2019; Perea, Morand, and Nigay, 2020) • Higher task load than voice/gaze input (S. Mayer, Laput, and Harrison, 2020) • Precision highly dependent on design parameters (e.g., cursor length) (Yin et al., 2019; Perea, Morand, and Nigay, 2020; Tanikawa et al., 2015) • Motion inputs often require system adaptations for usable rotation and pointing/selecting (Samini and Palmerius, 2016; Perea, Morand, and Nigay, 2020) • Touchscreen alone limits interaction capabilities (H. Bai, G. A. Lee, et al., 2014; Nazri and Rambli, 2015; M. Kim and J. Y. Lee, 2016)

Table 3.4: Mapping the most appropriate inputs to distinct tasks on headworn displays: advantages and disadvantages.

Input Method	Advantages	Disadvantages
Hand	• Most intuitive (Sadri et al., 2019; Frutos-Pascual, Creed, and I. Williams, 2019; Whitlock et al., 2018; Kang, J.-h. Shin, and Ponto, 2020) • Useful for object manipulation tasks **(translation/rotation/scale)** (Piumsomboon, Altimira, et al., 2014; Frutos-Pascual, Creed, and I. Williams, 2019) • Effective in moderation (Belkacem, Pecci, and Martin, 2019) • Accurate for **selection when content is in arm's reach** (Özacar et al., 2016; Whitlock et al., 2018) • Gesture metaphors work well for viewport control (Satriadi et al., 2019; Tung et al., 2015)	• Prone to induce fatigue (W. Xu, Liang, He, et al., 2019; Franco and Cabral, 2019; Krupke et al., 2018; Esteves, Y. Shin, and Oakley, 2020; Belkacem, Pecci, and Martin, 2019; Özacar et al., 2016) • Difficult for **abstract** interactions (Adam S. Williams, Garcia, and F. Ortega, 2020; Piumsomboon, Altimira, et al., 2014) • Challenging for small, distant, or dense content (Özacar et al., 2016; Piumsomboon, Altimira, et al., 2014) • Scaling sometimes not intuitive (Piumsomboon, Altimira, et al., 2014; Frutos-Pascual, Creed, and I. Williams, 2019) • Not ideal when time/error is critical (Sadri et al., 2019; W. Xu, Liang, Y. Chen, et al., 2020) • Lacks tangible support and affected by social acceptance (Plasson et al., 2020; Alallah et al., 2018; Tung et al., 2015)
Head	• Effective **pointing/selection** (Franco and Cabral, 2019; Krupke et al., 2018; Esteves, Y. Shin, and Oakley, 2020; Chittaro and Sioni, 2018) • Less physically demanding than hand input (Krupke et al., 2018; Esteves, Y. Shin, and Oakley, 2020) • Best for hands-free applications (Krupke et al., 2018; Esteves, Y. Shin, and Oakley, 2020; Belkacem, Pecci, and Martin, 2019) • Discreet nods/tilts work for **menu-based/abstract** interactions (Yu et al., 2019; Prilla, Janßen, and Kunzendorff, 2019; X. Lu et al., 2019) • Helps improve accuracy/prediction models (Wolf et al., 2019; Henrikson et al., 2020) • Faster than hand for **translation/scale** (Sadri et al., 2019)	• Dwell is slower and more demanding than external controllers (Franco and Cabral, 2019; Krupke et al., 2018; Esteves, Y. Shin, and Oakley, 2020; Chittaro and Sioni, 2018) • Less intuitive; short learning curve (X. Lu et al., 2019; Kang, J.-h. Shin, and Ponto, 2020) • Affected by social acceptance (X. Lu et al., 2019; Alallah et al., 2018) • **Rotation tasks** are difficult (Sadri et al., 2019)
Eye	• Fast for many **target-acquisition** tasks (Kytö et al., 2018; Blattgerste, Renner, and Pfeiffer, 2018; Meng, W. Xu, and Liang, 2022), especially for **pre-selection** in multimodal input (hand/pinch/button confirms) (Sidenmark, Z. Sun, and Gellersen, 2024; L. Lu et al., 2021; Pfeuffer, B. Mayer, et al., 2017) • Lowers physical fatigue and keeps the hands free (Pfeuffer, Mecke, et al., 2020; Bao et al., 2023; Sidenmark, Clarke, et al., 2020) • Socially acceptable and suitable for mobile or busy scenarios (e.g. walking, passengers) (Khamis et al., 2018; H. S. Lee et al., 2024; Schramm et al., 2023)	• Higher error rates than head input for small, distant or long-range targets (Y. Y. Qian and Teather, 2017; Jihyeon Lee, J. Kim, and Jeongmi Lee, 2023) • Vulnerable to involuntary activation ("Midas-touch") without explicit confirmation (Mohan, W. B. Goh, et al., 2018; Pfeuffer, B. Mayer, et al., 2017) • Prolonged fixation or dwell triggers eye fatigue and lower preference (Pfeuffer, Mecke, et al., 2020; Mathias N Lystbæk et al., 2022a) • Depends heavily on tracker quality and calibration accuracy (Pathmanathan et al., 2020; Bao et al., 2023)
Speech	• Most appropriate for **abstract interactions** (Z. Chen et al., 2017; Adam S. Williams and Francisco R. Ortega, 2020; Adam S. Williams, Garcia, and F. Ortega, 2020) • Effective **hands-free selection/menu** (Pourmemar and Poullis, 2019; Wolf et al., 2019) • Aids scaling/rotation alongside hand input (Adam S. Williams, Garcia, and F. Ortega, 2020; Piumsomboon, Altimira, et al., 2014) • Lets user focus on task, not input method (Wolf et al., 2019) • Not affected by distance to elements (Whitlock et al., 2018)	• Hard to express rotation/translation by speech (Piumsomboon, Altimira, et al., 2014; Whitlock et al., 2018) • Low social acceptance (Whitlock et al., 2018) • High error rates with short utterances (Manuri and Piumatti, 2015)
Hardware-based	• Allows discrete, indirect interactions (Franco and Cabral, 2019) • Provides tangible support (Plasson et al., 2020) • Outperforms others for **pointing/selecting** (W. Xu, Liang, He, et al., 2019; Franco and Cabral, 2019; Ganapathi and Sorathia, 2018) • Often least tiring for selection (Franco and Cabral, 2019; Krupke et al., 2018)	• Requires extra hardware (less practical) (W. Xu, Liang, He, et al., 2019) • Less accurate for selecting distant content (Whitlock et al., 2018)

Verweij, et al., 2017), hand for object manipulation (Sadri et al., 2019), and speech for abstract tasks and commands with headworn displays (Adam S. Williams and Francisco R. Ortega, 2020). A plausible mapping for head pointing interactions on HMDs is ray-casting (Yin et al., 2019), or rod techniques (Tanikawa et al., 2015) for handheld devices, and again hand interaction could be used for more intuitive and enjoyable interactions (Waldow et al., 2018; Plasson et al., 2019). 6-DoF gestures could notably be beneficial when used in conjunction with headworn devices, such as for applications in gaming (Dong et al., 2020) or for object manipulation (Zhu and Grossman, 2020), which could prove to be more usable and precise than touchscreen gestures for interaction (Dong et al., 2020).

As well as highlighting some advantages and distadvantages of input techniques for diverse tasks and use cases, the review has revealed some of the research gaps that need addressing. Based on the 102 papers reviewed, the following recommendations for future work are provided to prompt research directions. Aspects of these recommendations are also addressed within the thesis and justify the studies conducted in subsequent chapters. As defined in Section 3.7, focus is given to 1) how factors related to the activity/situation of the user will impact interaction, and 2) how interaction techniques could be employed interachangably based on this.

Test with a wider variety of user groups.

Although participant demographics and past experience was often noted, user group is not generally a primary consideration. However, different users may have contrasting preferences surrounding input techniques. By conducting user studies with a more diverse range of user groups, patterns may be presented surrounding preferences for inputs, which could make it more straightforward to adapt interaction to each user. Different user groups can be defined by considering a combination of factors, such as age, gender, cultural background, ability/disability and technology usage. Through creating mappings of how these considerations affect interaction approaches and user preferences (i.e. through tree data structures), we can work towards making immersive technologies more

personalised for individual users, and more representative of a true population.

Pay closer attention to task scenarios.

As well as considering user demographics, we should pay close attention to the scenarios that users will be applying immersive technologies. This is because results suggest that the suitability of different interaction techniques depends on the context that applications are being employed (i.e. for fun/at leisure, or for more serious tasks where time and error considerations are of high importance). By considering the context of different immersive applications, and how AR/VR technologies will be used for a range of consumer use cases, we can better understand the advantages and disadvantages of input methods. The design of applications can then be tailored, to ensure they are transferable for the range of scenarios that immersive technologies will be used.

Consider how users' activity/situation will impact interaction.

Building on the task scenario, we should also consider under what activities and situations a user will interact. Key factors such as impairments, whether permanent or due to a users situation/activity, will directly impact the most appropriate interaction techniques. Consequently, it is important to understand how users adapt behaviours and interactions, depending on their circumstances, so designers can adapt input techniques accordingly. Because immersive technologies offer a broad range of use cases, the influence of activity/situation will be important to consider, and account for, when designing interaction techniques.

Further explore environmental and social constraints.

As well as understanding the impacts of users' activity/situation, we must also explore how the environment (and how the social acceptance associated with this environment) will impact interaction preferences. Usability studies should focus on testing in, or simulating, real-world scenarios, under diverse conditions. This will help to maximise social acceptance of immersive technologies and system usability/robustness. Research should therefore be exploring how input approaches are affected by different social and environ-

mental factors. It will also be important to consider how these factors can be measured and, depending on these variables, how different input modalities can harmonise the nature and flow of interaction. Although testing in real conditions is not generally practical for scientific research, it is important to deliver more theoretical studies that focus on the future of interaction with these technologies. By understanding how different variables related to society and environment impact interaction, we can design input techniques that are more appropriate for realistic use cases/conditions.

Consider the provisions of emerging and future technologies.

Although it is important to research what is currently achievable, we should also be considering what we expect to be possible with XR technologies in the future (keeping this suggestion in mind will also help to address all of the recommendations provided). This could be achieved by designing studies that eliminate the issues surrounding current technologies, or systems could be adapted/enhanced by modifying existing equipment. Adopting such techniques will ensure researchers are more in line with what is achievable when novel technologies are released. As opposed to recycling input approaches, we can focus on constantly making them better, as the technologies used for AR/VR are continuously improving.

Investigate how inputs/devices could be employed simultaneously.

As detailed in section 3.3, few studies have been designed to consider different combinations of input (primarily only 2 modalities), and how they can be used simultaneously, to improve usability. The findings of this review suggest that multimodal input can improve interaction by decreasing fatigue, improving system understanding and providing more interaction capabilities. Benefits of using multiple displays simultaneously were also noted, which can provide multimodal inputs across two platforms (i.e. a smartphone coupled with a headworn display). Therefore, we recommend considering how more intuitive forms of interaction, such as hand gesture and hardware-based input, can be best used alongside inputs like speech and head/gaze, for different tasks in immersive environments.

Investigate how inputs/devices could be employed interchangeably.

Although we recommend that multimodal inputs should become more widely explored, to utilise all forms of input inherent to consumer devices more frequently, multimodal input is not always required/useful for all types of interaction. Therefore, it is also important to understand how to balance the use of unimodal and multimodal inputs, to maximise the effectiveness, usability and flow of interactions. Even though different inputs are more suited to certain tasks (as defined in tables 3.3 and 3.4, and discussed in Chapter 2), it is important to consider how to best employ techniques interchangeably, to minimise negative consequences such as fatigue, frustration and cognitive load. This also applies to different devices (i.e. some tasks are more suited to handheld displays and others better employed with headworn displays). Although the benefits for using inputs interchangeably have been highlighted, it is still unclear how to best design and employ different input methods across a range of use cases.

Further explore similarities/differences between AR and VR interaction.

Exploring to what extent AR interaction is transferable to VR (and vice versa) is another important research direction. Although there are differences between AR and VR which affect interaction, they also require the consideration of very similar factors, especially regarding input methods and tasks. Therefore, it will be interesting to highlight and explore the factors that impact the appropriateness of different interaction techniques in AR and VR. Researchers can then better establish to what extent a common set of interaction guidelines could be mapped and adopted for the spectrum of XR technology.

Revisit approaches to Elicitation studies.

To effectively understand how different input modalities can be employed simultaneously and interchangeably, for different XR environments, we must reconsider how interactions are designed and delivered. This can be achieved by employing carefully designed elicitation studies, that go beyond providing standard referents, to place minimal restrictions on users. By exploring how a range of users adapt their input choices and behaviours (in different representative scenarios, environments and conditions), we can begin to under-

stand how to adapt system behaviours accordingly.

3.5 Limitations

Although reviewing a corpus of 102 papers has provided an overview of the trends surrounding explicit interaction, this research (as with other reviews) is limited by the search criteria, the databases employed and the publication dates included.

Furthermore, the review does not consider the citation count for particular papers and therefore the potential significance of each paper discussed. If this were considered, papers deemed most influential could be prioritised and potential richer insights found. However, as citations accumulate over time, it is most likely that this approach would exclude, or negatively bias, more recent papers (which could prove influential in future development of immersive technologies (Dey et al., 2016)).

Sample size for each study was also considered but was not used as part of the inclusion/exclusion criteria. Again, this may have impacted the potential significance of the results, however, we believe that this leads to a more representative review of publications.

Another possible limitation is that both AR and VR technologies were considered for the review. Although these technologies share many similarities, especially surrounding input techniques, their differences will impact users' preferences and approaches (due to factors such as the provided level of embodiment/awareness and variations in interaction approaches with real and virtual content).

Finally, owing to the proliferation of some input paradigms (notably hand/manual input) the review has a higher number of studies using specific devices/inputs. While this may inherently skew or bias some of the findings, it is representative of published data. However, we still recommend further exploration of alternative modes for inputs

(i.e head, eye, speech) for future research in immersive technology.

3.6 Summary

Despite the limitations highlighted above, this review (alongside Chapter 2) has addressed RQ1. 102 papers were systematically reviewed, examining explicit, peripheral-free interaction techniques across immersive technologies, and mapping them against display types, input methods, study types, use cases and tasks. The review identified how commonplace peripheral-free input techniques are being explored, and highlighted their advantages and disadvantages based on different contextual factors. From this analysis, a set of recommendations for future research were presented, offering directions for how immersive interaction could be further developed and evaluated.

3.7 Implications for Empirical Studies

While all of the recommendations outlined in section 3.4 are valuable for advancing interaction design in immersive environments, this thesis narrows its scope to exploring factors related to two recommendations in particular:

Consider how users' activity/situation will impact interaction: Although this recommendation encompasses a broad set of factors and circumstances, here it is considered based on users' pose (sitting, standing), distance to content (near-field within arm's reach and far-field beyond arms reach across proxemic zones), and locomotion behaviours (how users choose to move and reposition themselves in room-scale environments). These aspects align with the proxemic dimensions of distance and movement, which prior research has shown to be crucial considerations when interacting with spatial technologies (Greenberg, Marquardt, et al., 2011; Ballendat, Marquardt, and Greenberg, 2010; Huang et al., 2022).

Investigate how inputs could be employed interchangeably: While interchangeability could extend across many modalities and devices, this thesis focuses on freehand gestures, head-gaze, and eye-gaze, comparing how different unimodal techniques afford selection across different distances on HWDs, as well as how users employ locomotion with different techniques. These modalities are widely supported in XR (Hertel et al., 2021; Gallardo et al., 2023), and comparing them will highlight their strengths and weaknesses, leading to informed recommendations for how techniques could be designed and adapted based on two key contextual factors: interaction distance and user movement.

Although the systematic review considered both AR and VR, the empirical work that follows focuses specifically on AR. Unlike VR, where interaction takes place in fully simulated environments, AR requires users to negotiate digital content in relation to the physical world (Pfeuffer, Abdrabou, et al., 2021). This makes distance and movement particularly crucial, as they impact not only how virtual content is selected but also how users physically reposition themselves within real world spaces (Huang et al., 2022). Furthermore, while the review shows that handheld displays are still widely used and valuable for AR interaction, the user studies focus on interaction with HWDs. This is because HWDs free the user's hands and capable of capturing head, hand, and eye input natively (Gallardo et al., 2023; Xiaoan Liu et al., 2025). This makes them better suited for studying how techniques could be adapted based on distance and movement than handheld devices, which tend to constrain interactions to their screens.

The three user studies conducted as part of the thesis focus on distance and movement for three main reasons. First, these two factors are fundamental to interaction in AR, directly influencing the performance and usability of input modalities such as freehand, head-gaze, and eye-gaze (Whitlock et al., 2018; Norouzi et al., 2019). Second, they can be continuously and reliably sensed by current AR devices, making them practical input streams for adaptive interaction (Gallardo et al., 2023). Third, the review revealed that despite their importance, distance and movement have sparsely been exam-

ined across a spectrum of peripheral-free input methods, in relation to input technique affordance and suitability (Spittle, Frutos-Pascual, et al., 2022).

To investigate the potential of referencing distance and movement as contextual factors, the thesis focuses on selection as the most fundamental explicit task in AR. Prior work has shown that both the distance and scale of virtual content affect how pointing and selection techniques are performed and perceived (Norouzi et al., 2019; Whitlock et al., 2018), yet it remains unclear which input modalities are most effective across different distances, how users choose to position themselves with different input techniques, or how different techniques support interaction whilst moving. The empirical work therefore examines freehand gestures, head-gaze, and eye-gaze - modalities that are inherently impacted by distance, involve both pointing and selection stages, and can be directly compared. Although speech can be employed unimodally and is recognised as a valuable peripheral-free AR input (Jaewook Lee et al., 2023), it is not influenced by distance and movement in the same way as these other modalities, and has been found most effective when used for multimodal interaction (Adam S. Williams, Garcia, and F. Ortega, 2020; T. Chen et al., 2023; Monteiro et al., 2023).

The empirical research begins in the next chapter with a controlled user study exploring freehand, head-gaze, and eye-gaze techniques for near-field selection in seated AR. This provides a foundation for understanding performance and user experience of freehand and gaze-based modalities in the intimate proxemic zone, forming the foundation for subsequent studies that expand to far-field and room-scale interaction where distance and movement play an increasingly central role.

Chapter Four

Impact of Technique on Augmented Reality Interaction (Study 1)

Contents

4.1 Introduction **82**
4.2 Background Research **83**
4.2.1 Extended Workspaces 84
4.2.2 Interaction on the go 86
4.3 Comparing Techniques for Near-Field Selection in Seated AR **88**
4.3.1 Method 89
4.3.2 Results 95
4.3.3 Discussion 99
4.4 Summary **102**

Note: This chapter is adapted from previously published work

Spittle, B., Frutos-Pascual, M., Creed, C. and Williams, I. (2025). Exploring the Impact of Distance on Extended Reality Selection Techniques. In Proceedings of the 27th International Conference on Multimodal Interaction (ICMI '25), October 13–17, 2025.

4.1 Introduction

In the previous chapters, an in-depth exploration of interaction within immersive environments was conducted. Chapter 3 highlighted a significant observation: that AR studies and applications tend to restrict users to a singular, pre-defined interaction method and context. Interaction paradigms that often mirror those used for human-human communication, such as freehand gestures and gaze are becoming commonplace (Hertel et al., 2021), reshaping how we interface with digital content. As highlighted in the previous chapter, freehand techniques have emerged as a go-to approach for interacting with virtual content in immersive environments, being adopted extensively as a flexible technique that can be designed to mimic real-world actions such as pressing, grabbing, and palming (T. Wang et al., 2021). Many consumer-grade HMDs additionally provide head-gaze as an indirect pointing mechanism (Frutos-Pascual, Gale, et al., 2021), with eye techniques also gaining increased attention, demonstrated by the implementation of eye-gaze into state-of-the art consumer-level headsets such as the HoloLens 2, Meta Quest Pro and Apple Vision Pro (Gallardo et al., 2023).

The expected proliferation of AR technologies presents more diverse use cases, necessitating further exploration of alternative interaction techniques, and consideration for how they can be best employed across a range of interaction scenarios (Grubert et al., 2017). As highlighted in Chapter 3, a key consideration surrounding the use case is the user's pose/activity, or whether they are interacting whilst seated, standing, or walking (Lages and Doug A. Bowman, 2019; Spittle, Frutos-Pascual, et al., 2022; Andersson and Y. Hu, 2023). Despite this, there is currently a lack of understanding around the advantages and limitations of interaction techniques in these contexts.

To begin considering how to maximise the advantages and limitations of commonplace AR input methods, this chapter examines the applicability of freehand and gaze-based techniques for seated AR experiences. Despite the extensive application of seated near-field interaction, such as for work (McGill, Kehoe, et al., 2020), entertain-

ment (Ng et al., 2021), and gaming (Tung et al., 2015), a gap in knowledge remains when understanding and optimising AR input techniques in different interaction contexts. To address this, a study involving 32 participants is presented, which aims to highlight the effects of different interaction techniques on fundamental AR object selection tasks in a simulated "observe and interact" environment. Utilising the capabilities of the HoloLens 2, the study compares two freehand techniques, 'Press' and 'Hover,' against two gaze-based methods, 'Eye' and 'Head', for selection within the intimate (>0.5m) proxemic zone (E. T. Hall, 1966), where Press is the baseline (or the standard peripheral-free input technique employed at the time the study was conducted for near-field selection on commercial HWDs such as the Hololens 2 and Meta Quest series (Microsoft, 2023c; Meta, 2023)). The objective is to establish a grounding for how these techniques can be used interchangeably, focusing on their respective strengths and limitations. Techniques are assessed based on selection time, error rate, task load, user experience, and user preference.

The chapter is structured as follows: first, a review of near-field AR interaction research in seated environments is presented. This is followed by a description of the research methodologies employed for the study. The results are then reported and analysed, leading to the identification of key insights, with the discussion section considering the broader implications of providing AR interaction techniques in seated contexts.

4.2 Background Research

This section explores research involving seated interaction, where participants are engaged in an AR experience while sitting down. As selection is one of the most fundamental tasks performed in immersive environments, able to be employed in standalone (e.g., to select a menu component (Pourmemar and Poullis, 2019)), or as the first stage to complete more complex tasks (e.g. for indicating an object to manipulate (Adam S. Williams, Garcia, and F. Ortega, 2020)), focus is given to research involving selection, considering the range

of potential applications and use cases, and the advantages and limitations of different interaction techniques.

AR selection tasks can be employed in a range of seated scenarios, with many applications presenting interactive content at a close distance to users (<0.5m) within the intimate proxemic zone (see Chapter 1). This is commonly referred to as near-field interaction, denoting tasks performed on content within the user's reach (J. Hu, Dudley, and Kristensson, 2022).

4.2.1 Extended Workspaces

An area commonly explored is using AR for extending desktops and workspaces, where immersive applications can be employed to create personalised multi-display work environments, allowing users to multitask and access more information without the need for physical displays (McGill, Kehoe, et al., 2020). An early prototype illustrating an AR workspace environment, ARWIN (see Figure 4.1), mimicked and extended traditional desktop features such as the clock, calendar and web and file hierarchies, allowing users to spatially visualise information. By having an extra dimension in their volumetric workspace, users were able to interact with these applications more intuitively, as they could easily control physical relationships such as their proximity to virtual content. Although ARWIN employed marker-based interactions, researchers noted the value of exploring a range of interaction techniques, such as those employing gesture recognition, along with more advanced interaction and data visualisation technologies (Verdi, Nurmi, and Höllerer, 2004).

More recently, Cheng, Gebhardt, and Holz (2023) demonstrated how AR content can be adapted based on the physical characteristics of the real world, to optimise the layout of UI elements. They highlight the benefits of positioning AR interfaces relative to nearby features such as vertical walls, horizontal surfaces, or in line with physical objects and displays. Furthermore, they considered how this impacts the appropriateness

Figure 4.1: ARWin desktop (Verdi, Nurmi, and Höllerer, 2004) showing a range of applications such as a weather report, business card, web browser and clock. These are presented as they would be viewed through a video seethrough head-worn display.

of interaction techniques, with their system providing direct touch, pinching, or remote cursor control interchangeably based on the interaction context (see Figure 4.2).

Several interaction techniques have been explored for interfacing with virtual workspaces. This includes using head gaze and orientation for navigating and interacting with 2D panels (McGill, Kehoe, et al., 2020) and freehand interaction for performing a range of everyday tasks. Such tasks could involve multi-window selection and layout manipulation, knowledge work, planning, map exploration, or internet browsing (Cheng, Gebhardt, and Holz, 2023). Similar commonplace tasks were also explored by F. Lu, Pavanatto, and Doug A Bowman (2023), where focus was given to understanding the

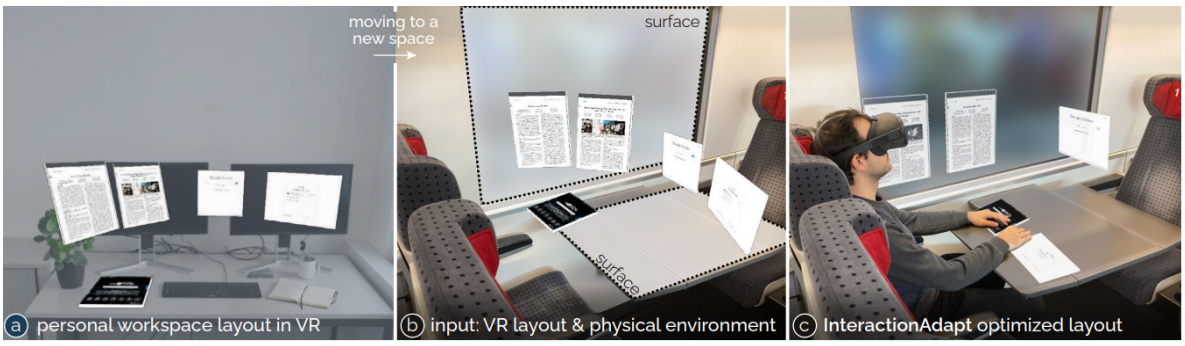

Figure 4.2: InteractionAdapt (Cheng, Gebhardt, and Holz, 2023) optimises the position and orientation of virtual elements to retain the user's workspace confguration as much as possible, and uses this information to provide direct or indirect input techniques, in different seated contexts.

practicality of eye-gaze interaction. Participants could glance at certain elements and quickly perform actions via blink or dwell (e.g., deleting an email, marking a to-do list, or setting a timer as shown in Figure 4.3). They also employed a controller to perform some interactions, such as app positioning and customisation. However, authors note the limited suitability of hand-held controllers, highlighting their impracticality in many everyday use cases, and proposing the value of exploring alternative techniques such as freehand and gaze.

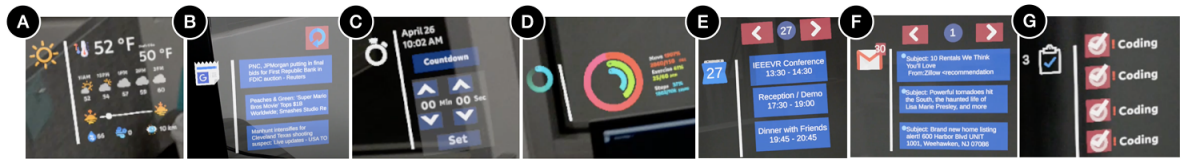

Figure 4.3: Applications considered in an interactive Glanceable AR system: (a) Weather; (b) News; (c) Clock; (d) Activity; (e) Calendar; (f) Email; and (g) Task.(F. Lu, Pavanatto, and Doug A Bowman, 2023)

4.2.2 Interaction on the go

An instance where it may be impractical to use controllers is when traveling, for example, by bus, car, train (as illustrated in Figure 4.2), or plane (Ng et al., 2021). When employing AR in homes and offices, a range of hardware solutions can be provided more easily due to the fixed nature of the environment. However, this is less ideal when users are on the go, especially when users are only remaining seated for a short duration. To increase their portability and practicality, most AR head-worn displays provide peripheral-free, touchless interactions (McGill, Williamson, et al., 2019). Despite this, existing interaction techniques, notably those employing freehand interaction, often fail to take into account the limited space users have to move around due to obstructions within the vehicle interior, such as windows, seats, and other passengers (Ng et al., 2021; McGill, Williamson, et al., 2019).

Instead, Ng et al. (2021) explored using head gaze for interface control in an aeroplane environment, emphasising the importance of providing users with flexible in-

teractions (see Figure 4.4). They note how hands-free techniques allow users to interact whilst encumbered, meaning they can be employed for a wide range of use cases that are not attainable with freehand techniques. As well as discussing the restrictions of manual interaction, especially in constrained interaction zones, McGill, Williamson, et al. (2019) also bring attention to additional physical and social implications when interacting on public transport, including the increased potential for motion sickness, and concerns around the social acceptability of interactions in shared transit.

Figure 4.4: Potential display layouts for interaction in an aeroplane environment (Ng et al., 2021).

There is a much stronger consideration needed for social acceptance when interacting with AR in public. Tung et al. (2015) looked beyond the capabilities of current sensors to explore user-defined game input for smart glasses in public settings (i.e. a coffee shop). Although some users indicated that moving a finger in front of their face (as depicted in Figure 4.5) was "weird and not socially acceptable", results show that non-touch and non-handheld interactions, notably freehand gestures, were significantly preferred over using handheld input devices. Users still tended to gravitate towards discrete interactions due to concerns with social acceptance, however, few participants were found to define interactions incorporating head or eye gaze, despite gaze being deemed as a more unobtrusive and socially acceptable technique than freehand gesture (Heo et al., 2020).

Although gaze techniques have been considered less intuitive to employ than freehand gestures (Mathias N Lystbæk et al., 2022a), which may impact the tendency for users to employ them, issues have also been presented with their social acceptance, where head-gaze has been considered as indiscreet as freehand gesture (Alallah et al., 2018). For example, if virtual content is positioned so it aligns with a bystander's physical display

Figure 4.5: Participant interacting in a public space, performing an in-air gesture to drag an object in a coffee shop (Tung et al., 2015)

(i.e. laptop or mobile phone) interacting via head gaze could appear to others as shoulder surfing, or if aligned with another person, it may seem like they are being stared at (Ng et al., 2021). This highlights the importance of displaying content relative to the real environment based on both physical and social factors.

Summary This section has highlighted a range of AR applications involving near-field selection tasks while seated. However, further investigation is required to better understand how different interaction techniques perform and how they are received by users. The following section presents a study exploring the strengths and limitations of freehand and gaze-based interaction techniques.

4.3 Comparing Techniques for Near-Field Selection in Seated AR

Building on the work discussed in Section 4.2, a study exploring the appropriateness of commonplace interaction techniques is now presented. Focus is given to fundamental selection within a seated AR environment, where objects are placed within the intimate

proxemic zone (<0.5m). The task explored is centered around a widespread "observe and interact" scenario, where a user may be referring to one interface component to gather information (e.g. a text document/set of instructions, specification, video, or design concept), before interacting with another AR component to perform a task (e.g. selecting a virtual menu panel/button or an object to interact with) (Cheng, Gebhardt, and Holz, 2023; Lischke et al., 2016). Details of the study are provided below.

4.3.1 Method

Apparatus

An application was developed for the Microsoft HoloLens 2 in C#, using Unity 2020.3.30f1 and MRTK 2.7.2.0. Windows device portal was used throughout the study to remotely open applications, guide and monitor participants, and track the stability and refresh rate of the HWD. First-person, live Mixed-Reality captures were screen recorded to aid with data analysis. Ambient lighting was also routinely monitored with a lux meter to ensure it fell within the recommended levels of 500-1000 lux (Microsoft, 2022a).

Environment

The research was conducted in a controlled, lab environment between August and September 2022. Once comfortably wearing the HoloLens and seated at the desk, participants were asked to look straight ahead at a marker whilst the application loaded. Following this, a text prompt instructed participants to select a start button to begin (using the technique explored) and the environment would alternate between the observe and interact stage as shown in Figure 4.6.

"Observe" stage For each trial, participants were instructed to turn 45° to the right and view the reference stimulus (target), which was displayed for 5 seconds based on

BT.500 guidelines (Union, 2023). Following this, a text prompt was displayed, instructing the participant to face forward and select the target from an array of cubes (as considered in previous work (Doug A. Bowman and Hodges, 1999)), at which point the trial would begin.

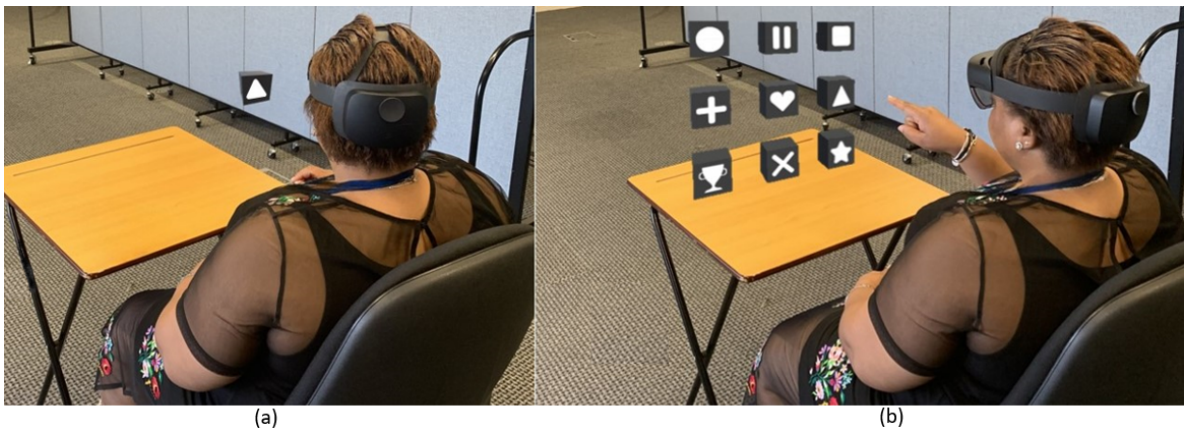

Figure 4.6: Example of the environment during (a) the observe stage, and (b) interact stage. Objects are representative and are not to scale.

"Interact" stage After the observe stage, the object array appeared, which measured 25x25cmx5cm and was positioned 0.5m away from the user. Each cube, spaced 5cm apart, had a visual angle of 5.7°. Targets were represented by 9 different symbols and the array (see Figure 4.6) was shuffled for each trial. Participants selected symbols/positions in a randomised order. The array was positioned to be within the appropriate interaction region, so all cubes were identifiable within the HoloLens 2 FOV (centred -0.12m below the horizon line relative to the height of individual participants (Microsoft, 2021a)). Selections could be made by interacting with any point of the target cube. All cubes were black by default and turned blue when targeted, with visual feedback initiated after an onset of 200ms. If a selection was successful, the target cube turned green and a confirmation sound was triggered from the MRTK sound library. In cases where participants interacted with the wrong cube, the target turned red and produced an MRTK error sound.

To generate results representative of natural interaction approaches with each technique, users could assume any hand pose they wished between trials. To minimise

data loss, a reset task button was also provided to the users left, which was used in cases when participants failed to recall the target object. After all targets were selected an "End of Condition" message appeared to prompt participants to cease the process.

Interaction Techniques

Figures 4.7 b, c, d, and e illustrate the four explored interaction techniques. These techniques were developed by referencing HoloLens 2 guidelines (as this was the device used for the study) and utilising MRTK 2.7.2.0 (Microsoft, 2022b). MRTK provides assets to support accelerated cross-platform development. Therefore, to maximise replicability, default inputs, feedback, and parameters provided by the MRTK were employed, including cursor sizes and settings. Details are provided for each of the techniques below.

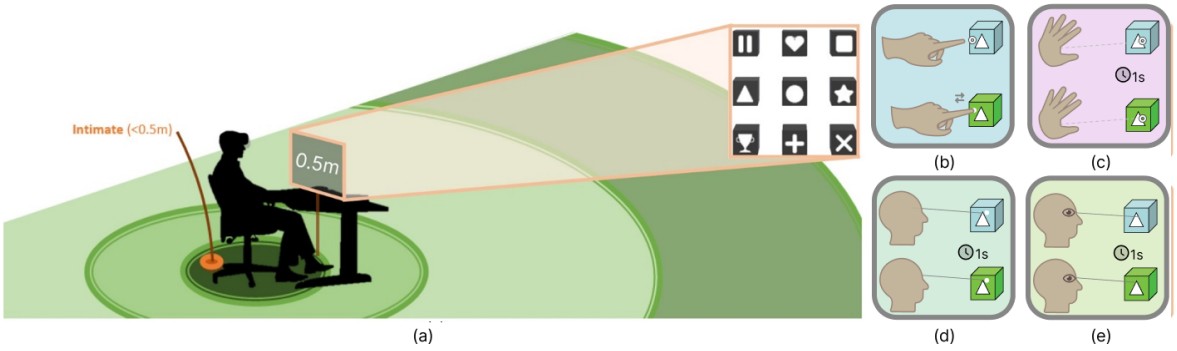

Figure 4.7: Selections were made within the intimate proxemic zone (a) using four interaction paradigms: *Press* (b), *Hover* (c), *Head* (d) and *Eye* (e).

Press required participants to position their finger over the target object before employing a "Press" gesture to execute the selection. The cursor would transition from a torus to a solid circle to communicate depth, with selections being triggered upon release (pulling pack the hand).

Hover employed a "dwell" or "hover" time, requiring that the default MRTK cursor, which was attached to the end of a ray rendered from the palm, continuously intersected the object for one second to trigger a selection.

Head employed a cursor attached to the end of an invisible ray, which extended from the HWD towards the viewing direction. Selections were triggered when the cursor had continuously intersected the object for one second.

Eye relied on visual feedback to indicate object targeting and selection. A cursor was attached to the end of a ray, which extended from the eyes toward the viewing direction. No cursor or ray was visualised with *Eye* to avoid an effect described as "fleeing cursor" (Microsoft, 2023b). Selections were triggered when participants' gaze continuously intersected the object for one second.

Dwell time for *Hover*, *Head*, and *Eye* was determined by following HoloLens 2 recommendations (Microsoft, 2023a).

Task

Participants performed a total of 9 selection tasks with each technique. They were positioned 0.5m from the object array (see Figure 4.7 a) and were instructed to employ a natural approach to select the correct target as quickly and accurately as possible. After selecting the correct target, participants would repeat the process, viewing the target object to their right (observe stage) before selecting it from the array (interact stage), until every position had been selected once.

Participants

32 right-handed participants (21 male and 11 female), aged between 22 and 44 (M = 29.28, SD = 6.01) were recruited for the study. Participants were from a population of university students and staff from the College of Computing. Based on self-reported experience, participants ranged from novice AR/VR users who had never (n = 6) or rarely (n = 8) used immersive technologies, to more experienced users who used them monthly (n = 12), weekly (n = 5) or daily (n = 1). No participants with colour blindness

and/or with corrected visual acuity below 0.80 were included in the analysis (see Section 4.3.1 for details on screening). All participants were compensated with a £10 gift voucher for their time.

Protocol

Each study session lasted around 60 minutes, with the study protocol being designed in line with local COVID-19 regulations and previously receiving IRB approval. The protocol was broken down into the following steps:

Pre-test Informed consent and demographic information, including experience, was attained from participants before each study session. After welcoming participants, the purpose of the study and test protocol was explained in detail. Following this, the researcher measured visual acuity and colour blindness via a Snellen chart and Ishihara test. Participants were then asked to take a seat at the desk and adjust and wear the HWD so it was comfortable.

Training After the experimenter explained how to employ the technique being tested, participants would experience the training phase, where they were given as much time as they required to practice and ask questions. The training simulated the main experimental task, however, only three selections were made and a different set of symbols were used. This was to prevent any artificial learning effects, where the user may have selected a target in the training and anticipated the target to appear in the same array position during the recorded trials. Participants were able to repeat the training phase, with no participants completing more than three repetitions. Each participant ran eye calibration before the first training phase to maximise the performance of the HoloLens 2. If performance was not as expected during any of the training phases, calibration was repeated.

Test After attaining verbal confirmation that participants clearly understood the technique and were comfortable with the interaction space, they were presented with the main experimental task (see Figure 4.7). All Participants completed 9 selections with each *technique*, which produced 1,152 trials (32 participants x 9 selections x 4 techniques = 1,152). After each condition, participants completed NASA-TLX and UEQ questionnaires and noted what they liked and disliked about the technique. They would then have a short break before beginning the next condition. The study followed a within-subjects design, where *technique* (counterbalanced using a latin square) was defined as the independent variable and *selection time*, *error rate task load*, *user experience* and *preference* were captured as dependent variables.

Post-test Following the completion of all four conditions, a semi-structured interview was conducted to gather subjective feedback and overall preference rankings. This was to better understand interaction approaches and contextualise what participants liked and disliked about each technique.

Metrics

The following metrics were captured for each technique:

Task Completion Times represent the duration in ms from when each trial began (following the "observe" stage as defined in Section 4.3.1) to when the correct object was selected.

Error Rate defines the number of trials where an incorrect selection was made before the correct selection.

Task Load was measured via the official iOS NASA-TLX (Task Load Index) application, where raw scores were analysed for individual subscales as well as weighted scores

for overall task load (Index, 2020).

User Experience was analysed using a standardised User Experience Questionnaire (UEQ) (Schrepp, 2015).

Preference Rankings are based on ordinal data, where participants rated the interaction techniques from best to worst.

4.3.2 Results

This section presents the results from the study, where each interaction technique (*Press*, *Hover*, *Head*, *Eye*) is compared for making selections in the intimate proxemic zone (<0.5m). As all data followed a non-normal distribution, Aligned Rank Transform (ART) (Wobbrock, Findlater, et al., 2011) repeated-measures ANOVAs were used.

After screening trial-level data (removing invalid/error trials and outliers), the remaining trial-level observations were analysed in R using ARTool (Wobbrock, Elkin, et al., 2024). The data was first aligned and ranked per effect with the Aligned Rank Transform. The aligned ranks were then fit with repeated-measures mixed-effects models including a fixed effect of *Technique* and a random intercept for *Participant*. No pre-analysis aggregation to per-participant summaries was performed. This is becuase modelling at the trial level with a participant random effect preserves within-participant dependence without inflating Type I error and avoids aggregation bias, while ART accommodates non-normal, long-tailed timing distributions and retains valid factorial tests of main effects and interactions (Wobbrock, Findlater, et al., 2011).

Overall ART F-tests for each main effect and interaction were obtained from the ART framework and, where effects were significant, pairwise contrasts used Bonferroni adjustments to control family-wise error ($\alpha = .05$). Subjective measures collected once per condition (e.g., NASA-TLX/UEQ) were analysed at the *Participant*×*Condition* level

using the same ART framework, as trial-level modelling is not applicable for single-rating outcomes.

Selection Times

Selection times were based on individual trials as described in Section 4.3.1. Trials were removed where participants made an incorrect selection (n = 8, 0.69%). Outliers were also omitted (n = 16, 1.39%), which were defined as any attempts above three standard deviations from mean selection times (mean ± 3std.).

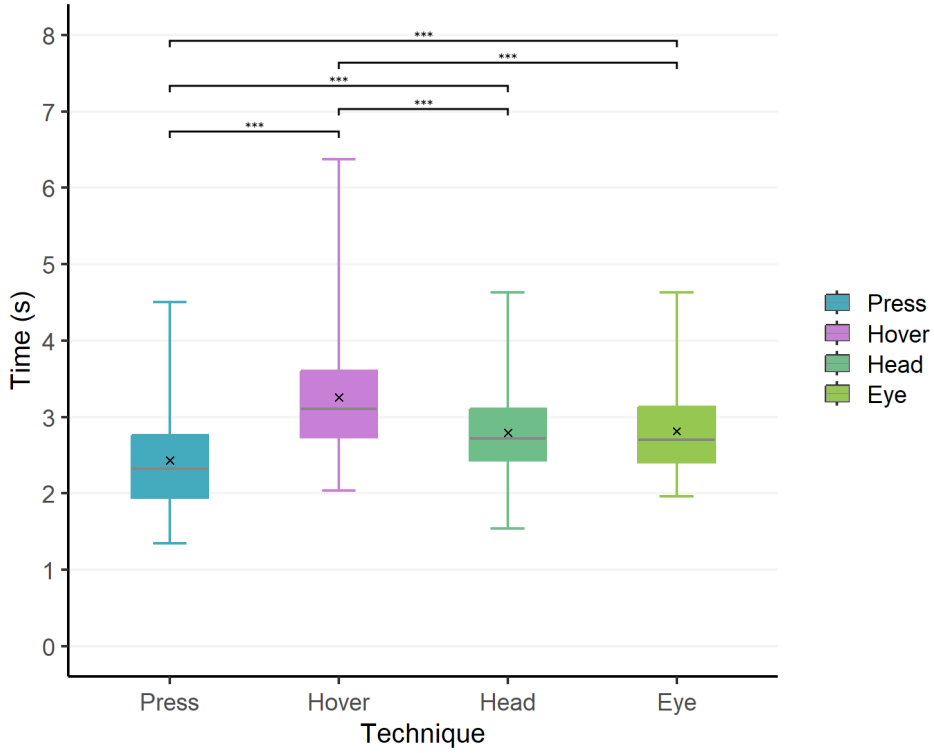

Figure 4.8: Boxplot showing distribution of *selection times* with significant differences for *Press*, *Hover*, *Head* and *Eye* in the intimate proxemic zone. Mean times are indicated by 'X'.

When analysing the resulting 1,128 trials, a significant main effect was found between *technique* and *selection time* ($F_{3,1093.4} = 93.709, p < .001, \eta_p^2 = 0.184$). *Press* was faster than all techniques ($p < .001$) with *Hover* having the worst performance ($p < .001$). No significant differences were found between *Head* and *Eye*. Figure 4.8 presents an overview of *selection time* for each interaction technique.

Error

Error rate was calculated based on the number of trials where an incorrect selection was triggered before a correct selection (N = 8). *Hover* produced the most errors within the intimate proxemic zone (N = 5), however no significant effects were found. Press, Head and Eye produced one error each.

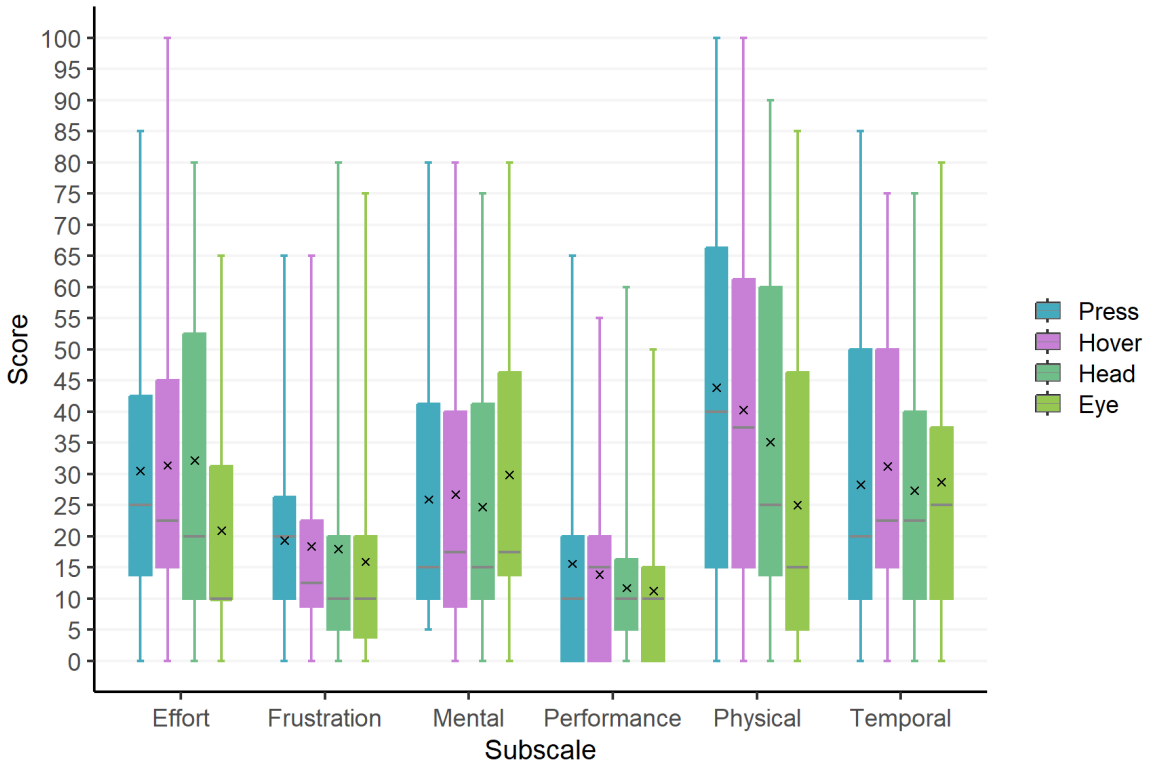

Figure 4.9: Boxplots showing distribution of NASA-TLX subscale scores for *Press*, *Hover*, *Head* and *Eye* within the intimate proxemic zone. Mean scores are indicated by 'X'. All outliers were included in the analysis.

Task Load

Eye was found to require the least *task load* (M = 25.57, SD = 16.34), followed by *Head* (M = 29.56, SD = 18.62), *Press* (M = 30.73, SD = 18.58) and *Hover* (M = 30.78, SD = 18.82). No significant main effects were found between techniques, however, ART ANOVA tests revealed differences for the Physical ($F_{3,93} = 5.57, p < .01$) and Effort ($F_{3,93} = 4.76, p < .01$) subscales. Post-hoc pairwise comparisons suggest that *Eye* is less

Physically Demanding than *Press* ($p < .01$) and *Hover* ($p < .05$). *Eye* also required less Effort than all other techniques explored ($p < .05$). A breakdown of NASA-TLX results for each technique is provided in Figure 4.9.

User Experience

When analysing UEQ Scores, ART ANOVA tests revealed that *technique* produced a significant main effect ($F_{3,93} = 3.31, p < .05$). Post-hoc pairwise comparisons show that *Eye* provided a better *user experience* than *Hover* ($p < .05$). ANOVA tests also indicated differences between *technique* for the Attractiveness ($F_{3,93} = 3.56, p < .05$), Novelty ($F_{3,93} = 3.56, p < .05$) and Stimulation ($F_{3,93} = 3.78, p < .05$) subscales. This was a result of *Eye* being rated higher for Attractiveness, Novelty and Stimulation than *Hover* ($p < .05$). Figure 4.10 provides an overview of the UEQ scores.

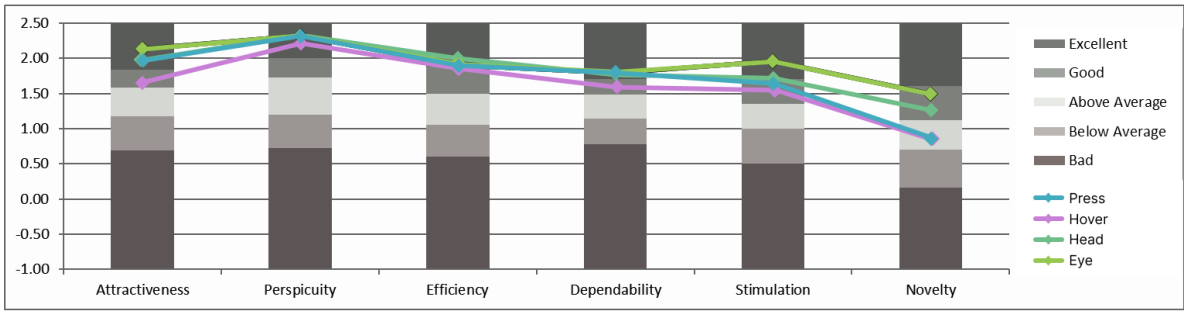

Figure 4.10: UEQ Subscale Scores for *Press*, *Hover*, *Head* and *Eye* within the intimate proxemic zone. Mean ratings from Bad to Excellent.

Preference

After completing all selection trials, participants ranked techniques in order of preference. Eye was ranked first most frequently (43.75%), followed by Head (28.13%), Press (21.88%) and Hover (6.25%).

When considering open subjective responses, Eye was often deemed as **easy/intuitive** (n = 14) - P1:"It seemed really intuitive", P11:"Eye just felt very easy", **comfortable/requiring minimal fatigue** (n = 7) - P27:"Eye gaze required little effort", and/or

efficient (n = 6) - P15:“It’s the most efficient, so I’d choose eye”.

In contrast, although Head was regarded as **straightforward to employ** (n = 16) - “Really easy, I could just point and select”, P13:“You just move your head and it identifies quickly”, the technique was also noted to be more **physical/fatiguing** (n = 12) - P19:“Head adds physical demand”.

Press interaction was viewed as **natural/familiar** (n = 11) - P2:“Press is more natural, like clicking a real button”, however, the **fatigue** experienced caused an issue for some users (n = 8) - P24:“I kept pushing and it got tiring”.

When compared to the other input methods, Hover was most often regarded as **difficult to employ** (n = 14) - P15:“The major challenge with hover is that it’s hard work using the cursor”.

4.3.3 Discussion

Study 1 explored the appropriateness of four interaction techniques (*Press*, *Hover*, *Head* and *Eye*) for performing fundamental selections within the intimate proxemic zone (0.5m). This section highlights the key findings of the experiment, discussing the design implications of employing freehand and gaze-based techniques in seated AR environments.

The most prominent finding was that preference did not correlate with performance. Although *Press* was found to provide the best overall performance, this is likely because it is not restricted by time-based selection (Mathias N. Lystbæk et al., 2022b). By employing discrete selection, users were permitted heightened control when compared to the *Head*, *Eye* and *Hover* techniques, as they were not constrained by dwell-times. Further, increased selection times with the Hover technique were primarily due to users moving their hand towards the cube to touch it, resulting in the cursor passing through the object and breaking the hover time. This suggests that visualising the raycast could be detremental to interaction performance when content is within arms reach. Although

depth perception was also an issue presented by *Press*, where users occasionally misjudged the distance of interactive content, this was far less prevelant when compared to *Hover*.

Previous studies have also found direct freehand interaction most effective when objects are comfortably within arms reach, especially when applying selection as the first step to complete more complex tasks such as object manipulation (i.e. translation, rotation or scaling) (Piumsomboon, Altimira, et al., 2014). Whereas hands are considered to be more effective for deliberate inputs, gaze techniques; especially eye gaze, have often been found more appropriate for completing initial pointing steps to indicate a target to interact with, as they naturally focus on objects that we aim to manipulate (Mathias N. Lystbæk et al., 2022b).

Although *Press* provided the best performance, *Eye* was reported to require significantly lower Physical Demand and Effort and was the most preferred interaction technique overall. As *Head* and *Eye* interactions are used implicitly to conduct search tasks, they have been found to require less Physical Demand for explicit interaction than freehand techniques (Kytö et al., 2018). Even though direct freehand interaction has often been deemed preferable for applications where the level of enjoyment, exploration or immersion are a priority (Piumsomboon, Altimira, et al., 2014; Tung et al., 2015; T. Wang et al., 2021), as the interaction context was centered around a visual search task (i.e. "Observe and Interact"), users could have found the additional effort required with Press unnecessary, preferring eye which required the least overall task load. This highlights that there is a strong consideration for the interaction context, where users do not always value performance over user experience, even when their goal is to complete the task as quickly and accurately as possible.

Although providing comparable performance, *Head* was considered to require more Effort than *Eye*, which is likely linked to the technique requiring more extreme head movements to correctly position the cursor. Contrarily, providing that the virtual content was within view, users could employ *Eye* separately to successfully select an object, reducing the extent of head motion required. This finding is in line with the research

of Blattgerste, Renner, and Pfeiffer (2018), where participants were found to only move their head using eye gaze if the target element was not visible. Table 4.1 summarises the findings from the study conducted.

Table 4.1: Advantages and disadvantages of interaction techniques in the Intimate zone (<0.5 m).

Technique	**Findings**
Press	+ Fastest technique + Interaction pace control – Most physically demanding
Hover	– Slowest technique – Worst user experience – Least preferred – Physically demanding
Head	+ On par with *Eye* performance – As physically demanding as *Press* and *Hover*
Eye	+ Least effort and physical demand + Best user experience + Most preferred

Overall, the study has highlighted the importance of considering performance and preference relative to the interaction context in which input techniques are applied. In the intimate zone, where targets are within arm's reach, discrete freehand input such as Press affords speed and control. However, participants' preference for gaze techniques demonstrates that in a context framed around a visual search task, users may value reduced effort and task load over raw efficiency. This indicates that context-aware AR systems should not only adapt techniques based on spatial factors (for example, distance to content) but also account for contextual factors surrounding the task (such as visual search versus object manipulation), as users' expectations and comfort can impact the perceived appropriateness of different modalities.

4.4 Summary

This chapter has reported a study involving 32 participants, highlighting the effects of freehand and gaze-based interaction techniques on fundamental AR object selection tasks. Results suggest that, in the simulated "observe and interact" environment explored, most users had a preference for gaze techniques over freehand, despite Press providing the best performance overall. This highlights the importance of considering how to balance performance with user experience, and provide the most appropriate techniques depending on a range of contextual factors surrounding the interaction. This includes considering not only distance, but the wider aspects relating to the virtual environment, and a users primary motivations and goals when conducting different tasks. In the next chapter, the exploration is extended to far-field interaction, where the suitability of techniques are considered for selecting distant content (which is beyond arms reach) across personal, social and public proxemic zones.

Chapter Five

Impact of Distance on Augmented Reality Interaction (Study 2)

Contents

5.1 Introduction **104**

5.2 Background Research **105**

5.2.1 Home Environments 106

5.2.2 Interaction on the go 107

5.2.3 Spectator Experiences 109

5.3 Comparing Techniques for Far-Field Selection in Seated AR **110**

5.3.1 Method 111

5.3.2 Results 117

5.3.3 Discussion 122

5.3.4 Summary 127

Note: This chapter is adapted from previously published work

Spittle, B., Frutos-Pascual, M., Creed, C. and Williams, I. (2025). Exploring the Impact of Distance on Extended Reality Selection Techniques. In Proceedings of the 27th International Conference on Multimodal Interaction (ICMI '25), October 13–17, 2025.

5.1 Introduction

Building on the initial insights presented around near-field interaction in Chapter 4, the focus is now shifted towards understanding the affordances of freehand and gaze-based interactions for far-field selection in seated AR environments. Several additional challenges have been highlighted when selecting distant AR content, including reduced visual angle of targets, ambiguous pointing, and how to effectively implement paradigm adaptations (Whitlock et al., 2018). As a result, it can become difficult to accurately interact with distant objects, and users generally experience an increase in error rates and selection times as the distance of virtual content increases (Bhowmick, Kalita, and Sorathia, 2020). This complicates the transferability of findings from near-field interaction studies to far-field interactions, necessitating a better understanding of how distance impacts a range of techniques (Whitlock et al., 2018).

To continue exploring how to capitalise on the advantages and limitations of commonplace AR input methods, this chapter reports on a second study involving 32 participants. The study aims to highlight the impact of distance on fundamental AR object selection tasks, where the baseline Press interaction paradigm considered in the previous study was modified to the baseline pinch gesture (or "Airtap" gesture as defined by HoloLens 2 device manufacturers (Microsoft, 2021b)), the standard input technique employed at the time the study was conducted for peripheral-free far-field selection on commercial HWDs such as the Hololens 2 and Meta Quest series (Microsoft, 2023c; Meta, 2023). Experiments were split into two studies to prevent extra cognitive load, where combining near and far interactions would have forced participants to switch between Airtap and Press. This would have added complexity that might have biased comparisons at this stage of research.

The study employs the same "observe and interact" environment explored in the previous chapter and again compares two freehand techniques, 'Airtap' and 'Hover,' against two gaze-based methods, 'Eye' and 'Head,', however, selections are performed

across Personal (0.5m-1.0m), Social (1.0m-4.0m) and Public (>4.0m) proxemic zones (E. T. Hall, 1966; Whitlock et al., 2018). This works towards understanding how interaction techniques can be adapted to the distance and angular size of interactive content, to further enhance the adaptability and usability of AR systems (Grubert et al., 2017). In line with the first study presented in Chapter 4, techniques are assessed based on selection time, error rate, task load, user experience, and user preference.

The chapter is organised as follows: first, an overview of research on far-field AR interactions is provided, followed by an outline of the research methods employed for the study. The findings are then presented, providing insights on the impact of distance, with the discussion exploring the wider significance of these results and how they can inform the design of far-field AR interaction approaches. The chapter ends by summarising the key findings and contributions, outlining how they inform the subsequent research conducted in Chapter 6.

5.2 Background Research

Distance has been found to affect the suitability of some AR techniques more than others, thus playing a crucial role in selecting appropriate techniques for different AR interactions (Whitlock et al., 2018; Kytö et al., 2018). In this section, focus is given to exploring selection tasks across far-field scenarios, where interactive content is placed beyond arms reach within Personal (0.5m-1.0m), Social (1.0m-4.0m) and Public (>4.0m) proxemic zones (E. T. Hall, 1966; Whitlock et al., 2018). A range of potential applications and use cases are considered, helping to highlight the advantages and limitations of freehand and gaze-based techniques in different seated interaction contexts.

5.2.1 Home Environments

Nijholt, Zwiers, and Peciva (2007) focused on how AR interaction could bring a new dimension to everyday activities in smart home environments, including those conducted while seated, where users could participate in AR interactions with mobile robots, virtual pets and virtual humans, and could retrieve, browse, and replay content in more intelligent and entertaining ways. Although Hoffmann et al. (2019) found hand gesture to be the least appropriate technique over speech and touch screen interfaces for smart home control, Vogiatzidakis and Koutsabasis (2020) established mid-air interaction with multiple home devices to be feasible, fairly easy to learn and apply, and enjoyable. They exemplified how AR could be used to interact with home devices such as air conditioning, blinds, lights, and audio/visual media systems (see Figure 5.1). Further, Cottin et al. (2016) point to the potential of combining eye gaze interaction and AR for smart home control systems. Their study focused on how this could allow users to interact with virtual controls overlaid on real-world objects, which they demonstrated by controlling LED strip lights.

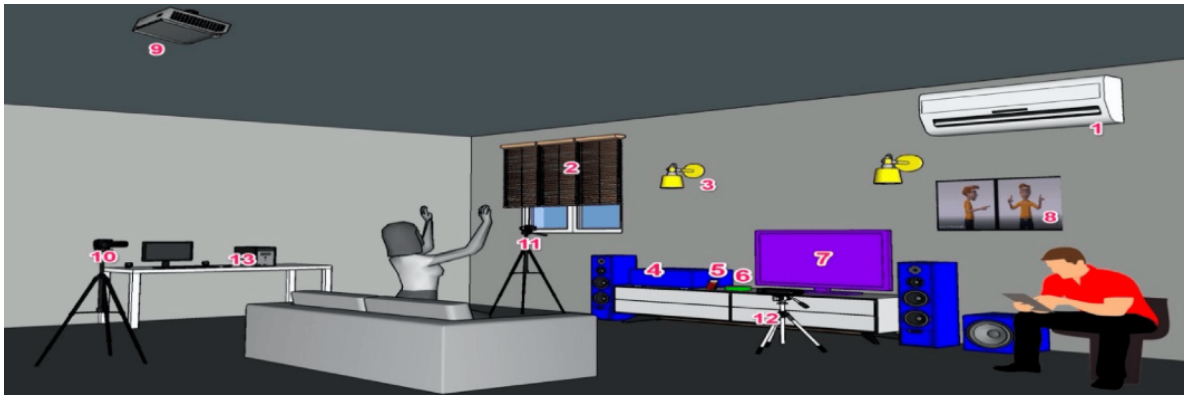

Figure 5.1: Environment mock up for an AR application to control serveral smart home devices: (1) Air Conditioner, (2) Blinds, (3) Lights, (4) Amplifier with speakers, (5) Audio player, (6) Media-video player, and (7) TV (Vogiatzidakis and Koutsabasis, 2020)

5.2.2 Interaction on the go

Users are also expected to be interacting with AR in seated contexts in public spaces, such as parks, towns and cities. This includes interactions from benches, chairs, gazebos, bus stops and rest areas, as well as when visiting shops, restaurants and cafes (Grubert et al., 2017; Abdullah, Faredzuan Mohd Noor, and Raziff Abd Razak, 2023; Regenbrecht et al., 2024) (see Figure 5.2). In public spaces, the appropriateness of interaction techniques is arguably more so affected by social acceptance than activities conducted in the home (Heo et al., 2020). This emphasises the importance of not only distance (Whitlock et al., 2018), but also the location and situation of the user in determining the most suitable techniques (X. B. Liu et al., 2024).

Figure 5.2: User interacting with a gaze and gesture interface that provides more discreet and socially acceptable interactions for public environments (Heo et al., 2020).

AR could also support interactions in a range of private and shared passenger environments, such as in cars, trains, subways and planes (Ng et al., 2021). As shown in Figure 5.3, Medeiros, McGill, et al. (2022) note that a benefit of AR technologies during travel is the provision to better arrange interactive content and displays. They highlight how mobile devices, such as laptops, tablets and smartphones, can be limited in size and lack the capacity to position content ergonomically, introducing issues such as neck fatigue and general discomfort. However, AR content can be placed and interacted with

more flexibly, within or outside of the vehicle, meaning users are not limited to direct interactions within the Intimate proxemic zone (Togwell et al., 2022). The ability to interact with content beyond arms reach, and have multiple levels of information, could therefore be highly beneficial, and permit passengers to use their time in new ways for productivity (Medeiros, McGill, et al., 2022; Ng et al., 2021) and entertainment (Togwell et al., 2022).

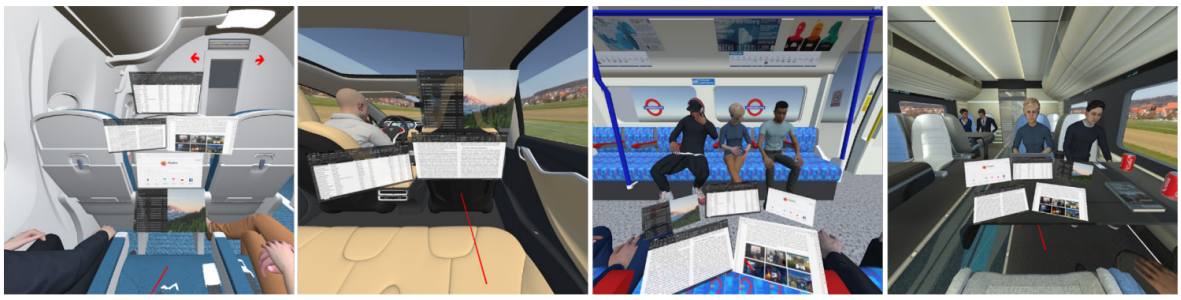

Figure 5.3: Examples of AR workspace layouts proposed by participants across four transport environments: airplane, car, subway and train (Medeiros, McGill, et al., 2022).

This concept is especially prominent as we approach the adoption of autonomous transportation, where users will be provided further freedom and versatility to interact during travel (Medeiros, McGill, et al., 2022). For example, Togwell et al. (2022) explored the potential of in-car AR games, considering how experiences could be provided within an AR vehicle context. They investigated how elements of reality (e.g. bus stops and other cars) could be appropriated into gameplay, and how the appearance of reality could be altered in relation to game events (e.g. augmented cracks in car windows). Their findings suggested that AR could allow for more engaging travel experiences and unique gaming opportunities (see Figure 5.4).

Figure 5.4: Examples of potential in-car AR games: controlling a virtual car ahead out the windscreen, and platforming on real surfaces out a side window (Togwell et al., 2022).

5.2.3 Spectator Experiences

AR also holds potential in the event space. Although visual overlays are becoming standard in sport broadcasting (Stropnik et al., 2018; Sawan et al., 2020), Zollmann et al. (2019) highlight that live spectators attending sport events within the stadium do not have access to these augmentations. They discuss how AR interaction could significantly enhance the experience of on-site spectators, demonstrating an approach that could provide information and statistics in real time. This could include names of players, pathways of individual players or groups, heat-maps showing ball possessions or activity on the field, as well as the integration of cues highlighting important aspects of the game (such as line markings in American football as shown in Figure 5.5).

Building on this, Lo, Zollmann, and Regenbrecht (2021) implemented a cursor central to the users field-of-view, enabling interaction via head gaze, where viewers could select particular elements by looking towards them. They state that this allowed for relevant information to be provided in-situ and on-demand, as opposed to viewing potentially cluttering information at all times. As well as providing augmented information to spectators, Zollmann et al. (2019) also considered how AR technologies could be employed for crowd-based visualisations and interactions. This includes interactive games or entertainment for the audience during breaks.

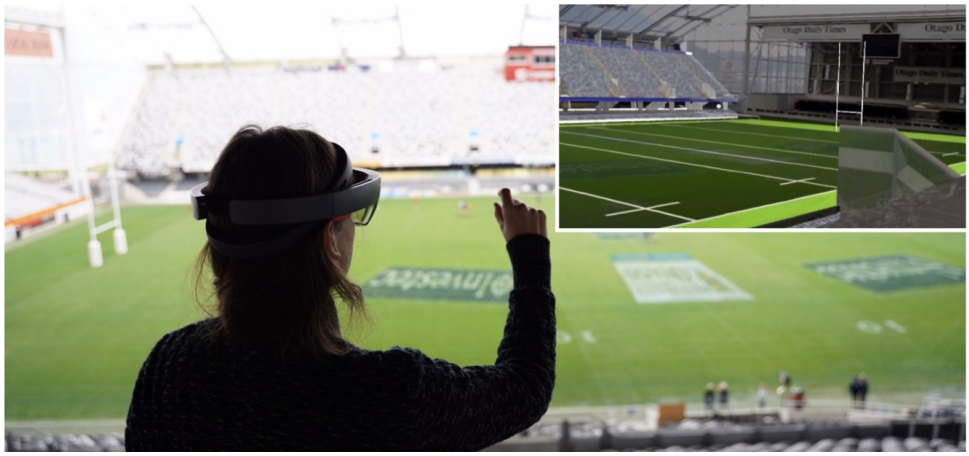

Figure 5.5: ARSpectator: example use case showing 3D structures of the stadium as well as line markings overlaid on the pitch that the user can interact with (Zollmann et al., 2019).

Similar concepts have also been explored in the realm of theatre and live performance. Pike (2023) discuss the interactive potential afforded by AR technologies, highlighting how they present more possibilities for active engagement and participation by audience members. They underscore how AR not only enables users to passively observe, but actively interact with the digital and physical elements of a performance or narrative. This could make a range of events and performances more engaging, involved and enjoyable; for example, in large venues where some audience members are seated far away and perhaps less immersed in the event (Zollmann et al., 2019; Lo, Zollmann, and Regenbrecht, 2021).

Summary This section has begun to consider some of the advantages and limitations of far-field freehand and gaze-based techniques, and a range of seated scenarios covered by literature were also highlighted. With limited research considering the impact of distance in far-field seated interaction scenarios, there is still uncertainty around the most appropriate techniques to employ when interacting in different contexts. The next section aims to better understand the appropriateness of interaction techniques at different distances in a seated environment, and how this could help provide adaptive interaction methods to users based on user-object distance (Hussain, Park, and H. K. Kim, 2023; X. B. Liu et al., 2024).

5.3 Comparing Techniques for Far-Field Selection in Seated AR

Building on the work discussed in Section 5.2, a study exploring fundamental far-field selection within a seated AR environment is now presented. The task is again centered around an "observe and interact" scenario (Cheng, Gebhardt, and Holz, 2023; Lischke et al., 2016), however, objects are placed across the Personal (0.5m-1.0m), Social (1.0m-4.0m) and Public (>4.0m) proxemic zones. Details of the study are provided below.

5.3.1 Method

Apparatus

An application was developed for the Microsoft HoloLens 2 in C#, using Unity 2020.3.30f1 and MRTK 2.7.2.0. Akin to Study 1, Windows device portal was used throughout the study to remotely open applications, guide and monitor participants, and track the stability and refresh rate of the HWD. First-person, live Mixed-Reality captures were screen recorded to aid with data analysis. Ambient lighting was also routinely monitored with a lux meter to ensure it fell within the recommended levels of 500-1000 lux (Microsoft, 2022a).

Environment

The research was conducted in the same controlled, lab environment as Study 1, between May and June 2022. Once comfortably wearing the HoloLens and seated at the desk, participants were asked to look straight ahead at a marker whilst the application loaded. Following this, a text prompt instructed participants to select a start button to begin the session.

"Observe" stage Before each trial, participants were instructed to turn 45° to the right and view the reference stimulus (target) until the text instruction to start was displayed.

"Interact" stage Following this, the 25cmX25cmX5cm array of objects was displayed. The array was world-anchored in accordance with interaction as applied in the real world (E. T. Hall, 1966; Whitlock et al., 2018). The array was placed 1.0m away from the user for interactions within the Personal zone (targets = 2.9° visual angle), 2.5m away for interactions within the Social zone (targets = 1.1° visual angle) and 5.0m away from the user for the Public zone (targets = 0.6° visual angle). Interaction and feedback were

identical to that employed for Study 1. Selections could be made by interacting with any point of the target cube, with all cubes being black by default and turning blue when targeted. Visual feedback was initiated after an onset of 200ms, and if a selection was successful, the target cube turned green and a confirmation sound was triggered from the MRTK sound library. In cases where participants interacted with the wrong cube, the target turned red and produced an MRTK error sound. Targets were represented by the same 9 symbols, and the array (see Figure 5.6) was shuffled for each trial. Targets were presented to participants in randomised order (*cube position in array* x *distance*).

The array was again positioned to be within the appropriate interaction region, so all cubes were identifiable within the HoloLens 2 FOV (centred -0.12m below the horizon line relative to the height of individual participants (Microsoft, 2021a)). Users could assume any hand pose they wished between trials. To minimise data loss, a reset task button was also provided to the users left, which was used in cases when participants failed to recall the target object. After all targets were selected an "End of Condition" message appeared to prompt participants to cease the process.

Interaction Techniques

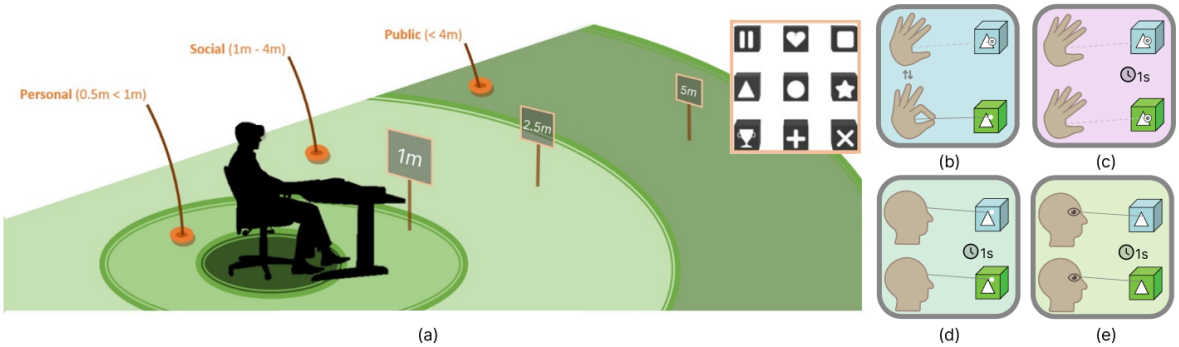

Figure 5.6: Selections were made across Personal, Social, and Public proxemic zones (a) using four interaction paradigms: *Airtap* (b), *Hover* (c), *Head* (d) and *Eye* (e). The study involved selecting 9 cubes from an array of objects with each technique, as shown in (a).

Figures 5.6 b, c, d, and e, present the four interaction techniques explored. Hover, Head and Eye techniques were identical to those employed for Study 1. For Hand Airtap

(which replaced the Hand Press near interaction paradigm considered in the first study) participants pointed a ray at the object of interest before performing an Airtap gesture to confirm the selection. The visual feedback and task presentation remained consistent with Study 1, where interaction techniques were developed by referencing HoloLens 2 guidelines (as this was the device used for the study); using the default inputs, feedback, and parameters, including cursor sizes and settings. Details are provided for each of the techniques below:

Airtap involved a two-step interaction process. First, users pointed at the object of interest, then employed a "Pinch" gesture to execute the selection. This was achieved by controlling a cursor, which was attached to the end of a ray rendered from the palm. A solid ray and cursor indicated successful gesture registration, with selections being triggered upon release.

Hover employed a "dwell" or "hover" time, requiring that the default MRTK cursor, which was attached to the end of a ray rendered from the palm, continuously intersected the object for one second to trigger a selection.

Head employed a cursor attached to the end of an invisible ray, which extended from the HWD towards the viewing direction. Selections were triggered when the cursor had continuously intersected the object for one second.

Eye relied on visual feedback to indicate object targeting and selection. A cursor was attached to the end of a ray, which extended from the eyes toward the viewing direction. No cursor or ray was visualised with *Eye* to avoid an effect described as "fleeing cursor" (Microsoft, 2023b). Selections were triggered when participants' gaze continuously intersected the object for one second.

Task

The task employed was the same as that described in Chapter 4, with all 9 cubes within the array (see Figure 5.6) being selected once at each distance. Participants were instructed to employ a natural approach to select the correct target as quickly and accurately as possible. After selecting the correct target, participants would repeat the process, alternating between the observe stage and interact stage, until every position at each distance had been selected once.

Participants

32 right-handed participants (22 male and 10 female), aged between 22 and 44 (M = 29.13, SD = 6.03) were recruited for the study. Participants were from the same population of university students and staff as Study 1 (26 participants took part in both studies). Based on self-reported experience, users ranged from novices who had never (n = 5) or rarely (n = 6) used immersive technologies, to more experienced users who used them monthly (n = 15), weekly (n = 5) or daily (n = 1). Participants were again screened for colour blindness, with analysed data being based on users who had corrected visual acuity of 0.80 or more (see Section 5.3.1 for further details on screening). All participants were compensated with a £10 gift voucher for their time.

Protocol

The protocol primarily followed that desicribed in 4.3.1, however each session lasted between 90 and 120 minutes including rest periods for participants. The study was designed in line with local COVID-19 regulations and had previously received IRB approval, being broken down into the following steps:

Pre-test Informed consent and demographic information, including experience, was attained from participants before each study session. After welcoming participants, the purpose of the study and test protocol was explained in detail. Following this, the researcher measured visual acuity and colour blindness via a Snellen chart and Ishihara test. Participants were then asked to take a seat at the desk and adjust and wear the HWD so it was comfortable.

Training After the experimenter explained how to employ the technique being tested, participants would experience the training phase, where they were given as much time as they required to practice and ask questions. The training simulated the main experimental task, however, only three selections were made and a different set of symbols were used. Participants were able to repeat the training phase, with no participants completing more than three repetitions. Each participant ran eye calibration before the first training phase to maximise the performance of the HoloLens 2. If performance was not as expected during any of the training, calibration was repeated.

Test After attaining verbal confirmation that participants clearly understood the technique and were comfortable with the interaction space, they were presented with the main experimental task (see Figure 4.6). Each participant completed 9 selections using each technique at 3 distances, which were selected to cover multiple proxemic zones (Personal, Social and Public) as considered in previous work (Whitlock et al., 2018) (Figure 5.6a defines the distances that were considered). This produced 3,456 trials (32 participants x 9 selections x 3 distances x 4 techniques = 3,456). After each condition, participants completed NASA-TLX and UEQ questionnaires and noted what they liked and disliked about the technique. They then had a short break before beginning the next condition. The study followed a within-subjects, mixed-factorial design, where *Technique* and *Distance* were defined as independent variables and *selection time*, *error*, *task load*, *user experience* and *preference* were captured as dependent variables.

Post-test Following the completion of all four conditions, a semi-structured interview was conducted to gather subjective feedback and overall preference rankings. This was to better understand interaction approaches and contextualise what participants liked and disliked about each technique.

Metrics

Identical to those reported in Chapter 4, the following metrics were captured for each technique:

Task Completion Times represent the duration in ms from when each trial began (following the "observe" stage as defined in Section 5.3.1) to when the correct object was selected.

Error Rate defines the number of trials where an incorrect selection was made before the correct selection.

Task Load was measured via the official iOS NASA-TLX (Task Load Index) application, where raw scores were analysed for individual subscales as well as weighted scores for overall task load (Index, 2020).

User Experience was analysed using a standardised User Experience Questionnaire (UEQ) (Schrepp, 2015).

Preference Rankings report how participants rated the interaction techniques from best to worst.

5.3.2 Results

This section presents the results from Study 2, where each interaction technique (*Airtap*, *Hover*, *Head*, *Eye*) is compared for performing selections across the Personal, Social and Public proxemic zones. As all data followed a non-normal distribution, ART (Wobbrock, Findlater, et al., 2011) repeated-measures ANOVAs were used.

After screening trial-level data (removing invalid/error trials and outliers), the remaining trial-level observations were analysed in R using ARTool (Wobbrock, Elkin, et al., 2024). The data was first aligned and ranked per effect with the Aligned Rank Transform. Aligned ranks were then fit with repeated-measures mixed-effects models including fixed effects of *Technique*, *Distance*, and their interaction (*Technique*×*Distance*), plus a random intercept for *Participant*. No pre-analysis aggregation to per-participant summaries was performed. This is becuase trial-level modelling with a participant random effect preserves within-participant dependence without inflating Type I error and avoids aggregation bias, while ART accommodates non-normal, long-tailed timing distributions and retains valid factorial tests of main effects and interactions (Wobbrock, Findlater, et al., 2011).

Overall ART F-tests for each main effect and interaction were obtained and, where effects were significant, pairwise contrasts used Bonferroni adjustments to control family-wise error ($\alpha = .05$). Subjective measures collected once per condition (e.g., NASA-TLX/UEQ) were analysed at the *Participant*×*Condition* level using the same ART framework, as trial-level modelling is not applicable for single-rating outcomes.

Selection Times

Selection times were based on individual trials as described in Section 5.3.1. After removing incorrect selections (n = 54, 1.56%) and outliers (n = 56, 1.62%), selection times were evaluated for 3346 trials.

Table 5.1: Task Completion Times: Results of pairwise comparisons for Interaction Technique and Object Size, where *** p=<0.001, ** p= 0.01 and * p= 0.05. Arrows indicate which technique provided the fastest selection time, H-AT: Airtap, H-H: Hover, H-G: Head or E-G: Eye

Zone		Personal				Social				Public			
	Technique	H-AT	H-H	H-G	E-G	H-AT	H-H	H-G	E-G	H-AT	H-H	H-G	E-G
Personal	H-AT		↑***	↑***	↑***	←***	←	↑***	↑***	←***	←***	←	←***
	H-H			↑***	↑***	←***	←***	↑	←*	←***	←***	←***	←***
	H-G				↑	←***	←***	←***	←***	←***	←***	←***	←***
	E-G					←***	←***	←***	←***	←***	←***	←***	←***
Social	H-AT						↑***	↑***	↑***	←***	←	↑***	←
	H-H							↑***	↑***	←***	←***	←	←***
	H-G								←***	←***	←***	←***	←***
	E-G									←***	←***	←***	←***
Public	H-AT										↑***	↑***	↑***
	H-H											↑***	←
	H-G												←***
	E-G												

Significant interaction effects were found between *technique* and *distance* ($F_{6,3303.2} = 93.082, p < .001, \eta_p^2 = 0.130$). Post-hoc analysis revealed that, for selections within the Personal proxemic zone, *Head* and *Eye* outperformed *Hover* and *Airtap* ($p < .001$). The performance of *Hover* also surpassed *Airtap* ($p < .001$).

At the Social distance, *Head* outperformed all techniques ($p < .001$). *Eye* was also significantly faster than *Hover* ($p < .001$) with both *Eye* and *Hover* having better performance than *Airtap* ($p < .001$).

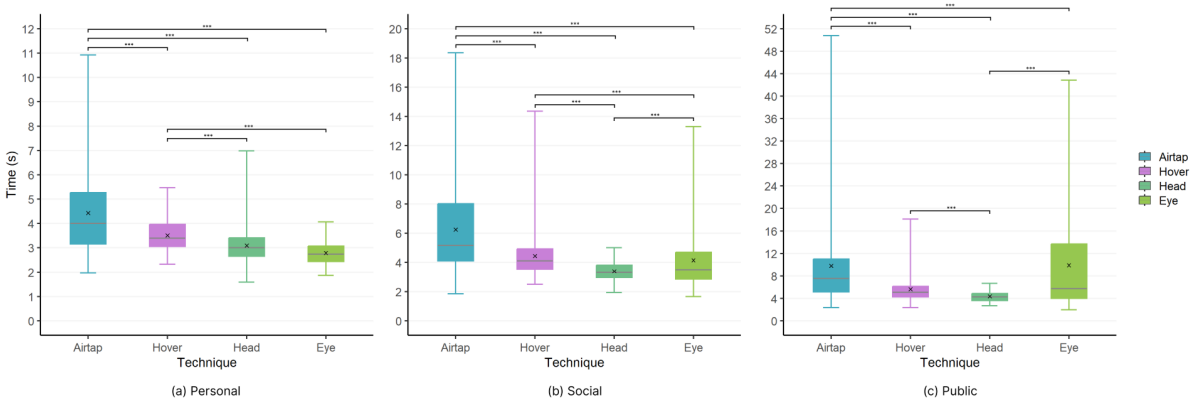

Figure 5.7: Boxplots showing distribution of selection times for Airtap, Hover, Head and Eye in the Personal (a), Social (b) and Public (c) proxemic zones. Mean times are indicated by 'X'. *Please note that y-axis scales differ across subfigures.*

Within the Public proxemic zone, *Head* again exceeded the other techniques ($p <$

.001). *Hover* and *Eye* also surpassed the performance of *Airtap* ($p < .001$). Selection times are presented based on proxemic zone in Figure 5.7, with a full breakdown of pairwise comparisons defined in Table 5.1.

Error

Overall, 54 trials resulted in errors (where an incorrect selection was triggered before the correct selection). Statistically significant interaction effects were found between Interaction Technique and Distance ($F_{6,341} = 36.387, p < .001, \eta_p^2 = 0.369$), with post-hoc analysis revealing that *Eye* performed significantly worse at the Public distance (n = 33) than every other technique at all distances ($p < .001$). All other technique/distance comparisons resulted in minimal errors (between 0 and 3). Although Hover produced the most errors in the Personal proxemic zone (n = 5), no other significant differences were found.

Task Load

The *technique* requiring the least *task load* was *Head* (M = 35.12, SD = 19.93), followed by *Hover* (M = 38.97, SD = 18.83), *Eye* (M = 50.00, SD = 23.07) and *Airtap* (M = 57.05, SD = 16.03). ART ANOVA tests revealed a significant difference between *techniques* ($F_{3,93} = 16.711, p < .001$), with post-hoc pairwise comparisons suggesting that users experienced less overall *task load* with *Head* than with *Eye* ($p < .001$) and *Airtap* ($p < .001$). *Hover* was found to require significantly less *task load* than *Airtap* ($p < .001$) and *Eye* ($p < .05$).

Significant differences were also found for the Physical ($F_{3,93} = 9.1252, p < .001$), Performance ($F_{3,93} = 11.743, p < .001$), Effort ($F_{3,93} = 13.711, p < .001$) and Frustration ($F_{3,93} = 21.519, p < .001$) subscales. Notably, *Head* and *Eye* techniques were considered less Physically Demanding than *Airtap* ($p < .001$), while *Head* and *Hover* received lower scores for Frustration than *Airtap* and *Eye* ($p < .001$). *Head* was also considered to

require less Effort than *Eye* and *Airtap* ($p < .001$), with *Hover* receiving lower Effort scores than *Airtap* ($p < .01$). *Head* scored significantly better than *Eye* and *Airtap* in terms of Performance ($p < .001$), with *Hover* being deemed to have better Performance than *Airtap* ($p < .01$) and *Eye* ($p < .05$). A breakdown of NASA-TLX results for each technique is provided in Figure 5.8.

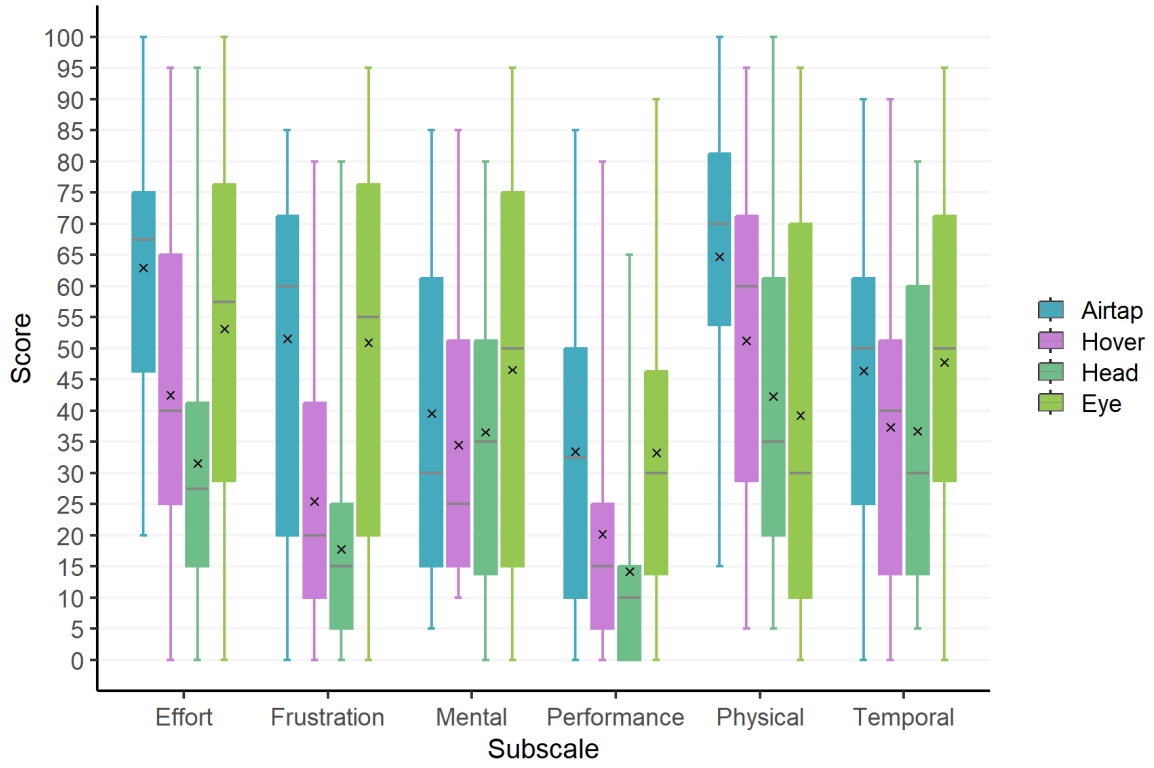

Figure 5.8: Boxplots showing distribution of NASA-TLX subscale scores for *Airtap*, *Hover*, *Head* and *Eye* across Personal, Social and Public proxemic zones. Mean scores are indicated by 'X'. All outliers were included in the analysis.

User Experience

When analysing mean UEQ Scores (as in Figure 5.9), ART ANOVA tests revealed that *technique* produced a significant main effect ($F_{3,93} = 14.888, p < .001$). Post-hoc pairwise comparisons revealed that *Eye*, *Head* and *Hover* had significantly better overall *user experience* than *Airtap* ($p < .001$). Significant differences were also found between interaction techniques on all UEQ subscales. Figure 5.9 provides a breakdown of results for Attractiveness ($F_{3,93} = 11.787, p < .001$), Perspicuity ($F_{3,93} = 18.731, p < .001$), Effi-

ciency ($F_{3,93} = 13.667, p < .001$), Dependability ($F_{3,93} = 12.395, p < .001$), Stimulation ($F_{3,93} = 10.242, p < .001$) and Novelty ($F_{3,93} = 5.0467, p < .01$).

Head, *Hover* ($p < .001$) and *Eye* ($p < .05$) were found more Attractive and Perspicuous than *Airtap*. *Head* also received higher scores for Perspicuity than *Eye* ($p < .001$). *Head* and *Hover* were considered more Efficient and Dependable than *Airtap* ($p < .001$), with *Head* also being deemed more Efficient and Dependable than *Eye* ($p < .01$). Furthermore, *Head* ($p < .001$), *Eye* ($p < .001$) and *Hover* ($p < .01$) received better scores for Stimulation, with *Eye* being considered more Novel than *Airtap* ($p < .01$).

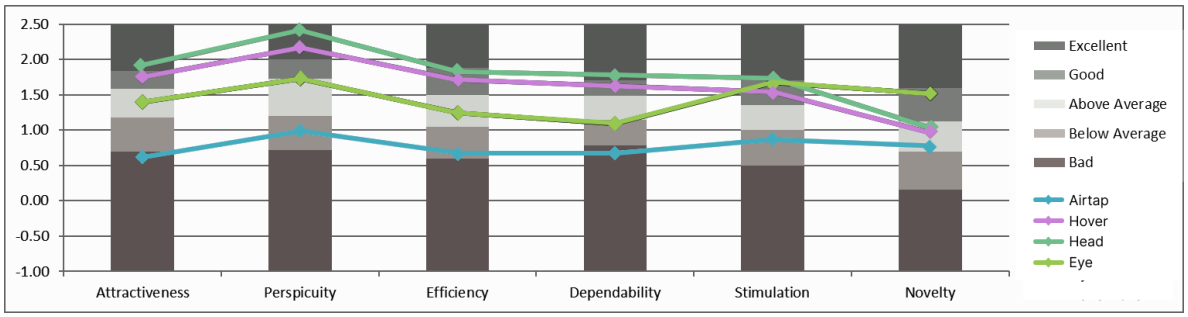

Figure 5.9: UEQ Subscale Scores for *Airtap*, *Hover*, *Head* and *Eye* for interactions across Personal, Social and Public proxemic zones. Mean ratings are shown from Bad to Excellent.

Preference

After completing all trials, users ranked the interaction techniques in order of preference. As distances were counterbalanced, participants were asked to provide rankings for objects that were closest to them (Personal) and for objects that were farthest away (Public). Within the Personal zone, Eye was ranked first most often (50%), followed by Head (31.25%), Hover (15.63%) and Airtap (3.13%). When objects were in the Public zone, most participants preferred Head (65.63%), followed by Hover (21.88%), Eye (9.38%) and Airtap (3.13%).

Participants tended to refer to Eye as being quick/easy (n = 12) and accurate (n = 5) when objects were closer - P11: “it was fast and spot-on”, P30: “there’s very little physical interaction”. Conversely, the technique was often considered difficult to employ

(n = 9) or the least accurate (n = 8) when objects were placed farther away - P10: "It goes from working to just not working, it falls off a cliff", P21: "It didn't seem like it knew what I was looking at".

Head, on the other hand, was reported to be quick/easy (n = 8) and accurate (n = 9) even when selecting objects in the Public zone - P13: "It's consistent, didn't let me down even with smaller targets", P14: "With head (compared to eye) I had a pointer, so I know where I'm aiming".

Hover was deemed more accurate/easier to control than Airtap (n = 8) - P2: "Hover was way quicker than the tap, which seems unnecessary", P22: "doing the gesture gets tiring, but hover was quite a bit easier and less time-consuming". Many found that Airtap required unnecessary effort for close objects (n = 7) as well as those far away (n = 8), with the technique considered to be awkward/frustrating (n = 6) - P16: "like why is this (performs gesture) so complicated?", P19: "I had a lot of trouble with that pinch".

Unlike other techniques, Head was not ranked last by any participants, highlighting its consistency and reliability - P25: "head gaze obviously was better", P26: "Head will be winner".

5.3.3 Discussion

This chapter has explored the appropriateness of four interaction techniques (*Airtap*, *Hover*, *Head* and *Eye*) for performing fundamental selections across Personal (0.5m-1.0m), Social (1.0-4.0m), and Public (>4.0m) proxemic zones. This section highlights the key findings of the experiment, discussing the design implications of employing freehand and gaze-based techniques for far-field interaction in seated AR environments.

Head Gaze is the Most Reliable Technique

Head and *Eye* were found to provide significantly faster *selection times* than *Airtap* and *Hover*, with *Head* being the most consistent technique overall. There were no significant differences found between *Head* at the public distance to *Airtap* in the personal zone, further emphasising the reliability of the *Head* technique. Head and Eye techniques are likely faster as they implicitly incorporate both the search task and the selection stage, whereas Airtap and Hover techniques require users to employ a visual search before employing their hand to select the correct object. Some users also had difficulties accurately positioning the cursor with Airtap and Hover, where the cursor kept going behind objects when participants lifted their hand upwards from a resting position. This resulted in users taking additional time to correct their aim, which was not an issue experienced with *Head* (Whitlock et al., 2018).

Further, in line with previous research, eye tracking performance was found to degrade considerably as the visual angle of objects decreased (Kytö et al., 2018). Even though some participants had little trouble selecting targets with *Eye* in the Public zone, several participants felt forced to adapt their behaviours in response to the technique's limitations. Notable changes that were observed and reported included leaning forward, tilting the head, standing up, or squinting. This could have been a result of limitations in the tracking capabilities of *Eye*, and/or due to the technique not providing a cursor (as per developer guidelines), which is a key factor for effective system feedback on the interacting layer (Venkatakrishnan et al., 2023).

Several participants found eye gaze difficult to employ, especially as target distance increased. As eye provided no cursor due to constraints known as fleeing cursor (see Section 6.3.1), improved system feedback could have made users feel more confident and in control of interactions, reducing error rate and completion time. Venkatakrishnan et al. (2023) highlight how interaction can be significantly enhanced by visual feedback, and although this was not a primary consideration of the presented research, results from

both studies exploring near-field and far-field interaction suggest more testing is needed to refine and improve system feedback. This will help ensure the system remains in alignment with the user's expectations at every interaction stage.

Although distance hugely impacted its performance, with Eye producing significantly more errors than any other technique, Eye still produced comparable selection times to the *Hover* technique in the Public zone, and exceeded the speed of *Airtap* across all distances explored. *Eye* also received comparable *task load* scores to *Airtap* and was ranked significantly lower for Physical Demand. Preferences (see Section 5.3.2), primarily reiterate the objective results. Even though *Eye* received low overall subjective feedback, likely due to heightened performance issues when selecting content in the Public zone, it was deemed the preferred technique for selecting objects in the Personal zone, with *Head* being preferred in the Public zone. In line with results from Whitlock et al. (2018), this could partly be because *Head* was the easiest to control even at a distance; with the cursor always appearing central to the user's field of view, and also due to the robustness and reliability of the technique (Uzor and Kristensson, 2021).

Despite being the baseline technique, a primary issue with *Airtap* was the lack of flexibility it provided. Many participants gradually adopted their own hand poses and approaches. For example, when using a "puppet hand" (bringing all fingers together to meet the thumb) the gesture maintained functionality, however, when tilting the hand sideways or upside down, input was ineffective. Users also often resorted to a "comfort grip" (Pfeuffer, B. Mayer, et al., 2017; Xinyi Liu et al., 2022) between trials, and gestures were frequently ignored or misinterpreted. Furthermore, for smaller distant objects, selections still often took multiple attempts, even when gestures were correctly inferred. This was due to the *Airtap* action causing the cursor to move off-target. Limited hand-tracking areas also meant hands were sometimes not recognised by the system. Although these are inherent limitations of the hardware and software used, prior research suggests that these issues could be minimised by referencing trajectories as opposed to predefined gestures (Pham et al., 2018) and providing larger interaction zones (W. Xu, Liang, Y.

Chen, et al., 2020) as provided by Apple Vision Pro.

Hover Provided better Usability than Airtap

Hover was found to have better performance and usability than *Airtap* across all distances, and received comparable scores for *task load* to *Head* which was ranked best overall. This suggests that separating the selection mechanism of unimodal distal freehand interaction (i.e. using diectic pointing alongside an alternative selection technique such as blink or speech) could decrease task load and improve the usability of selection tasks.

Although previous research highlights that there is a learning curve involved with *Airtap* (Pfeuffer, B. Mayer, et al., 2017; Pourmemar and Poullis, 2019), most users (68.75%, n = 22) had previous experience using freehand gestures (i.e. with HoloLens, Kinect or Leap Motion). Even though adding the confirmation gesture arguably decreases the likeliness of accidental selection compared to *Hover*, especially when objects are larger and closer to the user, the opportunity for input to be misinterpreted was heightened by introducing an added layer of complexity with Airtap. Despite being restricted by "hover" times, even frequent AR users preferred the simplicity of *Hover* over employing *Airtap*, as it decreased the level of frustration experienced.

Overall, these results suggest that the practicality of AR interfaces could be significantly enhanced by simplifying or separating pointing and selection mechanisms. Combining pointing methods with alternative confirmations, such as speech (Whitlock et al., 2018), blinks (Xinyi Liu et al., 2022; F. Lu, Pavanatto, and Doug A Bowman, 2023) or button presses (Whitlock et al., 2018) (depending on the use case) could limit the time required for users to maintain the stability of the pointing techniques and further improve their performance.

Table 5.2: Advantages and disadvantages of interaction techniques in the Personal (0.5m-1m), Social (1m-4m), and Public (>4m) zones.

Technique	Findings
Airtap	- Slowest technique - Worst ratings for user experience - Highest task load - Least preferred
Hover	+ Preferred over *Eye* (Public) + Less overall task load than *Airtap* and *Eye* - Cursor can be difficult to accurately position (Social, Public)
Head	+ Most consistent across distances + Fastest technique (Social, Public) + Most preferred (Public)
Eye	+ Fastest technique (Personal) + Most preferred (Personal) + Better user experience than *Airtap* - Less interaction control (Social, Public)

Distance Impacts Techniques

On the whole, results show that the distance and angular size of objects are key contextual factors in defining the suitability of selection techniques (Whitlock et al., 2018; Hussain, Park, and H. K. Kim, 2023). This suggests that providing interaction techniques interchangeably based on distance, as opposed to adapting a single input method (i.e. switching between press and airtap) to cover all distances, could improve the learnability, usability and flow of interactions.

For example, results suggest that by providing *Head* and *Eye* pointing interchangeably based on distance, issues surrounding the heightened task load experienced with *Head* when interacting with content close to the user (i.e. placed between the Intimate and Personal proxemic zones) and *Eye* when content is far away (i.e. placed beyond the Social zone) could be mitigated, maximising the value of both techniques. This concept is discussed further in Chapter 7, where findings from both seated interaction studies are considered in paralell to provide recommendations for interaction design based on distance. Table 5.2 summarises the results from the study presented.

5.3.4 Summary

This chapter has reported a study involving 32 participants, highlighting the advantages and limitations of freehand and gaze-based interaction techniques for far-filed interaction. Results show that user distance to interactive content significantly influences technique suitability. Focus has been given to comparing the baseline unimodal technique (freehand gesture) against common time-based techniques, to better understand the reliability of different pointing methods across proxemic zones.

Notably, *Head* interaction was deemed to be the most suitable/reliable for selecting virtual content across multiple proxemic zones, with *Eye* being appropriate in the Personal zone. The baseline freehand technique, *Airtap*, was found to be less practical and required too much Physical Demand to be employed comfortably in the explored context. This underscores the crucial role of spatial considerations in enhancing the practicality of AR technologies.

Building on this research, the next chapter looks beyond distance to also consider user-defined movement approaches in room-scale AR. This improves understanding around the affordances of commonplace interaction methods, and how user locomotion behaviours are impacted by the affordances of freehand and gaze-based techniques.

Chapter Six

User-Defined Locomotion - Distance and Movement (Study 3)

Contents

6.1 Introduction **129**
6.2 Background Research **131**
6.2.1 Implicit Interaction 131
6.2.2 Explicit Interaction 134
6.2.3 Summary 137
6.3 User-Defined Distance and Movement Approaches **137**
6.3.1 Method 138
6.3.2 Results 145
6.4 Discussion **164**
6.5 Summary **169**

Note: This chapter is adapted from work currently under review

Spittle, B., Frutos-Pascual, M., Creed, C. and Williams, I. (2025). Walk This Way: How Augmented Reality Interaction Techniques Influence User Movement and Spatial Positioning. Special Issue on Spatial Computing in the Journal of Behavior and Information Technology (under review)

6.1 Introduction

Chapters 4 and 5 considered the influence of distance on interaction techniques across proxemic zones in a seated context. However, as one of the primary benefits of AR technologies is the ability to interact with virtual content placed throughout the real world, an essential piece of contextual information is users locomotion approaches, which will inherently influence the suitability of different interaction techniques (Lages and Doug A. Bowman, 2019; Grubert et al., 2017; Whitlock et al., 2018). Understanding how users employ locomotion in AR will therefore be crucial for optimising spatial understanding, immersion, exploration, interactivity, and flow across interaction scenarios (Lages and Doug A. Bowman, 2019; Norouzi et al., 2019).

Locomotion can be defined as the ability to move from one place to another, and involves the dynamic interplay of human intention, perception, and environmental awareness (Sanz et al., 2015). As an integral part of our daily routines, we instinctively navigate diverse spaces and interact with people and objects in our surroundings, with AR technologies introducing possibilities to enhance our interactions within the physical world, by integrating digital information, such as text, objects, and auditory cues, throughout the environment (X. Li et al., 2022; Xinyi Liu et al., 2022).

As highlgihted by the range of applications discussed in Chapters 4.2 and 5.2, AR is poised to become a pervasive technology, where it will be used seamlessly across multifaceted domains, such as for work, education and entertainment, in a range of contexts and settings (Grubert et al., 2017; Hertel et al., 2021). Many current AR technologies, including Apple Vision pro, HoloLens 2 and Magic Leap 2, encompass a wide array of sensors that enable environmental understanding, motion tracking, and a myriad of interaction possibilities (Gallardo et al., 2023). Nevertheless, most applications still restrict users to a single interaction context, failing to leverage the benefits of using multiple interaction techniques interchangeably (Hertel et al., 2021; Spittle, Frutos-Pascual, et al., 2022). This arguably restricts the potential of AR, as providing users with the flexibil-

ity to engage in dynamic contexts, based on factors such as their activity, environment, and interaction scenario, will be crucial for establishing AR as a ubiquitous technology (Grubert et al., 2017; Lages and Doug A. Bowman, 2019).

Previous work has shown that people have a similar understanding of their spatial relations with technology as they do in the real world (Ballendat, Marquardt, and Greenberg, 2010; Greenberg, Marquardt, et al., 2011). Despite this, limited research has considered how users choose to employ locomotion when interfacing with virtual content (Lages and Doug A. Bowman, 2019), meaning that there is currently a lack of understanding surrounding how multiple interaction techniques could be used to adapt to users distance from virtual content, position within the interaction space, and walking trajectories. To address this research gap, this work begins to explore Proxemic Dimensions beyond distance, considering the impact of users movement (or lack thereof) in the context of AR. The chapter considers how distance and movement could be referenced to provide adaptive interactions that lead to more flexible and impactful AR experiences (Grubert et al., 2017; Lages and Doug A. Bowman, 2019; X. B. Liu et al., 2024).

A user study explores to what extent user locomotion approaches vary when employing different AR interaction techniques. This is achieved by investigating the processes taken by 40 participants to complete a series of fundamental selection tasks. Akin to Chapter 5, the study compares two freehand techniques (Airtap, Hover) and two gaze-based techniques (Eye, Head), but instead considers the methods followed by participants from a standing position to select different size objects (5cm, 15cm, 25cm) in a room-scale environment. The advantages and limitations of each technique are highlighted with respect to user-defined distance, position, locomotion approach and speed, selection time, task-load, user experience and preference.

The chapter is structured as follows: first, a review of AR interaction research involving locomotion is provided. This is followed by a description of the research methodology employed to conduct the study. The results are then reported and analysed, identifying key insights surrounding user-defined distance and movement, with the discussion

section considering how these findings could be considered to provide AR techniques interchangeably.

6.2 Background Research

This chapter explores research focused on locomotion, and how users employ it to interact with AR content either implicity or explicitly. A range of use cases involving locomotion are highlighted, as well as the advantages and disadvantages of freehand and gaze-based techniques.

6.2.1 Implicit Interaction

Locomotion is inherently employed to adjust position relative to world-locked content (Norouzi et al., 2019; Ballendat, Marquardt, and Greenberg, 2010), and plays a key role when interacting with applications employed in various settings, such as museums, cultural heritage sites, workplaces, and education (Lages and Doug A. Bowman, 2019; Pedersen et al., 2017). At present, the position of users relative to virtual objects is often referenced to affect how virtual content is displayed. This is generally applied in parallel to the real-world, by proportionately adapting the perspective and visual angle of static virtual objects (Pedersen et al., 2017), or the level of visual detail, to scale the amount of information provided to users (Ghaemi et al., 2022).

For instance, Toure, Welsch, and S. Mayer (2021) introduced how information displayed on a HWD could be scaled based on user's movements in a smart factory environment. This concept was proposed to address the challenge of in-situ information representation, by preventing information overload and reducing additional workload (see Figure 6.1). X. Li et al. (2022) also demonstrated how locomotion can be referenced to support real-world collaborative tasks, such as optimising parcel sorting in a warehouse. They achieved this by adapting virtual content on a HWD based on the distances be-

tween objects and multiple users, which was shown to promote collaboration and improve task performance. Similarly, Ghaemi et al. (2022) presented an approach where the density of information in map visualisations could be altered using proxemic zones, where manipulating the distance of the map would control the level of information provided. This offered flexibility, providing opportunities for users to overcome the limitations of different map types.

Figure 6.1: Smart factory maintenance scenario showing different levels of scaled information. When in close proximity to points of interest, more information and relevant interaction opportunities are available. Workers can see other machines in the distance, with coloured indicators showing their operational status (Toure, Welsch, and S. Mayer, 2021).

Some studies have gone beyond how to adapt content based on distance and locomotion to investigate how users position themselves relative to virtual elements. For example, Norouzi et al. (2019) considered user perception and behaviour when engaging with virtual pets on HWDs, concluding that having an AR companion impacts how users position themselves and their social interaction with other people, regardless of whether bystanders can see the virtual agent.

Furthermore, Novick and Rodriguez (2021) examined the similarities of virtual to real-world distance for conversational interaction, exploring how participants approach and converse with a virtual agent in three conditions: no crowd, small crowd, and large crowd. Their results suggest that users tend to position themselves closer to virtual agents in a virtual environment than they would to humans in the physical world, regardless of

whether other virtual agents were present or not. Huang et al. (2022) also considered how participants approached and positioned themselves relative to virtual objects and agents, as well as how they felt walking through them. They again explored this within a converational scenario, where users asked for directions (see Figure 6.2), and compared six representations, two males, two females, a humanoid robot, and an inanimate pillar. Their findings reaffirmed the impact of content type on interaction approaches (R. Li et al., 2019; Sanz et al., 2015). However, again, this was primarily focused on users passive actions and intentions (Ballendat, Marquardt, and Greenberg, 2010), with respect to surrounding people and virtual agents.

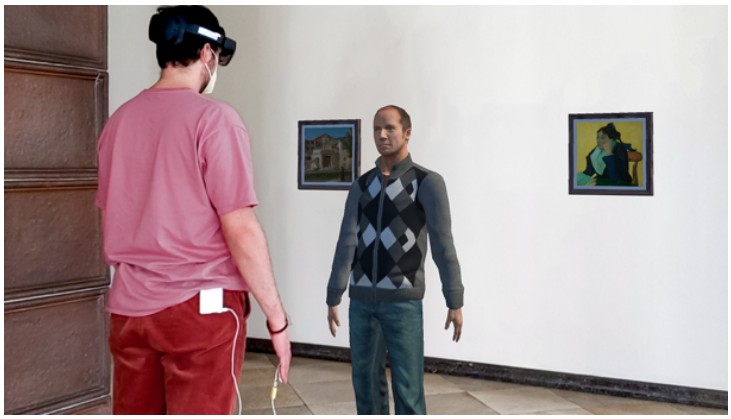

Figure 6.2: Participant respecting the personal space of an agent while asking for directions (Huang et al., 2022).

Previous research has also considered how locomotion impacts the viability and comfort of AR interactions. For example, Nijholt (2021) highlights the implications of experiencing social AR in public spaces, a key factor being to study the effect of changes in visual focus on noticing relevant objects during locomotion. Similarly, Lages and Doug A. Bowman (2019) considered how to support fluid AR interactions as users move and interact in the physical world. Here, different adaptation strategies that react to users orientation, position, or surfaces in the environment were explored, where focus was given to information retrieval (see Figure 6.3).

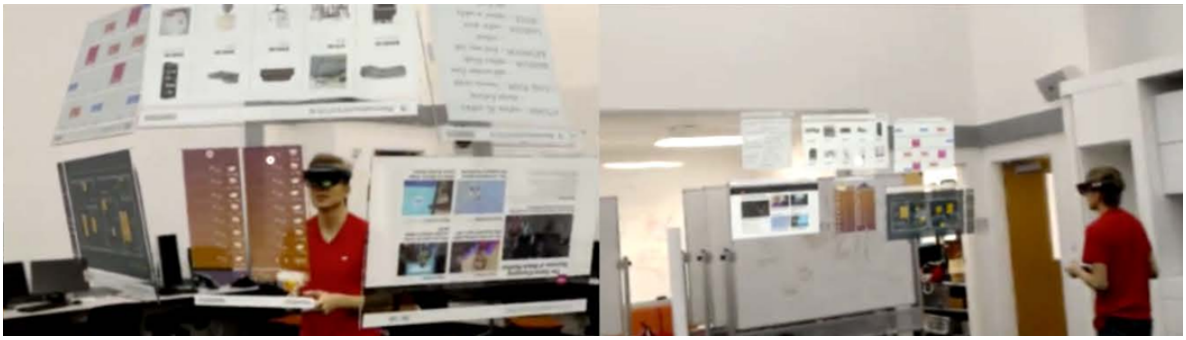

Figure 6.3: Adaptive Augmented Reality Workspace in use, where the virtual panels adjust when users walk and/or approach walls. (Lages and Doug A. Bowman, 2019).

6.2.2 Explicit Interaction

Research has also considered the role of walking for explicit AR interaction tasks, such as selections. For example, Bhowmick, Kalita, and Sorathia (2020) considered to what extent users employed locomotion for selecting objects in a dense, occluded VR environment. Even though the study allowed for navigation through real walking within a predefined tracking area, participants reported that walking to reach a target, especially when placed far away, introduced fatigue and discomfort, and increased the likeliness of colliding with real objects placed throughout the interaction space. This highlights the importance for AR technologies to reference locomotion and changes in spatial syntax to adapt explicit interactions to user environments and preferences (Ballendat, Marquardt, and Greenberg, 2010).

Interactions on the go will become a common scenario when experiencing continuous and multi-purpose AR interactions, such as those expected to be pervasively employed when traversing urban environments in the future (Grubert et al., 2017). Nijholt (2021) discussed how AR has the potential to facilitate and enhance public social interactions, such as through games and entertainment activities. Here, users could engage with virtual characters or participate in shared exploration games, where virtual content could be world anchored to features in the physical environment. This includes landmarks, buildings, billboards or public screens. For example, users could create virtual graffiti or leave messages that remain anchored to specific physical locations. This could remain visible

to users and be engaged with as they move around, creating new interactive experiences.

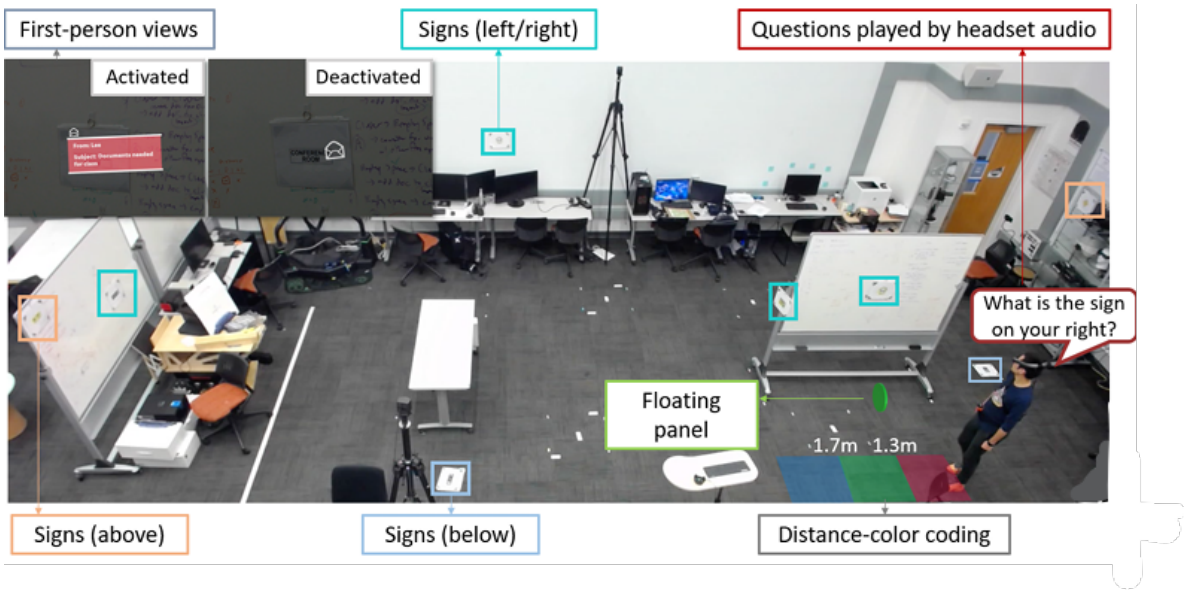

Figure 6.4: Partcipants were instrcuted to maintain their distance and follow a floating panel, encountering 8 physical signs placed throughout the real-world environment to interact with. (F. Lu, Davari, and D. Bowman, 2021).

Another commonplace application for where content will be interacted with during locomotion is when accessing everyday information, such as from weather, news, time, calendar, navigation and email applications, or task and activity trackers (F. Lu, Pavanatto, and Doug A Bowman, 2023). For example, F. Lu, Davari, and D. Bowman (2021) considered how to address explicit activation of virtual content whilst walking, where they proposed and evaluated five techniques. This was based on a task where users needed to access information, which was achieved by expanding icons or widgets that were mapped to content on physical posters/flyers attached to walls and surfaces (see Figure 6.4). They demonstrated the trade-offs of gaze, hand, and head-based methods, finding that the majority of participants preferred fixation glance, an eye-based technique that not only harnessed the natural tendency to fixate in the direction that objects of interest are positioned, but also at their depth.

Q. Zhou et al. (2020) investigated user performance of eyes-free target acquisition during walking, where they focused on using whole-hand grasping gestures to interact with content that followed users at head or torso level (see Figure 6.5). They provide

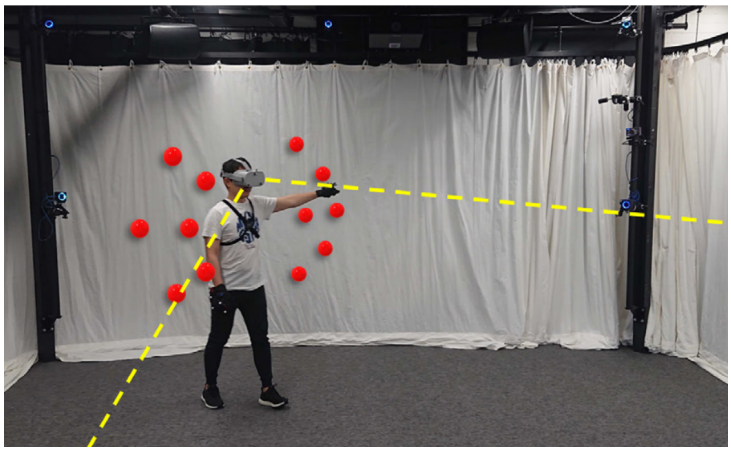

Figure 6.5: Participant performing target acquisition (grasping gesture) to perform object selections while walking in the tracked space (Q. Zhou et al., 2020).

a use case involving a room-scale blueprint-editing interface, which enables workers to grab and use different tools surrounding them (in the form of icons) whilst in motion. This method allows the user to always keep their attention focused on the task at hand, however, they found that body movements induced by walking negatively affected the accuracy of the technique. Müller et al. (2020) explored another approach, considering how to directly leverage walking as a hands-free interaction technique. As depicted in Figure 6.6, this was achieved by providing lanes on the ground, where lateral adjustments in the walking path (based on orientation) allowed users to interact with different AR applications, such as selecting controls in a media player.

Figure 6.6: Walk the Line: application that leverages lateral shifts of the walking path as an input modality. Options are visualised as lanes on the floor and users select options by shifting their walking path sideways. Following a selection, sub-options of a cascading menu appear as new lanes (Müller et al., 2020).

6.2.3 Summary

Overall, research exploring the influence of user locomotion approaches, especially when performing explicit interactions, is limited, and no studies have been found to consider this across a spectrum of AR interaction techniques. Prior work suggests that user's movement and position in the interaction space could be referenced to understand contextual information, such as the users activity and their distance from interactive content, which inherently impact the appropriateness of different techniques (Whitlock et al., 2018; Cheng, Gebhardt, and Holz, 2023). This information, similar to how F. Lu, Pavanatto, and Doug A Bowman (2023) adapts the position of AR content relative to physical characteristics of the interaction environment, could be harnessed by the system to provide the most appropriate interaction technique for excecuting explicit interactions depending on context.

This is partially demonstrated with interaction on some applicaions/HWDs, such as the Hololens 2, which triggers adaptations of freehand paradigms when predefined distances are breached (i.e. direct interaction being enabled when users are 45cm from a virtual object (Microsoft, 2023c)). However. even though this provides a system with some adaptability, no studies have been found to extend this concept to a range of interaction approaches, or consider how users adapt their locomotion behaviours when using different freehand and gaze-based selection techniques.

6.3 User-Defined Distance and Movement Approaches

Building on the concepts presented in Section 6.2, a study exploring the advantages and disadvantages of interaction approaches for selecting world-anchored AR content is now reported. This is based on comparing to what extent users employ locomotion, considering user-defined distance, position, locomotion approach and speed, selection time, task-load, user experience and preference, when using different interaction techniques

(Airtap, Hover, Head, Eye) to select different sized objects (5cm, 15cm, 25cm). The task is again centered around an "observe and interact" scenario (Cheng, Gebhardt, and Holz, 2023; Lischke et al., 2016), where a user refers to one interface component to gather information, before interacting with another virtual component to perform a task. The following section describes how the study was designed, developed, and conducted.

6.3.1 Method

Apparatus

An application was developed for the Microsoft HoloLens 2 in C#, using Unity 2020.3.30f1 and MRTK 2.7.2.0. Windows device portal was used throughout the study to remotely open applications, guide and monitor participants, and track the stability and refresh rate of the HWD. First-person, live Mixed-Reality captures were screen recorded to aid with data analysis, and third-person videos were captured using a 1080p webcam to provide deeper insight into participants' movements and approaches where needed.

Environment

The research was conducted in the same controlled lab environment as studies 1 and 2, between January and February 2023. Two lines were marked parallel to each other on the floor, 6m apart, which were used for alternating start positions. Once comfortably wearing the HoloLens and standing at the start line, participants were asked to look straight ahead at a marker whilst the application loaded.

"Observe" stage At the beginning of each trial, participants were instructed via text to select a start button. On selection, a cube with a symbol on was generated to indicate the target object. Unlike the first two studies, this cube was positioned directly infront of the user. A text prompt instructed participants to look at this cube, which, like study

1 and 2, was displayed for 5 seconds before it disappeared. Following this, participants were prompted via text to “go” and the trial would begin.

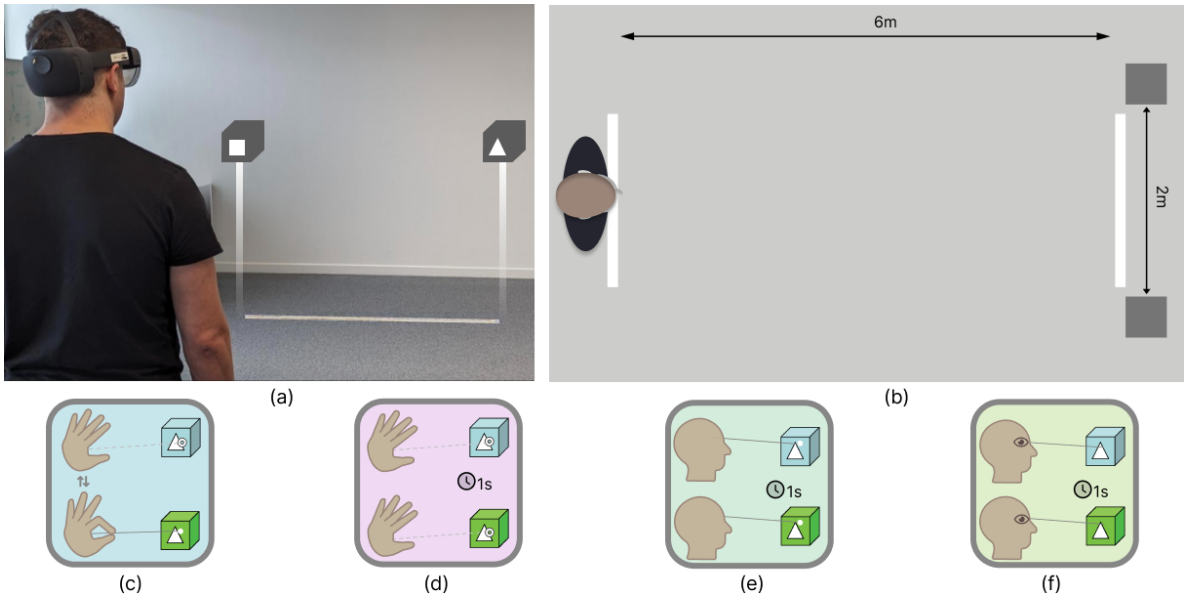

Figure 6.7: Environment and Techniques: For each trial, the participant began from standing behind the starting line. Following this, two objects appeared and a text prompt instructed participants to ”Go”. Participants were then free to move and select the target object. After selecting the correct object, the participant stood on the opposing line and faced the other direction to start the next trial.

“Interact” stage The interact stage was similar to that employed for studies 1 and 2, however, instead of an object array, two small (5cm), medium (15cm) or large (25cm) cubes were presented 6m away from the starting position on the opposing line, spaced 2m apart. The objects were positioned to be within the appropriate interaction region, centred -0.32 below the horizon line relative to the height of individual participants, which was based on HoloLens 2 comfort guidelines (Microsoft, 2021a). Spatial awareness was activated, overlaying the environment with a transparent mesh, and the “Pointing Extent” of each ray (see Section 6.3.1) was increased to 10m. The depth of targets was indicated to users via virtual poles, which extended from the base of the cubes to the ground. Figure 6.7 illustrates the interaction space and virtual environment employed.

Akin to the first two studies, selections could be made by interacting with any point of the target cube. All cubes were black by default and turned blue when targeted,

with visual feedback initiated after an onset of 200ms. If a selection was successful, the target cube turned green and a confirmation sound was triggered from the MRTK sound library. In cases where participants interacted with the wrong cube, the target turned red and produced an MRTK error sound. A reset task button was provided at the starting position, which was used in cases when participants failed to recall the target object. After all targets were selected an "End of Condition" message appeared to prompt participants to cease the process.

Interaction Techniques

The interaction techniques explored (Airtap, Hover, Head and Eye) are illustrated by Figure 6.7. Techniques were identical to those employed for Study 2, developed by referencing HoloLens 2 guidelines and using MRTK 2.7.2.0 (Microsoft, 2022b). Again, to maximise replicability, default cursors and input parameters provided by MRTK were used, as defined below.

Hand Airtap involved a two-step interaction process. First, users pointed at the object of interest, then employed a "Pinch" gesture to execute the selection. This was achieved by controlling a cursor, which was attached to the end of a ray rendered from the palm. A solid ray and cursor indicated successful gesture registration, with selections being triggered upon release.

Hand Hover employed a "dwell" or "hover" time, requiring that the default MRTK cursor, which was attached to the end of a ray rendered from the palm, continuously intersected the object for one second to trigger a selection.

Head Gaze employed a head-gaze cursor attached to the end of an invisible ray, which extended from the HWD towards the viewing direction. Selections were triggered when the cursor had continuously intersected the object for one second.

Eye Gaze relied on visual feedback to indicate object targeting and selection. An invisible cursor was attached to the end of an invisible ray, which extended from the eyes toward the viewing direction. Selections were triggered when participants' gaze continuously intersected the object for one second. No cursor was visualised with *Eye* to avoid the effect described as "fleeing cursor" (Microsoft, 2023b).

Task

Participants performed a total of 12 target selection tasks, 4 small, 4 medium, 4 large, with each technique. They started from a standing position within the public proxemic zone, 6m from the objects, and were instructed to employ a natural approach to select the correct target as quickly and accurately as possible. After selecting the correct target, participants would position themselves on the opposite line and rotate 180 degrees to face the other direction. They would repeat this process until all targets had been selected.

Participants

The study involved 40 participants: 21 males, 18 females, and 1 non-binary individual, aged between 22 and 41 (average age M=27.00, SD=4.72). Participants were university students and staff from the College of Computing. Their familiarity with immersive technologies varied. Some were novices (n = 10), who reported to have never used AR/VR technologies or to have used them rarely (n = 15). Others were more experienced, reporting to use them monthly (n = 10), weekly (n = 2), or daily (n = 3). All participants were right-handed, with none having colour blindness and/or corrected visual acuity below 0.80 (see Section 6.3.1 for further information on screening). Participants were compensated with a £10 voucher.

Protocol

Each study session lasted around 90 minutes, with the study protocol previously receiving IRB approval. The protocol was broken down into the following steps:

Pre-test Informed consent and demographic information, including experience, was attained from participants prior to each study session. After welcoming participants, the purpose of the study and test protocol was explained in detail. Following this, the researcher measured visual acuity and colour blindness via a Snellen chart and Ishihara test. Participants were then asked to adjust and wear the HWD so it was comfortable and stand at the initial starting position (see Figure 6.7).

Training After the experimenter explained how to employ the technique being tested, participants would experience the training phase, where they were given as much time as they required to practice and ask questions. The training simulated the main experimental task, however, only three selections were made (one of each size) and a different set of symbols were used. Participants were able to repeat the training phase, with no participants completing more than three repetitions. Each participant ran eye calibration before the first training phase to maximise the performance of the HoloLens 2. If performance was not as expected during any of the training, calibration was repeated.

Test After attaining verbal confirmation that participants clearly understood the technique and were comfortable with the interaction space, they were presented with the main experimental task. All participants completed 12 target selections with each technique, 4 small, 4 medium, 4 large, which produced 1,920 trials (40 participants x 12 selections x 4 techniques = 1,920). After each condition, participants completed NASA-TLX and UEQ questionnaires and noted what they liked and disliked about the technique. They would then have a short break before beginning the next condition. The study followed a within-subjects design, where Interaction Technique and the Size of the target cubes

were defined as independent variables. Interaction Technique was counterbalanced using a Latin Square and the cube position (left or right) and size (small, medium, or large) were randomised. User-defined distance, position, locomotion approach and speed, selection time, task-load, user experience and preference were captured as dependent variables.

Post-test

Following the completion of all four conditions, a semi-structured interview was conducted to gather subjective feedback and overall preference rankings. This was to better understand interaction approaches and contextualise what participants liked and disliked about each technique. Partcipants were asked to explain their ranking order and also draw on how each technique impacted their interaction approaches.

Metrics

The following metrics were captured for each technique:

Task Completion Times represent the duration in ms from when each trial began (following the "observe" stage as defined in Section 6.3.1) to when the correct object was selected.

Distance calculations were performed based on sets of 3D coordinates. For gaze techniques, this was based on the distance between the main camera (Hololens 2) and the virtual object. For Freehand techniques, it was based on the distance between the Mixed Reality hand pointer position (ray origin) and the virtual object.

Position was captured to show where in space users moved over time (their trajectories), as well as where selections were made (their end positions). This was based on data mapped from the position of the main camera (Hololens 2) in the z and x dimensions for

each technique. Users' end positions were considered in regards to Hall's proxemic zones: Intimate (less than 0.5 metres), Personal (0.5 to 1 metre), Social (1 to 4 metres), and Public (more than 4 metres) (E. T. Hall, 1966; Daza et al., 2021).

Locomotion Approach categorieses walking approaches based on two factors:

1) based on whether users walked or not - if the user took a step, or traveled more than 0.63m from the start position (defined by previously reported minimum step length (Koop et al., 2020)) they were classed as active. If not, they were classed as stationary.

2) For all active trials (where users moved more than 0.63m), it was also considered if users were primarily stationary, decelerating, or accelerating during different time windows (i.e. the 'Entire Trial', 'Last 5 Seconds', 'Last 3 Seconds', or the 'Last Second' before the selection was made). Analysis was conducted for full trial durations to provide an overview of participant interactions. The last second time window corresponds to the duration required for object selection using time-based techniques (Hover, Head, Eye), offering insight into participant behaviour as selections were being made. Further analyses explored three and five second windows to examine behaviours leading up to object selections, as well as highlighting how different techniques and user approaches influenced task completion times. Trials were classed as decelerating if there were more decreasing speed counts between samples than increasing, and vice-versa for accelerating. Stationary classification was again based on the 0.63m threshold defined above.

Locomotion Speed represents the average speed traveled by participants during each trial in meters per second (m/s).

Task Load was measured via the official iOS NASA-TLX (Task Load Index) application, where raw scores were analysed for individual subscales as well as weighted scores for overall task load (Index, 2020).

User Experience was analysed using a standardised User Experience Questionnaire (UEQ) (Schrepp, 2015).

Preference Rankings captures how participants rated the interaction techniques from best to worst.

6.3.2 Results

This section presents the research findings, where each interaction technique (Airtap, Hover, Head and Eye) and the locomotion approaches adopted with each technique are compared. As all data followed a non-normal distribution, each analysis was based on Aligned Rank Transform (ART) repeated-measures ANOVAs (Wobbrock, Findlater, et al., 2011).

After screening trial-level data (removing invalid/error trials and outliers), the remaining trial-level observations were analysed in R using ARTool (Wobbrock, Elkin, et al., 2024). The data was first aligned and ranked per effect with the Aligned Rank Transform. Aligned ranks were then fit with repeated-measures mixed-effects models including fixed effects of *Technique*, *Object Size*, and their interaction (*Technique*×*Size*), plus a random intercept for *Participant*. No pre-analysis aggregation to per-participant summaries was performed. This is because trial-level modelling with a participant random effect preserves within-participant dependence without inflating Type I error and avoids aggregation bias, while ART accommodates non-normal, long-tailed timing distributions and retains valid factorial tests of main effects and interactions (Wobbrock, Findlater, et al., 2011).

Overall ART F-tests for each main effect and interaction were obtained and, where effects were significant, pairwise contrasts used Bonferroni adjustments to control family-wise error ($\alpha = .05$). Subjective measures collected once per condition (e.g., NASA-TLX/UEQ) were analysed at the *Participant*×*Condition* level using the same ART framework, as trial-level modelling is not applicable for single-rating outcomes.

Task Completion Times

Before analysing completion times, data was removed where participants made an incorrect selection (N=15, 0.78%). Outliers were also omitted (N=30, 1.56%), which were defined as any attempts above three standard deviations from the mean selection time (mean ± 3std.). This resulted in a total of 1,875 trials being analysed out of the 1920 trials collected.

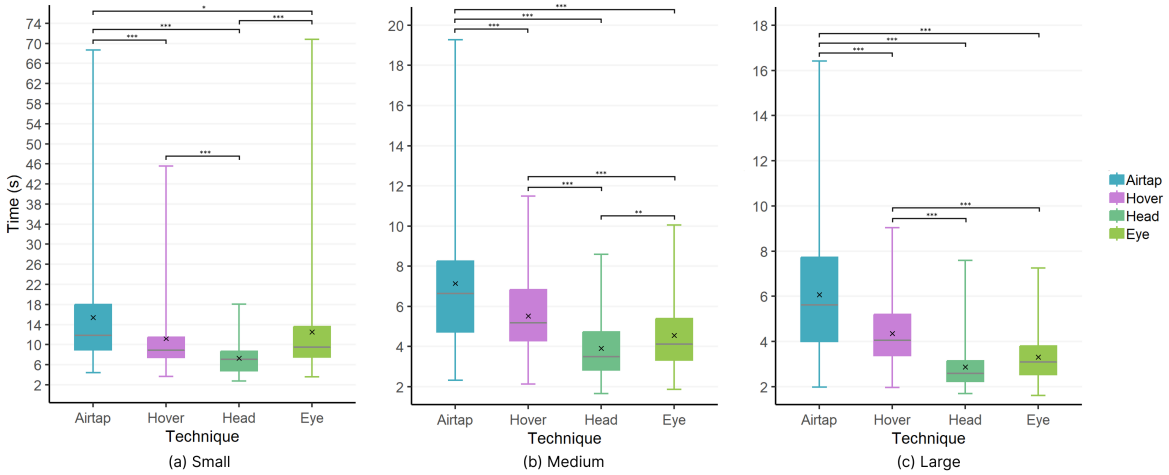

Figure 6.8: Box-plots showing the distribution of selection times for *Airtap*, *Hover*, *Head*, and *Eye*, across three object sizes: (a) Small, (b) Medium and (c) Large. Significant differences are provided, where *** p=<0.001, ** p= 0.01 and * p= 0.05. Mean times are indicated by 'X'. Outliers were not included. *Please note that y-axis scales differ across subfigures.*

When considering mean completion times for Small objects, Head was the fastest technique (M=7.30s, SD=3.22), followed by Hover (M=11.20s, SD=7.29), Eye (M=12.52s, SD=9.50) and Airtap (M=15.38s, SD=10.37). For Medium objects, Head (M=3.92s, SD=1.52) was again the fastest technique, however, this was followed by Eye (M=4.55s,

SD=1.73), Hover (M=5.52s, SD=1.81) and Airtap (M=7.14s, SD=3.28). Similarly, when selecting Large objects, Head (M=2.88s, SD=0.97) was faster than Eye (M=3.31s, SD=1.14), Hover (M=4.35s, SD=1.51) and Airtap (M=6.07s, SD=2.80). These findings are illustrated by Figure 6.8.

A significant main effect was found between Interaction Technique and Object Size ($F_{6,1824.3} = 23.221, p < 0.001, \eta_p^2 = 0.064$). Post-hoc analysis revealed that, when selecting Small objects, Head was more efficient than Airtap, Eye and Hover ($p < 0.001$). Hover ($p < 0.001$) and Eye ($p < 0.05$) were also faster than Airtap.

Distinctions were again found across all interaction techniques for Medium object selections ($p < 0.001$), however, the difference between Head and Eye ($p < 0.01$) was less significant than the other comparisons. Eye was found to be better than Hover and Hover better than Airtap, with Head being the fastest technique. Excluding the Head and Eye comparison which had no significant difference, the same effects were found across techniques for Large object selections ($p < 0.001$).

Table 6.1: Selection Times: Results of pairwise comparisons for Interaction Technique and Object Size, where *** p=<0.001, ** p= 0.01 and * p= 0.05. Arrows indicate which technique provided the fastest selection time, H-AT: Airtap, H-H: Hover, H-G: Head or E-G: Eye

Size		Small				Medium				Large			
	Technique	H-AT	H-H	H-G	E-G	H-AT	H-H	H-G	E-G	H-AT	H-H	H-G	E-G
Small	H-AT		↑***	↑***	↑*	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
	H-H			↑***	←	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
	H-G				←***	↑	↑***	↑***	↑***	↑***	↑***	↑***	↑***
	E-G					↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
Medium	H-AT						↑***	↑***	↑***	↑***	↑***	↑***	↑***
	H-H							↑***	↑***	←	↑***	↑***	↑***
	H-G								←**	←***	←	↑***	↑**
	E-G									←***	↑	↑***	↑***
Large	H-AT										↑***	↑***	↑***
	H-H											↑***	↑***
	H-G												←
	E-G												

Although Figure 6.8 shows a small number of comparatively long completion times relative to the majority of trials, especially for Airtap and Eye, these cases primarily oc-

curred when participants chose to remain at the starting line and persist with distant selections rather than walking closer to objects to reduce task difficulty. These values therefore reflect participant behaviour rather than technical issues or measurement error, and were treated consistently within the analysis. A full breakdown of pairwise comparisons for selection time is provided in Table 6.1.

Distance

When selecting Small objects, users interacted from farthest away with Head (M=4.71m, SD=2.11), followed by Hover (M=2.91m, SD=2.21), Eye (M=2.36m, SD=1.38) and Airtap (M=2.32m, SD=1.88). For Medium objects, Head (M=5.74m, SD=1.54) was again used from the greatest distance, however, this was followed by Eye (M=4.84m, SD=1.71), Hover (M=4.79m, SD=1.95) and Airtap (M=4.30m, SD=2.03). Likewise, for Large objects, Head (M=6.03m, SD=1.17) was employed from farther away than Eye (M=5.51m, SD=1.21), Hover (M=5.10m, SD=1.73) and Airtap (M=4.66m, SD=1.89). These findings are further illustrated by Figure 6.9.

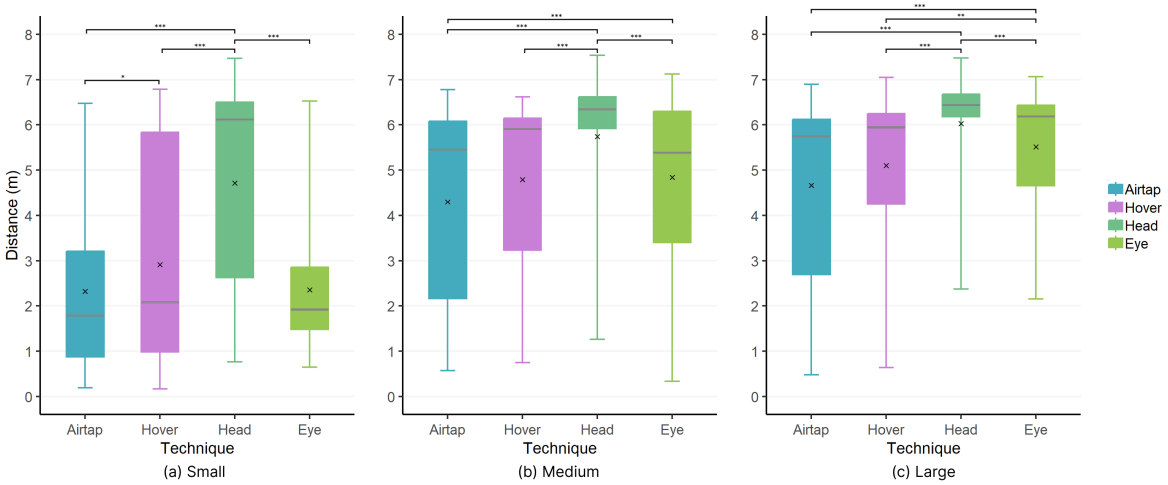

Figure 6.9: Box-plots showing the distribution of participants' distance to target when interacting with the four interaction techniques: *Airtap*, *Hover*, *Head*, and *Eye*, across three object sizes - (a) Small, (b) Medium and (c) Large. Distance is measured from the pointer origin to the target object in meters. Significant differences are provided, where *** p=<0.001, ** p= 0.01 and * p= 0.05. Mean Distances are indicated by 'X'.

A significant main effect was found for Interaction Technique and Object Size

($F_{6,1869} = 28.167, p < 0.001, \eta_p^2 = 0.046$), where post-hoc analysis revealed Head was used from a greater distance than all other techniques for each object size ($p < 0.001$). Participants were also more likely to make selections closer to Small objects with Hover than Airtap ($p < 0.01$) and got closer to Medium and Large objects with Airtap ($p < 0.001$), and Large objects with Hover ($p < 0.01$) when compared to Eye. A full breakdown of pairwise comparisons is provided in Table 6.2. On average, distance between users and the target object was decreased by 0.37m (SD=0.12) with Airtap and 0.36m (SD=0.12) with Hover when compared to the reference position of the HWD main camera. This was due to users extending their arm out in front when making selections.

Table 6.2: Distance from Target: Results of pairwise comparisons based on Interaction Technique and Object Size, where *** p=<0.001, ** p= 0.01 and * p= 0.05. Arrows indicate which technique was used from a greater distance, H-AT: Airtap, H-H: Hover, H-G: Head or E-G: Eye.

Size		Small				Medium				Large			
	Technique	H-AT	H-H	H-G	E-G	H-AT	H-H	H-G	E-G	H-AT	H-H	H-G	E-G
Small	H-AT		↑*	↑***	↑	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
	H-H			↑***	←	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
	H-G				←***	←***	←	↑***	←	←*	←	↑***	↑*
	E-G					↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
Medium	H-AT						↑	↑***	↑***	↑	↑***	↑***	↑***
	H-H							↑***	↑	←	↑	↑***	↑***
	H-G								←***	←***	←***	↑	←**
	E-G									←	↑	↑***	↑***
Large	H-AT										↑	↑***	↑***
	H-H											↑***	↑**
	H-G												←***
	E-G												

Position

As well as capturing the distances techniques were employed from, x, z coordinates of the user were collected to measure users positions within the interaction space relative to the virtual cubes. Movement trajectories are presented in Figure 6.10, illustrating the walking paths and end positions of participants. They also present the frequency of selections made in each proxemic zone: Intimate (less than 0.5 metres), Personal (0.5 to 1 metre), Social (1 to 4 metres), and Public (more than 4 metres) (E. T. Hall, 1966; Daza

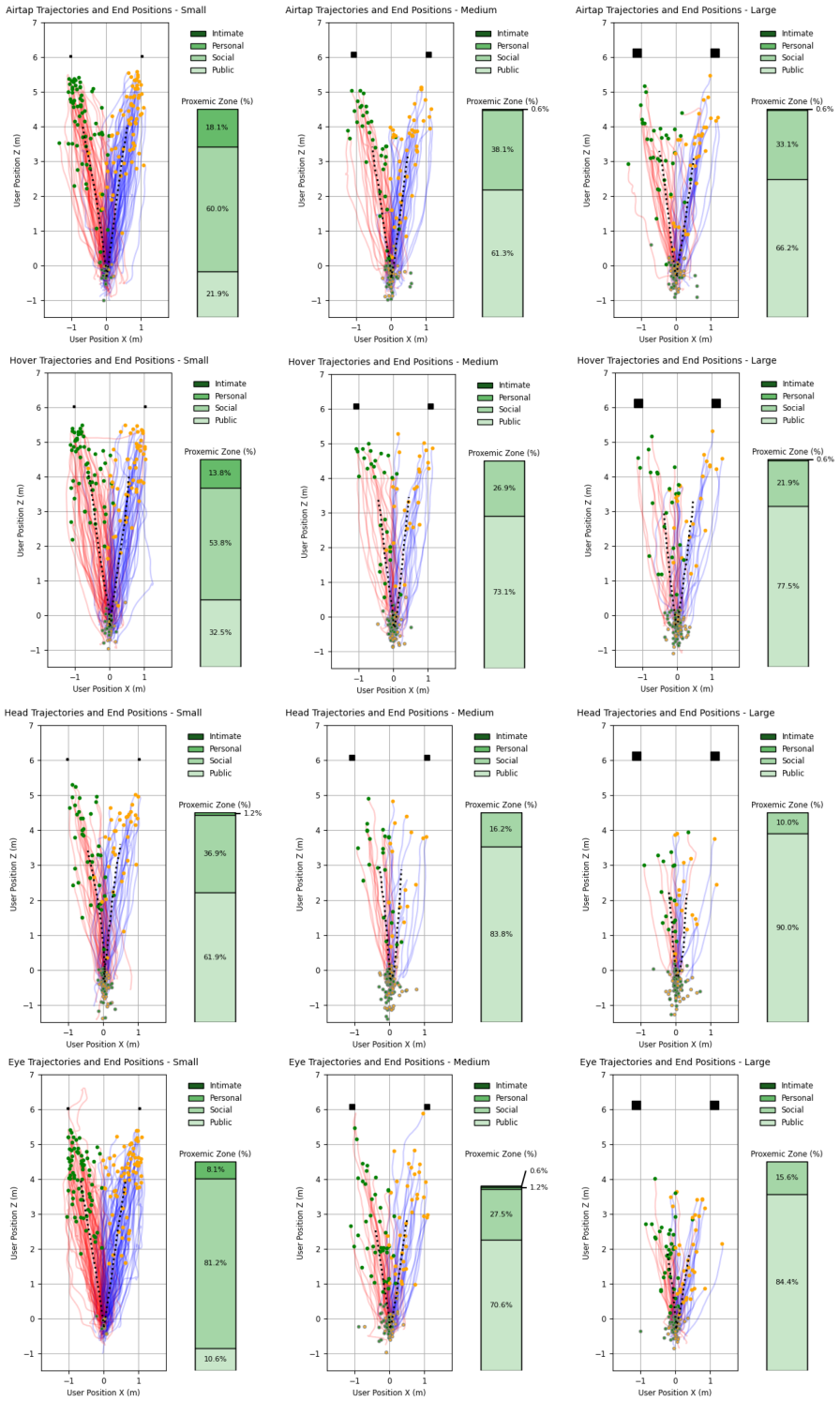

Figure 6.10: Movement Trajectories and End Positions. Black dashed lines show mean trajectories. Red trajectories and green points correlate to selections on the left and blue trajectories and orange points represent selections to the right. Half of the data was transposed to account for alternating start positions. Percentages of selections made in each proxemic zone are presented in stacked barplots.

et al., 2021).

Most users were found to interact from within Public or Social proxemic zones across all Interaction Techniques and Object Sizes. Whilst most selections were made in the Public zone with Head, even when interacting with Small objects (61.6%), most users breached the Social zone when selecting Small objects with all other techniques. Selections in the Personal proxemic zone were primarily performed using Airtap (18.1%) and Hover (13.8%) for interacting with Small objects. Several Small object selections were also made in the Personal zone using Eye (8.1%). When selecting Medium and Large objects, users were mostly interacting from the Public zone across all techniques.

Locomotion Approach and Speed

The end positions illustrated in Figure 6.10 were reached by following different approaches to locomotion. First, it was considered if users chose to walk or not, and if so, whether they were Stationary or primarily Decelerating or Accelerating in speed during different time windows leading up to each selection.

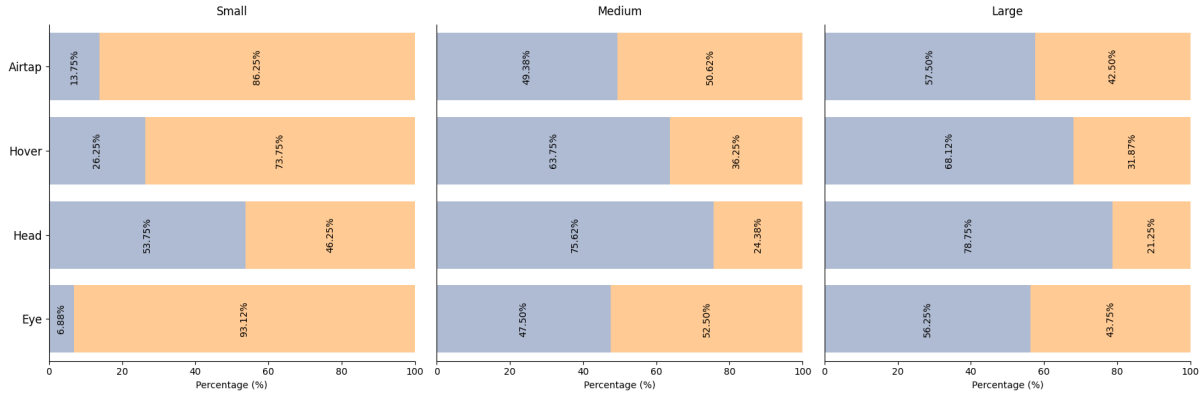

Figure 6.11: Percentage of users stationary or walking during the entire trial, showing the percentage of trials where participants were STATIONARY or ACTIVE. 100% is equal to 160 trials.

Approach Figure 6.11 presents the percentage of trials where the user was stationary (or, as defined in Section 6.3.1, did not take a step by walking more than 0.63m (Koop et al., 2020)) compared to Active (approached the target cube) for each Interaction Tech-

nique and Object Size. Participants interacted from a stationary position most frequently with Head, followed by Hover, Airtap and Eye. This was the case across all object sizes. Percentages for approaches taken using Airtap and Eye were are notably comparable, with only a 1.88% difference for Medium objects, and 1.25% difference for Large Objects.

When focusing on cases where participants did walk (were active), additional analysis was conducted to sort approaches into three categories. Approaches were considered by observing different time windows of data: 'Entire Trial', 'Last 5 Seconds', 'Last 3 Seconds' and 'Last Second' before the end of the trial.

When analysing 'Entire Trial Duration', trials were classified as 1) Decelerating: where users were primarily slowing down or 2) Accelerating: where users were primarily speeding up (see Section 6.3.1). For the remaining time windows, it was also considered if users were 3) Stationary: did not walk more than 0.63m in the defined duration.

		Small		Medium		Large	
Technique	Window	% of Total	% Excluded	% of Total	% Excluded	% of Total	% Excluded
Airtap	Last 5	99.4% (86.2%)	0.62% | 0.00%	71.3% (48.1%)	26.25% | 2.50%	59.4% (37.5%)	35.62% | 5.00%
	Last 3	100% (86.2%)	0.00% | 0.00%	98.1% (50.6%)	1.88% | 0.00%	85.0% (42.5%)	15.00% | 0.00%
Hover	Last 5	93.1% (73.8%)	6.88% | 0.00%	53.1% (31.2%)	41.88% | 5.00%	28.7% (22.5%)	61.88% | 9.38%
	Last 3	100% (73.8%)	0.00% | 0.00%	95.6% (36.2%)	4.38% | 0.00%	80.6% (31.9%)	19.38% | 0.00%
Head	Last 5	71.3% (45.6%)	28.12% | 0.62%	21.3% (18.8%)	73.12% | 5.62%	8.1% (7.4%)	78.12% | 13.75%
	Last 3	96.9% (46.2%)	3.12% | 0.00%	68.8% (24.4%)	31.25% | 0.00%	29.4% (18.7%)	68.12% | 2.50%
Eye	Last 5	96.3% (91.9%)	2.50% | 1.25%	33.8% (26.3%)	40.00% | 26.25%	10.7% (8.2%)	53.75% | 35.62%
	Last 3	100% (93.1%)	0.00% | 0.00%	82.5% (49.4%)	14.38% | 3.12%	53.8% (33.8%)	36.25% | 10.00%

Table 6.3: Percentage of data included in 'Last 5 Seconds' and 'Last 3 Seconds' time window analyses. For '% of Total' the first number is the percentage off all data (100% = 160 trials), with the number in brackets showing the percentage of this data which was Active. '% Excluded', represents the percentage of excluded trials that were STATIONARY or ACTIVE.

'Entire Trial Duration' and 'Final Second' included all Active data displayed in Figure 6.11. Any trials where the entire duration was less than the 'Last 3 Seconds' and 'Last 5 Seconds' thresholds were removed from the respective analyses. Table 6.3 presents a breakdown for the amount of data included/excluded in each analysis. This not only contextualises results, but also provides an insight into how task completion times were

impacted by the approach taken by users, based on Interaction Technique and Object Size.

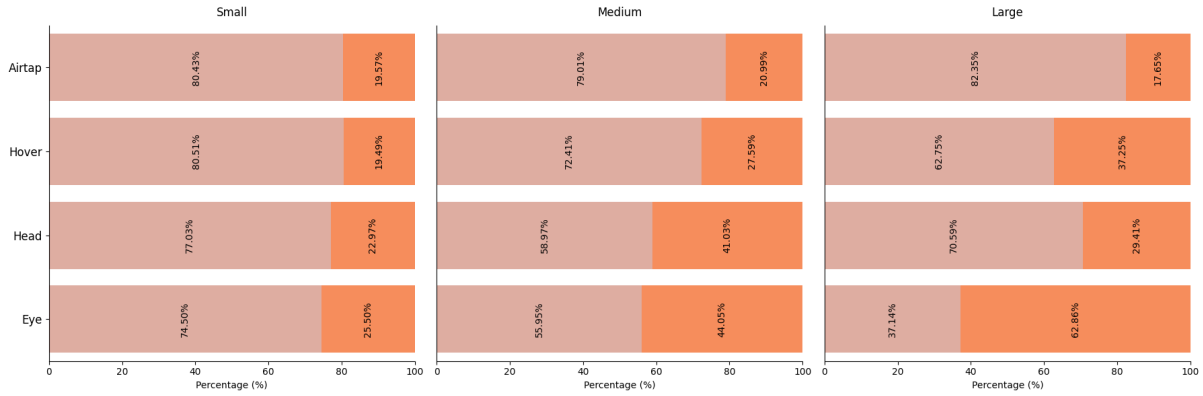

Figure 6.12: Walking patterns showcased by users across the Entire Trial duration, showing the percentage of trials where participants were DECELERATING or ACCELERATING.

Entire Trial When considering the 'Entire Trial' duration (see Figure 6.12), there was little difference in the percentage of Decelerating and Accelerating trials when selecting Small objects. Although users were more likely to accelerate with gaze-based techniques, this was more pronounced with Medium objects. While the number of Accelerating trials consistently increased with Hover and Eye as object size increased, more participants were found to be Accelerating with Airtap and Head for selecting Medium objects when compared to Large, with object size seeming to have minimal impact on walking patterns with Airtap. Participants consistently accelerated more with Eye than any other technique. This was particularly evident with Large objects, where over half (62.9%) of the trials were Accelerating.

Last 5 Seconds For the 'Last 5 Seconds' time window (see Figure 6.3.2), Airtap noticeably had more stationary trials across object sizes when compared to the other techniques. Object size was found to impact the number of trials where users remained stationary, consistently increasing as object size decreased. Similarly, the number of Accelerating trials consistently increased as object size increased across techniques. Participants again accelerated more with Eye across all object sizes, especially Large, where

50% of trials were Accelerating. No trials were classed as Stationary during the last 5 seconds when selecting Large objects with Head and Eye.

For Small object selections, less than 1% of trials were completed in less than 5 seconds with Airtap, 2.7% with Eye and 6.9% with Hover. Considerably more trials were faster than 5 seconds using Head (28.7%). Although Airtap still had the least excluded data (28.7%) when selecting Medium objects, this was followed by Hover (46.9%), Eye (66.2%) and Head (78.7%). This was also the case for Large objects: Airtap (40.6%), Hover (71.3%), Eye (89.3%), Head (91.9%). Whereas most of the excluded trials were stationary, there were noticably more active trials completed in less than 5 seconds with Eye than the other techniques.

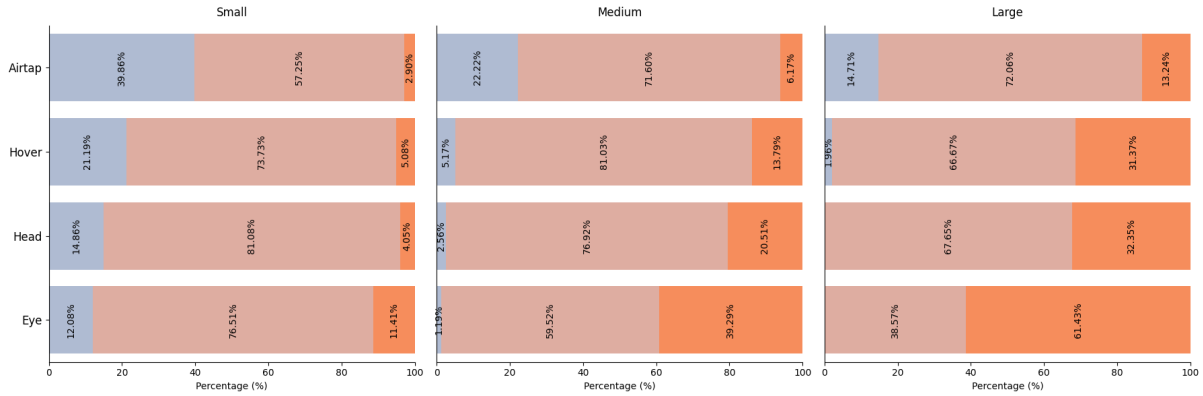

Figure 6.13: Walking patterns showcased by users in the Last 5 Seconds of the interaction, showing the percentage of trials where participants were STATIONARY, DECELERATING, or ACCELERATING.

Last 3 Seconds Most users were Stationary during the 'Last 3 Seconds' window when selecting Small objects with Airtap, Hover and Head, whereas the majority were still walking using Eye. Considerably more users had also became Stationary when selecting Medium objects with Airtap, Hover and Head, however, there was only a slight increase (0.1%) from the 'Last 5 Seconds' window with Eye. When selecting Large objects, more users had become stationary with Airtap and Hover, with some users also begining to stop with Head (6.7%) and Eye (3.7%). Participants were continuing to slow down across all techniques and object sizes, however, more participants were still Accelerating with Eye, especially when selecting Medium and Large objects (see Figure 6.14).

For Small object selections, only 3.1% were completed in less than 3 seconds using Head, with no trials being faster than 3 seconds with Airtap, Hover and Eye. When considering Medium objects, Airtap (1.9%) and Hover (4.4%) had minimal excluded data whereas Eye (17.5%) and Head (31.2%) had considerably more. This was also the case for Large objects: Airtap (15.0%), Hover (19.4%), Eye (46.2%), Head (70.6%). Most of the excluded trials were again stationary, where there were noticably more active trials completed in less than 3 seconds with Eye than the other techniques, especially for Large object selections.

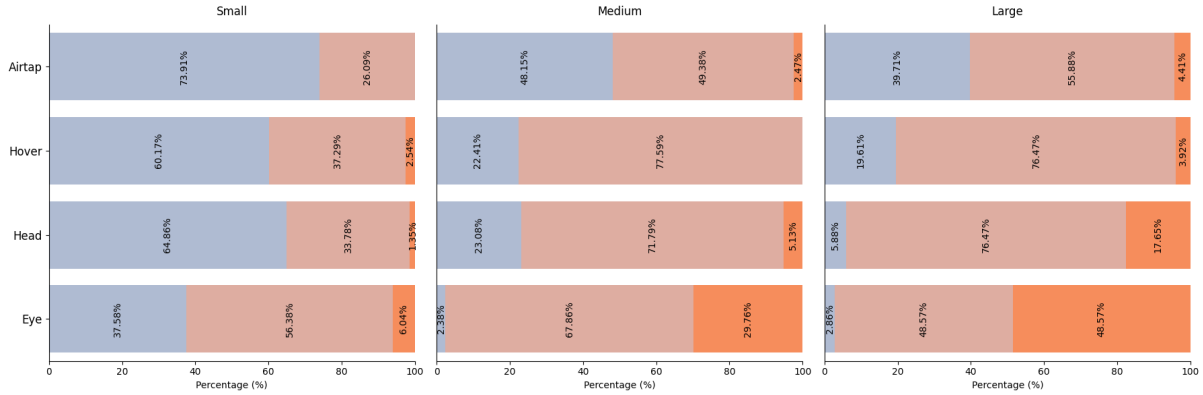

Figure 6.14: Walking patterns showcased by users in the Last 3 Seconds of the interaction, showing the percentage of trials where participants were STATIONARY, DECELERATING, or ACCELERATING.

Last Second During 'The Last Second' of the trial (see Figure 6.15), all users were stationary when selecting Small objects with Airtap, Hover and Head, however, some trials (7.4%) still involved walking with Eye as the selection was being made. Whereas most users were also Stationary when selecting Medium and Large objects with Airtap, Hover and Head, the opposite was true for Eye, where the majority of selections were made whilst walking. Although Accelerating trials were consistently reducing until this point, when compared to the previous time window, 3.45% more trials were Accelerating when selecting Medium object with Hover, as well as 1.47% more trials when selecting Large objects with Airtap.

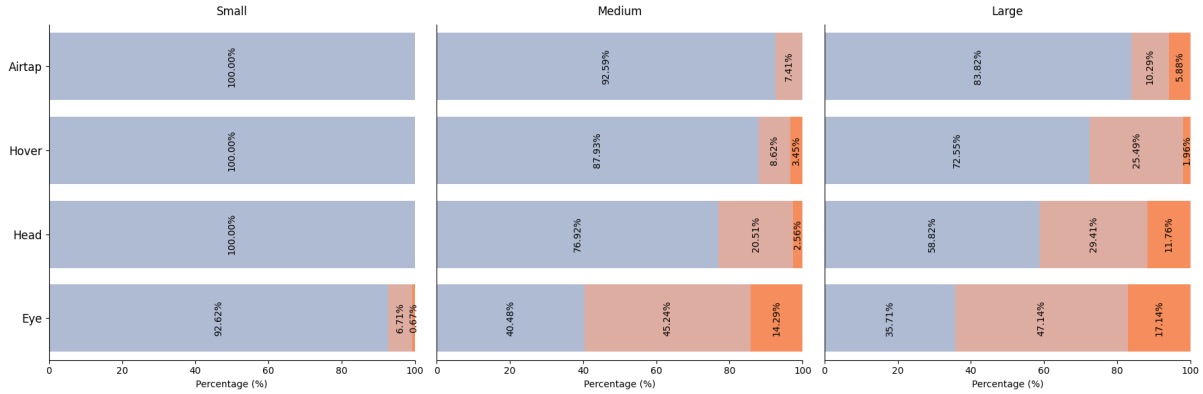

Figure 6.15: Walking patterns showcased by users in the Last second of the interaction, showing the percentage of trials where participants were STATIONARY, DECELERATING, or ACCELERATING.

Speed When considering the 'Entire Trial Duration' across the full data set (including stationary trials) for selecting Small objects, the highest average walking speed was Eye (M=0.44m/s, SD=0.23), followed by Airtap (M=0.35m/s, SD= 0.23), Hover (M= 0.33m/s, SD= 0.26) and Head (M= 0.23m/s, SD= 0.25). This was also the case for Medium objects: Eye (M= 0.34m/s, SD= 0.33), Airtap (M= 0.27m/s, SD= 0.28), Hover (M= 0.23m/s, SD= 0.30), Head (M= 0.15m/s, SD= 0.25), and Large objects: Eye (M= 0.28m/s, SD= 0.31), Airtap (M= 0.23m/s, SD= 0.28), Hover (M= 0.21m/s, SD= 0.30), Head (M= 0.14m/s, SD= 0.23).

Results are now presented based on only active trials, focusing on instances where users walked (see Figure 6.11 and Table 6.3). It is important to note that variences in sample size across groups may impact the statistical power of comparisons, e.g., in the 'Last 5 Seconds' time window, where Head (n = 12) and Eye (n = 13) have considerably fewer trials included in analyses than Airtap (n = 60) and Hover (n = 36).

Entire Trial As presented in Figure 6.16, when considering 'Entire Trial Duration' (excluding instances where users remained stationary) for selecting Small objects, the highest average walking speed was Eye (M=0.47m/s, SD=0.21) and Head (M=0.47m/s, SD= 0.17), followed by Hover (M= 0.45m/s, SD= 0.21) and Airtap (M= 0.40m/s, SD= 0.20). For Medium objects, Eye was again highest (M= 0.61m/s, SD=

0.23), however this was followed by Hover (M= 0.58m/s, SD= 0.24), Head (M= 0.55m/s, SD= 0.22) and Airtap (M= 0.50m/s, SD= 0.22). For Large object selections, the highest average walking speed was reached using Hover (M= 0.60m/s, SD= 0.25), followed by Eye (M= 0.59m/s, SD= 0.22), Head (M= 0.56m/s, SD= 0.20) and Airtap (M= 0.50m/s, SD= 0.23). All calculations in this time window were based on the data in Figure 6.11, where 100% of data represents 160 trials.

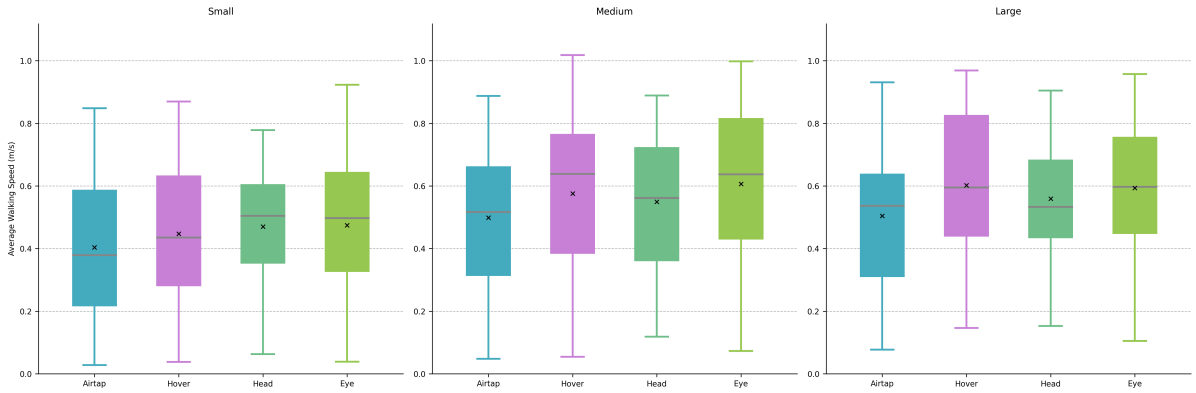

Figure 6.16: Locomotion Speed showcased by users throughout the Entire Trial. Mean times are indicated by 'X'.

A significant main effect was found for Technique on Average Walking Speed ($F_{3,917.58} = 16.916, p < 0.001, \eta_p^2 = 0.021$). There was also a significant main effect of Size ($F_{2,918.93} = 6.607, p < 0.001, \eta_p^2 = 0.059$). The interaction between Technique and Size was not significant ($F_{6,915.80} = 0.464, p = 0.835$), indicating that the influence of technique was not dependent on object size.

For the main effect of Technique, post-hoc comparisons with Bonferroni correction revealed that Eye resulted in significantly higher average speeds compared to Airtap ($p < 0.001$), Head ($p < 0.001$), and Hover ($p < 0.001$). No significant differences were found between Airtap and Head, Airtap and Hover, or Head and Hover. For Size, significant differences were found between Medium and Small objects ($p < 0.001$), with Medium objects resulting in higher average speeds.

Last 5 Seconds When analysing the 'Last 5 Seconds', for trials involing Small objects, the highest average walking speed was produced with Eye (M=0.50m/s, SD=0.27),

followed by Head (M=0.43m/s, SD= 0.23), Hover (M= 0.40m/s, SD= 0.27) and Airtap (M= 0.30m/s, SD= 0.27). Eye was also highest for Medium objects (M= 0.70m/s, SD= 0.27), however this was followed by Hover (M= 0.62m/s, SD= 0.26), Head (M= 0.60m/s, SD= 0.25) and Airtap (M= 0.48m/s, SD= 0.31). For Large object selections, the highest average walking speed was reached with Hover (M= 0.68m/s, SD= 0.29) followed by Head (M= 0.53m/s, SD= 0.22). Eye (M= 0.48m/s, SD= 0.27) and Airtap (M= 0.48m/s, SD= 0.28) produced the lowest walking speeds (see Figure 6.17). All calculations in this time window were based on the data in Table 6.3, where 100% of data represents 160 trials.

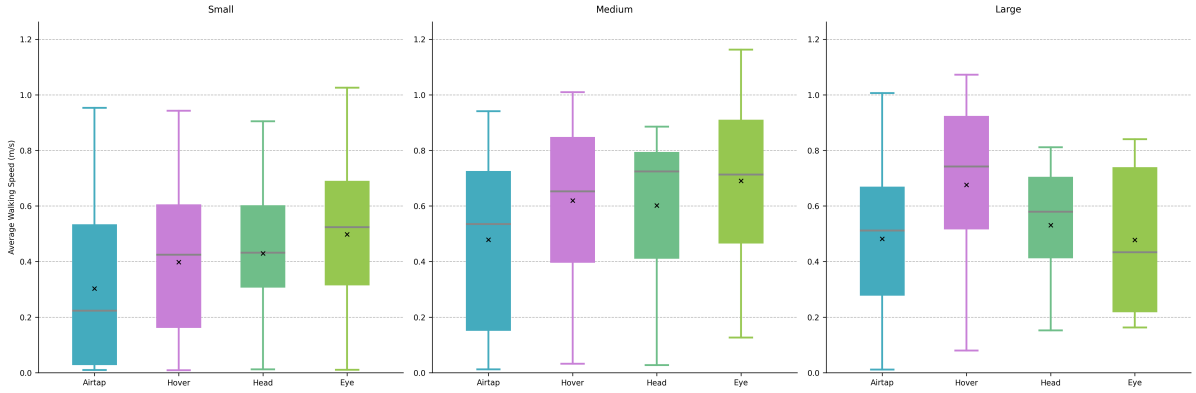

Figure 6.17: Locomotion Speed showcased by users in the Last 5 Seconds of interactions. Mean times are indicated by 'X'.

Akin to the 'Entire Trial' duration, a significant main effect was found on Average Walking Speed for Technique ($F_{3,748.39} = 13.267, p < 0.001, \eta_p^2 = 0.0308$) and Size ($F_{2,749.91} = 23.035, p < 0.001, \eta_p^2 = 0.0694$), with no interaction found between Technique and Size ($F_{6,745.92} = 1.269, p = 0.270$).

For Technique, post-hoc comparisons with Bonferroni correction revealed that Eye resulted in significantly higher average speeds compared to Airtap ($p < 0.001$) and Hover ($p < 0.01$). Additionally, participants were found to walk significanltly faster with Hover compared to Airtap ($p < 0.01$). No significant differences were found between Airtap and Head, Eye and Head, or Head and Hover. For Size, significant differences were again found between Small and Medium objects ($p < 0.001$), as well as between Small and Large objects ($p < 0.001$), with Medium and Large selections resulting in higher average

speeds than Small. No significant differences were observed between Medium and Large objects.

Last 3 Seconds During the 'Last 3 Seconds' (see Figure 6.18), participants walked fastest to select Small objects using Eye (M=0.36m/s, SD=0.28), followed by Hover (M=0.20m/s, SD= 0.20), Head (M= 0.19m/s, SD= 0.17) and Airtap (M= 0.15m/s, SD= 0.19). When selecting Medium objects, Eye again provided the highest average walking speed (M= 0.76m/s, SD= 0.30), however, this was followed by Head (M= 0.56m/s, SD= 0.35), Hover (M= 0.50m/s, SD= 0.31) and Airtap (M= 0.35m/s, SD= 0.31). Eye also provided the highest average walking speed for Large objects (M= 0.71m/s, SD= 0.30), followed by Hover (M= 0.62m/s, SD= 0.36), Head (M= 0.61m/s, SD= 0.29) and Airtap (M= 0.39m/s, SD= 0.34). All calculations in this time window were based on the data in Table 6.3, where 100% of data represents 160 trials.

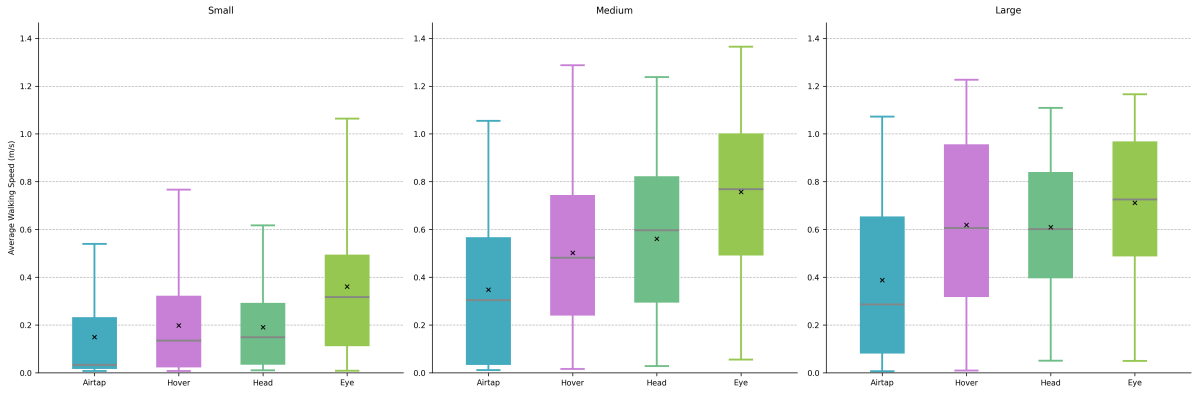

Figure 6.18: Locomotion Speed showcased by users in the Last 3 Seconds of interaction. Mean times are indicated by 'X'.

Unlike the previous time windows, a significant interaction effect was found between Technique and Object Size ($F_{6,893.07} = 6.395, p < 0.001, \eta_p^2 = 0.030$). Post-hoc analysis revealed that, when selecting Small objects, Eye produced faster average walking speeds than Airtap, Head, and Hover ($p < 0.001$). No significant differences were observed between Airtap, Head, and Hover.

For Medium object selections, Eye was significantly faster than Airtap ($p < 0.001$), Hover ($p < 0.001$) and Head ($p < 0.01$). Additionally, Head was significantly faster than

Airtap ($p < 0.01$). When selecting Large objects, Airtap was also significantly slower than Eye ($p < 0.001$), Head ($p < 0.001$), and Hover ($p < 0.01$). No significant differences were found among Eye, Head, and Hover, indicating similar walking speeds across these techniques. A full breakdown of pairwise comparisons is provided in Table 6.4.

Table 6.4: Average Walking Speeds (Last 3 Seconds): Results of pairwise comparisons based on Interaction Technique and Object Size, where *** p=<0.001, ** p= 0.01 and * p= 0.05. Arrows indicate which technique was used at greater speeds (m/s), H-AT: Airtap, H-H: Hover, H-G: Head or E-G: Eye.

Size		Small				Medium				Large			
	Technique	H-AT	H-H	H-G	E-G	H-AT	H-H	H-G	E-G	H-AT	H-H	H-G	E-G
Small	H-AT		↑	↑	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
	H-H			←	↑***	↑**	↑***	↑***	↑***	↑***	↑***	↑***	↑***
	H-G				↑***	↑*	↑***	↑***	↑***	↑***	↑***	↑***	↑***
	E-G					←	↑	↑	↑***	←	↑	↑*	↑***
Medium	H-AT						↑	↑**	↑***	↑	↑***	↑***	↑***
	H-H							↑	↑***	←	↑	↑	↑***
	H-G								↑**	←*	↑	↑	↑*
	E-G									←***	←**	←	←
Large	H-AT										↑**	↑***	↑***
	H-H											↑	↑
	H-G												↑
	E-G												

Last Second When considering the 'Last Second' of active trials (see Figure 6.19), Eye produced the highest walking speed (M= 0.18m/s, SD= 0.25), with Head (M= 0.04m/s, SD= 0.03), Hover (M= 0.04m/s, SD= 0.06) and Airtap (M= 0.04m/s, SD= 0.06) achieving the same average speed. For Medium object selections, the highest average walking speed was reached with Eye (M= 0.71m/s, SD= 0.37), followed by Head (M= 0.34m/s, SD= 0.39), Hover (M= 0.24m/s, SD= 0.30) and Airtap (M= 0.15m/s, SD= 0.24). This was also the case for Large object selections: Eye (M= 0.78m/s, SD= 0.38), Head (M= 0.52m/s, SD= 0.43), Hover (M= 0.38m/s, SD= 0.37), Airtap (M= 0.22m/s, SD= 0.30). All calculations in this time window were based on the data in Figure 6.11, where 100% of data represents 160 trials.

A significant interaction effect was again found between Technique and Object Size ($F_{6,920.54} = 47.895, p < 0.001, \eta_p^2 = 0.183$). Post-hoc analysis revealed that, when selecting

Small objects, users walked faster with Eye than Airtap, Head, and Hover ($p < 0.001$). Again, no significant differences were observed between Airtap, Head, and Hover for Small objects.

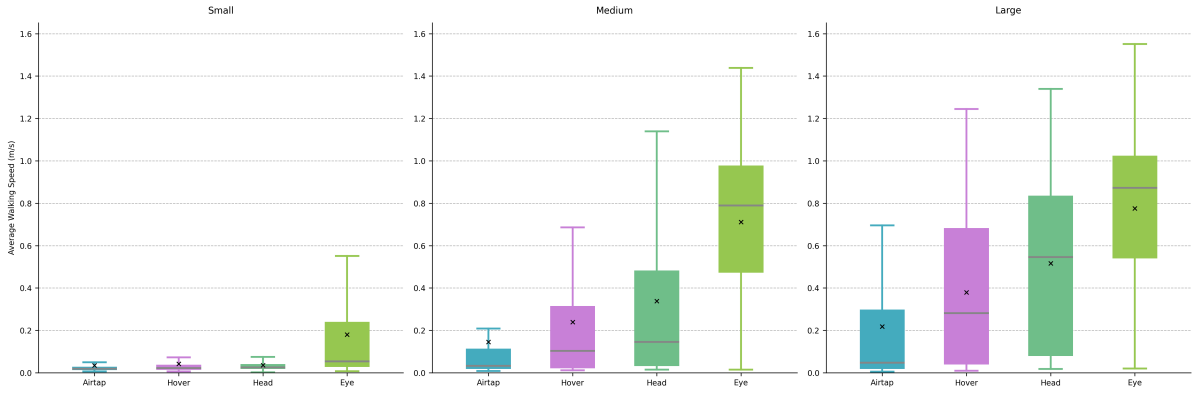

Figure 6.19: Locomotion Speed showcased by users in the Last second of interaction. Mean times are indicated by 'X'.

Table 6.5: Average Walking Speeds (Last Second): Results of pairwise comparisons based on Interaction Technique and Object Size, where *** p=<0.001, ** p= 0.01 and * p= 0.05. Arrows indicate which technique was used at greater speeds (m/s), H-AT: Airtap, H-H: Hover, H-G: Head or E-G: Eye.

Size		Small				Medium				Large			
	Technique	H-AT	H-H	H-G	E-G	H-AT	H-H	H-G	E-G	H-AT	H-H	H-G	E-G
Small	H-AT		↑	↑	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
	H-H			↑	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
	H-G				↑***	↑	↑***	↑***	↑***	↑***	↑***	↑***	↑***
	E-G					←**	←	↑	↑***	←	↑	↑***	↑***
Medium	H-AT						↑	↑***	↑***	↑	↑***	↑***	↑***
	H-H							↑	↑***	←	↑	↑***	↑***
	H-G								↑***	←	←	↑	↑***
	E-G									←***	←***	←	↑
Large	H-AT										↑	↑***	↑***
	H-H											↑	↑***
	H-G												↑
	E-G												

For Medium object selections, Eye was significantly faster than Airtap ($p < 0.001$), Head ($p < 0.001$), and Hover ($p < 0.001$), with participants also walking significantly faster with Head than Airtap ($p < 0.001$). No significant differences were observed between Head and Hover, or Airtap and Hover.

When selecting Large objects, Eye was significantly faster than Airtap ($p < 0.001$)

and Hover ($p < 0.001$). Head was also significantly faster than Airtap ($p < 0.001$). Table 6.5 provides a full breakdown of pairwise comparisons.

Task Load

When analysing average task load, significant main effects were found between techniques ($F_{3,117} = 11.93, p < 0.001$). Overall, Head was deemed to require the least taskload (M=34.46, SD=18.12) followed by Hover (M=40.20, SD=18.51), Eye (M=41.71, SD=21.04) and Airtap (M=54.09, SD=20.08). Significant differences were also found across subscales. When considering Mental demand, scores for Head and Hover were significantly lower ($p < 0.001$) than Airtap, with Eye being found more mentally demanding than Head ($p < 0.05$). Airtap was also deemed to be significantly more physically demanding than Head and Eye ($p < 0.001$).

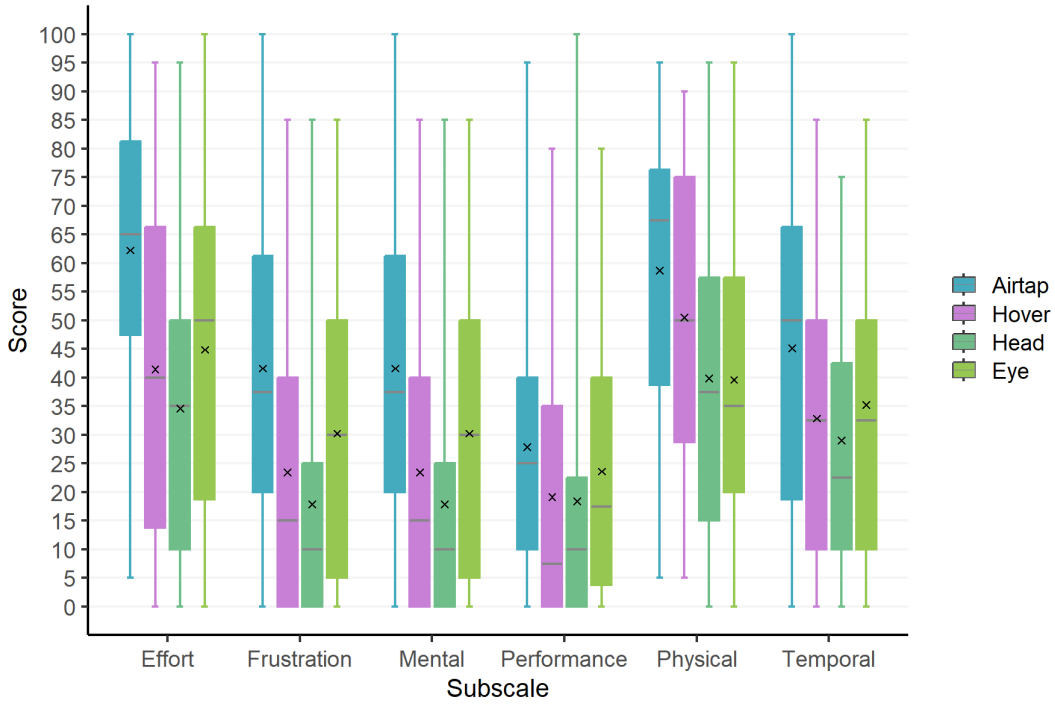

Figure 6.20: Boxplots showing distribution of NASA-TLX subscale scores for *Airtap*, *Hover*, *Head* and *Eye*. Mean scores are indicated by 'X'.

In terms of Performance and Temporal demand, Airtap was rated worse than Head ($p < 0.01$) and Hover ($p < 0.05$) and was considered to require significantly more Effort

than Head ($p < 0.001$), Hover ($p < 0.01$) and Eye ($p < 0.05$). Airtap was also found to be more Frustrating than Head and Hover ($p < 0.001$), with Eye being considered more Frustrating than Head ($p < 0.05$). A breakdown of NASA-TLX results is provided in Figure 6.20.

User Experience

Head (M=1.99, SD=0.81) was considered to offer a better overall user experience than Eye (M=1.73, SD=0.92), Hover (M=1.68, SD=0.88) and Airtap (M=0.93, SD=1.17). When analysing mean UEQ Scores, ART ANOVA tests revealed that Technique produced a significant main effect ($F_{3,117} = 17.23, p < 0.001$), with post-hoc pairwise comparisons showing that Head, Eye and Hover provided significantly better user experience than Airtap ($p < 0.001$).

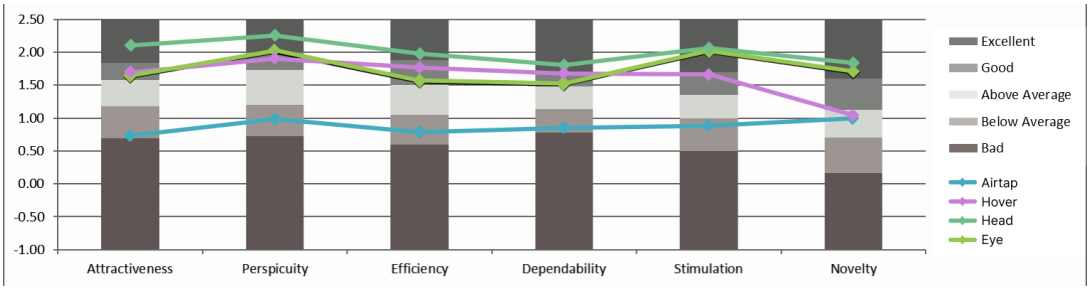

Figure 6.21: UEQ Subscale Scores for *Airtap*, *Hover*, *Head* and *Eye*. Mean ratings from Bad to Excellent.

Significant differences were also found between Interaction Techniques across all subscales: Attractiveness ($F_{3,117} = 12.52, p < 0.001$), Dependability ($F_{3,117} = 9.04, p < 0.001$), Efficiency ($F_{3,117} = 12.25, p < 0.001$), Novelty ($F_{3,117} = 9.73, p < 0.001$), Perspicuity ($F_{3,117} = 13.16, p < 0.001$) and Stimulation ($F_{3,93} = 15.11, p < 0.001$).

Head ($p < 0.001$), Hover ($p < 0.001$) and Eye ($p < 0.01$) were rated higher than Airtap in terms of Attractiveness. This was also the case for Dependability: Head ($p < 0.001$), Hover ($p < 0.001$), Eye ($p < 0.05$), Efficiency: Head ($p < 0.001$), Hover ($p < 0.001$), Eye ($p < 0.01$), Perspicuity ($p < 0.001$) and Stimulation ($p < 0.001$). Head ($p < 0.001$) and Eye ($p < 0.01$) were also deemed to have higher Novelty than Airtap

and Hover. Figure 6.21 provides an overview of UEQ scores.

Preference

After completing all conditions, users ranked the interaction techniques in order of preference. Head was ranked first most frequently (52.5%), followed by Eye (32.5%), Hover (10.0%) and Airtap (5.0%). Subjective feedback further highlights the advantages and limitations of each technique, where Head was deemed to be the "Easiest" and "Most Accurate" (N= 18). Participants also appreciated that they did not have to move closer to objects to select them (N= 8): "I found having to move pointless", "I didn't like moving so preferred head".

Although some participants found Eye to be "Fun" and/or "Enjoyable" (N= 4) and "Easiest to use when walking" (N= 7), several noted that it was difficult to select small objects (N= 7): "I couldn't select small objects from distance", "you had to walk closer to small objects". Some also commented on eye strain when using the technique (N= 6): "It made my eyes tired", "I felt I had to really strain my eyes".

Despite the limitations of Eye, Airtap was considered to be the most frustrating (N= 16) and to require the most effort (N= 9): "When trying to pinch, my hand kept moving so it wouldn't select", "It was over-complicated", "It was by far the most difficult".

Similarly, compared to Head and Eye, Hover was received to be difficult to employ (N= 7) and physically demanding (N= 6): "I found it tricky", "I had trouble getting the distance of the ray right", "It was more effort".

6.4 Discussion

This chapter has explored the affordances of four interaction techniques (*Airtap*, *Hover*, *Head* and *Eye*) and their impact on user-defined locomotion in AR when performing fun-

damental selection tasks. This section highlights the key findings of the study conducted, discussing how the methods followed by participants can be considered in the design of AR interaction techniques.

Head was the most accurate technique at distance

In line with Study 2 (Chapter 5), users were able to select objects with Head from farther distances more easily than the other techniques explored. This was especially the case for Small and Medium objects, where Head surpassed the performance of all other techniques. This is also in line with previous findings, where head gaze was found to provide high performance for targeting and selection, generally outperforming eye gaze and other combinations of modalities in terms of speed and accuracy (Kytö et al., 2018; Y. Y. Qian and Teather, 2017).

In the context explored, this ease of use likely contributed to quicker interaction times, as users could make selections without needing to adjust their position as frequently. This is emphasised by Table 6.3, which suggests that faster selection times were primarily achieved during trials where users remained stationary. This was something that was highly appreciated by users with the Head technique, many stating that they liked not having to move closer to objects to select them. Two participants likened this to interactions with a tv remote control, stating that they do not expect to have to get close to a tv set to use that effectively, and similarly, they did not want to feel forced to get closer with AR interaction techniques either. This finding is further supported by users primarily interacting from public or social proxemic zones across all interaction techniques and object sizes. It is also in line with previous research, where participants have reported that walking to reach a target, especially when placed far away, introduced additional fatigue and discomfort (Bhowmick, Kalita, and Sorathia, 2020).

Although Head was considered the most appropriate technique for interacting whilst stationary, some participants (n = 4) explicitly mentioned that they had difficulties

when attempting to make selections whilst walking with Head. This could be due to head rotations that occur along with the torso when walking, as well as head translatations, where at a moderate to fast speed (1.4–1.8 m/s), the head has been found to move up and down with an average frequency of 2 Hz and left and right with a frequency of 1 Hz (Manakhov et al., 2024). F. Lu, Davari, and D. Bowman (2021) found that using Head was the least favoured over Eye and Hand techniques when accessing virtual content during walking, participants reporting that they had to fully stop in order to perform the interaction. This reaffirms that the Head technique is likely more suitable for interactions where the user is stationary, and could be employed as a defacto pointing method for selection tasks where accuracy is a key requirement, especially when interacting with content that is small and far away.

Eye was the most appropriate for interacting whilst in motion

Users highlighted that a primary benefit of Eye was its affordance to interact in motion (n = 7), with results showing that users were more likely to select objects whilst walking with Eye than any other technique. Despite Eye and Airtap having a very similar number of Active trials (see Figure 6.11), when considering the 'Entire Trial' duration, the number of those that were Accelerating for Medium object selections more than doubled with Eye when compared to Airtap and more than tripled for Large objects, underscoring users confidence to walk while using the Eye technique.

Walking approaches across time windows show that participants consistently Accelerated the most with Eye across all techniques and object sizes. Users were also becoming stationary consistently soonest with Airtap and latest with Eye in the lead up to selections being made. During the 'Last Second' of trials, the benefits of employing Eye whist walking are most emphasised, where Eye was the only technique to afford interactions whilst in motion when selecting Small objects. Furthermore, more than half of those who appraoched Medium and Large objects with Eye were in motion as the selection was being made. This was considerabley more than those walking with other

techniques, especially Airtap.

This finding is further highlighted when analysing the outcomes of selection time and distance for interactions with large objects. Here, no statistically significant differences were observed in selection time between the Head and Eye techniques, even though users employed Eye considerably closer to target objects. This again indicates increased walking speeds employed with Eye, and points to the advantages of employing the technique for selection tasks whilst in motion.

The advantage to interact whilst walking with Eye could be attributed to the inherent human ability to fixate on objects whilst moving, known as the vestibulo-ocular reflex (Palomino-Roldan, Rojas-Cessa, and Suaste-Gomez, 2023), which would enable users to perform selections while either sustaining or increasing their walking speed more often. This reflex mechanism describes how eye movements are generated in the opposite direction of head movement to preserve the same amplitude of motion, enabling individuals to maintain visual focus on a target of interest.

Despite this benefit, subjective feedback in section 6.3.2 suggests that although users could select whilst walking, this was not always desirable, where some users reported feeling forced to approach smaller content to interact effectively with Eye. In line with Chapter 5, although still providing better performance than Airtap, the accuracy of the Eye technique was found to be heavily impacted by the angular size of objects. This finding is emphasised by results in Section 6.3.2, where it is evident that users were Accelerating consistently less when selecting Small objects with Eye compared to Medium and Large. These lower walking speeds could be attributed to the heightened precision and concentration required for the selection of smaller targets (C. Zhao, K. W. Li, and Peng, 2023; Whitlock et al., 2018).

Gaze techniques were preferred over Hand techniques

Currently, freehand interaction is most often used as the default interaction method when interfacing with AR applications (Hertel et al., 2021; Spittle, Frutos-Pascual, et al., 2022). Despite this, freehand interactions have recently been amongst the least preferred and/or efficient techniques in the context of fundamental selection tasks, with gaze-based selection often facilitating more efficient interactions (Pfeuffer, B. Mayer, et al., 2017; Heo et al., 2020; F. Lu, Davari, Lisle, et al., 2020).

Subjective results suggest that preference for gaze techniques could be attributed to them being straightforward to learn and employ, as they simply require users to direct their focus towards interactive components (Pfeuffer, Abdrabou, et al., 2021; Xinyi Liu et al., 2022). This makes selections more straightforward and efficient, as, when involving a search task, the interaction can seem like a one-step process. Conversely, with freehand techniques, users must first conduct a visual search task, before positioning their hand within a restricted interaction zone (W. Xu, Liang, Y. Chen, et al., 2020) and guiding the cursor over the object to make a selection. This is supported by results for NASA TLX, where Head provided the lowest overall taskload and Airtap the highest. Although Eye was considered to have low physical demand, the level of mental demand was relitevley high compared to Head and was considered more Frustrating, likely due to the issues around selecting Small targets. Further, when considering UEQ scores, participants found Airtap to be significantly worse than all other techniques, with gaze techniques providing the best overall user experience.

Although considerably fewer trials were Active with Head (see Figure 6.11), which will likely impact pairwise comparisons, results suggest that those who did walk with the technique were also able to walk as fast as Eye during the last 5 seconds of interaction, and faster than with Airtap for selecting Medium and Large objects during the Last 3 Seconds and Last Second time windows. This indicates that some users were relatively comfortable walking with Head, with results showing a higher percentage of Active trials

Table 6.6: Advantages and disadvantages of interaction techniques in a room-scale environment.

Technique	Findings
Airtap	- Highest overall task load and lowest UEQ ratings - Most physically demanding - Difficult to employ when walking - Least preferred
Hover	+ Provided better performance and user experience than *Airtap* - Physically demanding - Difficult to employ when walking
Head	+ Fastest for selecting small/medium objects at a distance + Lowest overall task load + High accuracy - Difficult to employ when walking
Eye	+ Most ideal for making selections while walking + Fast for selecting larger targets - More challenging to select small distant targets - Potential for eye strain/fatigue, especially when interacting with smaller objects

resulting in selections whilst in motion when compared to freehand techniques.

6.5 Summary

The exploration presented forms a critical component in developing a more user-centric approach to interactions in AR, where techniques can be adapted to key contextual factors like distance and movement. Findings underscore the importance of spatial considerations in enhancing the practicality of AR technologies, where different techniques have been found to have distinct advantages and disadvantages. This notably includes Head being an ideal pointing method for interacting with small, distant content, and Eye being beneficial where content is larger, closer and/or the user is in motion. Table 6.6 summarises the findings from the study conducted.

After considering the impact of user-object distance and user-defined distance and movement approaches in a room-scale environment, the next chapter compiles findings

from all studies and continues to explore the potential of referencing spatial context to provide adaptive interaction techniques. The chapter considers how context can be framed by dimensions of proxemic interaction, focusing on how systems can be adapted and techniques provided interchangeably based on not only Distance and Movement, but also Orientation, Identity and Location. This is timely as we move towards an era of human-computer collaboration, where research is driven by using intelligent systems that are capable of adapting to individual users, their preferences, and a range of dynamic use cases and environments (Grubert et al., 2017; Seeliger, Weibel, and Feuerriegel, 2022; X. B. Liu et al., 2024).

Chapter Seven

Discussion

Contents

7.1 Fulfillment of Research Questions **172**
7.1.1 RQ1: Gaps in Current Research 172
7.1.2 RQ2: Seated Near-field Selection 173
7.1.3 RQ3: Seated Far-field Selection 174
7.1.4 RQ4: User-Defined Distance and Movement 175
7.2 Distance and Movement **176**
7.2.1 Distance 176
7.2.2 Movement 179
7.3 Orientation, Identity and Location **183**
7.3.1 Orientation 183
7.3.2 Identity 186
7.3.3 Location 188
7.4 Limitations **189**
7.4.1 Users 190
7.4.2 Techniques 190
7.4.3 Technology 191
7.4.4 Task 191
7.4.5 Environment 192
7.4.6 Adaptation Possibilities 192
7.4.7 Qualitative Insights 193
7.5 Summary **193**

This thesis has investigated the potential for adapting peripheral-free input methods for AR interaction through two comprehensive literature reviews and three empirical studies. This chapter highlights how the research questions provided in Chapter 1 have been addressed and summarises the contributions provided. By considering the key takeaways from the three studies conducted, results are discussed and recommedations presented to provide guidelines on how to explore and facilitate more flexible AR experiences based on spatial factors. It is first considered how Distance and Movement could be referenced as primary streams of context, providing opportunities for how to adapt techniques when completing fundamental selection tasks. Following this, the potential of harnessing wider Proxemic Dimensions that were not considered as part of the empirical studies, Orientation, Identity and Location, is also explored. Limitations of the research and concepts presented are also discussed at the end of the chapter.

7.1 Fulfillment of Research Questions

This section provides an overview for how the Research Questions presented in Chapter 1 have been addressed:

7.1.1 RQ1: Gaps in Current Research

Initially, the work presents two literature reviews; 1) a narrative review on peripheral-free interaction and context-awareness in AR (Chapter 2) alongside 2) a systematic review of existing approaches to researching explicit AR interaction (Chapter 3), highlighting key gaps in current research. The work presented in Chapter 3 partly consists of a journal paper that is gaining traction in the XR research community:

B. Spittle, M. Frutos-Pascual, C. Creed and I. Williams, “A Review of Inter-

action Techniques for Immersive Environments," in IEEE Transactions on Visualization and Computer Graphics, vol. 29, no. 9, pp. 3900-3921, 1 Sept. 2023, doi: 10.1109/TVCG.2022.3174805.

This has addressed **RQ1.** — What approaches exist for working towards context-aware AR, how have input methods previously been explored for AR interaction, and what are the gaps in the current research landscape?

After surveying the state of the art in AR interaction, several research gaps were highlighted, and recommendations for future work were offered to provide directions for AR researchers. For the scope of the thesis, focus was given to 1) how users activity/situation will impact interaction, specifically their pose (sitting, standing), distance, and locomotion approaches when interacting with virtual content using different techniques, and 2) how interaction techniques could be employed interachangably based on these contextual factors. After establishing the focus of the research, three empirical studies were planned, designed and conducted to evaluate the performance and user experience of commonplace interaction techniques.

7.1.2 RQ2: Seated Near-field Selection

Chapter 4 reports on a user study involving 32 participants, which compares two freehand techniques, (Press and Hover), against two gaze-based methods, (Eye and Head), for selection within the intimate (>0.5m) proxemic zone (E. T. Hall, 1966). The objective was to provide a grounding for how these techniques could be used interchangeably, focusing on their respective strengths and limitations for near-field interaction. Techniques were assessed based on selection time, error rate, task load, user experience, and user preference. The chapter first highlights a range of AR applications involving near-field selection tasks while seated, with the study contributing a better understanding for how different interaction techniques perform, and how they are received by users. Results notably highlight the importance of considering how to balance interaction efficiency with user

experience, providing recommendations for how to employ the most appropriate techniques for fundamental selection techniques based on performance and preference (see figure 7.1).

This has addressed **RQ2.** — How effective are freehand and gaze-based techniques for near-field selection in seated AR environments?

After uncovering the advantages and limitations of Press, Hover, Head and Eye for interaction within the intimate proxemic zone, exploration is extended to far-field interaction, where the suitability of techniques are considered for interaction across personal, social and public proxemic zones when content is beyond arms reach.

7.1.3 RQ3: Seated Far-field Selection

Building on findings from Chapter 4, another user study involving 32 participants is presented in Chapter 5. The study again compares two freehand techniques, (Airtap and Hover), against two gaze-based methods, (Eye and Head), however, selections are performed across Personal (0.5m-1.0m), Social (1.0m-4.0m) and Public (>4.0m) proxemic zones (E. T. Hall, 1966; Whitlock et al., 2018). The baseline Press interaction paradigm considered in the previous study was modified to the baseline Airtap gesture, with techniques again being assessed based on selection time, error rate, task load, user experience, and user preference. The objective was to establish a grounding for how techniques could be used interchangeably, focusing on their respective strengths and limitations based on distance. The chapter begins by highlighting a range of AR applications involving far-field selection tasks while seated, and contributes to a better understanding for how interaction techniques can be adapted based on user-object distances. Results show that the distance and angular size of objects are key factors in defining the suitability of selection techniques, and again offers recommendations for how to employ input methods interchangeably based on performance and preference at different distances (see figure 7.1).

This has addressed **RQ3.** — How do freehand and gaze-based techniques perform for far-field selection across varying distances in seated AR environments?

After identifying the advantages and limitations of Airtap, Hover, Head and Eye techniques for far-field interaction across Personal, Social and Public proxemic zones, Study 3 looks beyond distance to also consider user-defined movement approaches in room-scale AR environments. This moves toward highlighting the affordances of commonplace interaction methods, and how user locomotion behaviours are impacted by the freehand and gaze-based techniques explored in Study 2.

7.1.4 RQ4: User-Defined Distance and Movement

The final study, presented in Chapter 6, considers user-defined approaches to interaction, exploring how users choose to position themselves and the extent to which they employ locomotion when using different AR interaction techniques in a room-scale environment. The study reports on the processes taken by 40 participants to complete a series of fundamental selection tasks. Akin to Chapter 5, the study compares two freehand techniques (Airtap, Hover) and two gaze-based techniques (Eye, Head), but instead considers the methods followed by participants from a standing position to select different sized objects (5cm, 15cm, 25cm). The advantages and limitations of each technique are highlighted with respect to user-defined distance, position, locomotion approach and speed, selection time, task-load, user experience and preference. Again, the objective was to understand how techniques could be used interchangeably, focusing on how distance and movement could be referenced to provide adaptive interactions based on the state and approach of the user. A range of use cases involving locomotion are provided at the start of the chapter, with the study leading to recommendations for how user-defined distance and movement could be considered to adapt techniques.

This has addressed **RQ4.** — How do user-defined distances and locomotion approaches vary with different freehand and gaze techniques, and what influence does this

have on their appropriateness in room-scale AR environments?

The above research questions consider two dimensions of Proxemic Interaction: Distance and Movement, with a focus on understanding how users interact with virtual content. Collectively, these studies have begun to reveal the trade-offs among freehand and gaze-based interaction methods, underscoring the importance of spatial context in designing effective AR experiences.

7.2 Distance and Movement

Through addressing the four research questions, the thesis has reaffirmed that a key aspect of interacting in AR is how we interpret depth and engage with virtual content placed throughout the real world (Whitlock et al., 2018). In some use cases, such as those defined in Sections 4.2 and 5.2, users will be seated and unable to easily adjust their distance relative to real world content. However, AR use cases also often call for users to physically move around an interaction space, adapting their positioning and locomotion approaches relative to the environment (Diaz et al., 2017). This encompasses two dimensions of proxemic interaction, Distance and Movement, where techniques were found to have distinct advantages and disadvantages based on user-object distances and the locomotion approach of the user.

7.2.1 Distance

Distance is a key consideration in AR environments, defined as *the continuous or discrete measure of the position of an entity with respect to another entity.* Previous research has highlighted the influence of distance on interaction performance and usability (Whitlock et al., 2018; Kytö et al., 2018), which is supported by the results from the studies conducted.

For example, the findings from Chapter 4, which assesses the effectiveness of freehand and gaze-based selection techniques for near-field interaction in seated AR environments, show that the Press technique provided the best performance for direct interactions within the intimate proxemic zone. This aligns with everyday behaviour, where Press closely resembles familiar physical actions like pressing a button or pushing an object. This means interacting with objects within arm's reach often feels intuitive and demands minimal cognitive effort (Pham et al., 2018). In contrast, Chapters 5 and 6 indicate that Airtap was the least effective technique and not readily adopted by users, resulting in a steeper learning curve and lower perceived usability (Piumsomboon, Altimira, et al., 2014; Pourmemar and Poullis, 2019; Pfeuffer, B. Mayer, et al., 2017).

The distinction between direct and indirect interaction significantly influences how users engage with virtual content, especially in freehand interaction scenarios (Whitlock et al., 2018). While direct interaction involves manipulating virtual objects in ways that closely resemble real-world actions, indirect interaction requires intermediary or abstract gestures that lack direct physical counterparts (Lilligreen, Henkel, and Wiebel, 2022). Although Airtap is based on a metaphorical interaction paradigm that simulates clicking a mouse (Frutos-Pascual, Creed, and I. Williams, 2019), such gestures have been found to increase cognitive load and may not feel as intuitive to users (Whitlock et al., 2018; Kytö et al., 2018). This can lead to reduced performance and increased frustration, as observed in Chapters 5 and 6.

Instead, findings suggest that an improved approach would be for systems to provide interaction techniques interchangeably, tailoring inputs to the spatial context of users. Results indicate that Head and Eye could provide more effective input streams for targeting objects compared to Freehand techniques, offering improved performance and usability for fundamental selection tasks when their use is aligned with users' distance to content. Switching between interaction methods based on a user's distance from virtual objects, rather than using one method (i.e. freehand gesture) for all distances, could therefore improve usability and interaction flow.

Chapters 4 and 5 revealed that Eye was preferred for interactions in the Intimate (<0.5m) and Personal (0.5m - 1m) zones, whereas Head was more suitable for interactions in the Social (1m - 4m) and Public (>4m) zones. When considering the appropriateness of *Eye*, the average performance threshold was found near the boundary of the Personal and Social zones. Based on the minimum recommended target size of 2° on HoloLens 2, the objects explored in Chapters 4 and 5 would be 1.4m from users, just breaching the Social zone. As *Head* was affected least by distance, it would be interesting to explore at which point *Head* becomes significantly more efficient than *Eye*, and whether this is in line with the proposed threshold. These insights could then be used to inform how Eye and Head should be used interchangeably, to maximise performance and usability depending on the distance and visual angle of virtual content.

Based on Chapters 4 and 5, Figure 7.1 provides recommendations for providing the most appropriate interaction techniques across different proxemic zones, considering performance and preference. Notably, *Head* interaction was deemed to be the most suitable/reliable for selecting virtual content across multiple proxemic zones. Although *Press* provided the best performance in the Intimate zone, the baseline freehand technique, *Airtap*, was found to be less practical and required too much Physical Demand to be employed comfortably in the explored context. This underscores the crucial role of task context and spatial considerations in enhancing the practicality of AR technologies.

These differences across proxemic distances can be considered to improve world-scale AR environments (as depicted in Figure 1.1). For example, a user could be selecting virtual content in the Public zone, and then switch planes to select an item in the Personal proxemic zone. Here, we can consider the design of more adaptive and interchangeable AR interactions, which offer potential to optimise the accuracy and experience for each selection within each zone (i.e. *Head* and *Eye*), or apply one selection method (i.e. *Head*) for commonality, albeit with a potentially reduced performance and user experience (see Figure 7.1). This interchangeable concept, coupled with the knowledge of the interplay between the technique and user-object distance in AR, can empower developers to provide

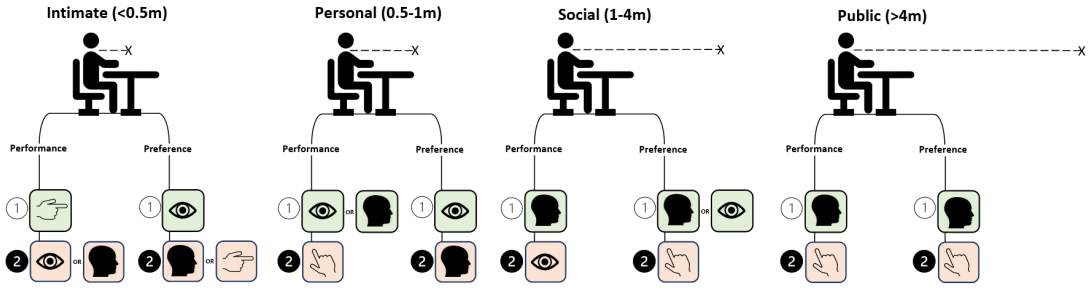

Figure 7.1: Appropriateness of interaction techniques in different proxemic zones: These are *Press*, *Eye*, *Head* or *Hover* depending on the distance.

improved interaction solutions to users.

Although the studies presented provide a reasonable grounding to explore adaptive techniques based on user-object distances, results affirm that to ensure the most effective experiences, distance thresholds would need to be accurately defined for individual users and contexts (Bhowmick, Kalita, and Sorathia, 2020; R. Li et al., 2019). Although Figure 7.1 provides recommendations based on results for performance and preference across proxemic zones, findings should also be considered relative to the technical limitations of current XR devices like the HoloLens 2. Factors such as sensor fidelity, tracking ranges, and field-of-view dictate what is achievable and therefore influence how interaction methods perform and would be recieved across different distances (W. Xu, Liang, Y. Chen, et al., 2020; Kytö et al., 2018). Redefining proxemic zones for AR interaction will therefore require more understanding of the relationship between depth and size of virtual objects in immersive interfaces across a range of users, devices and applications. This includes consideration for factors surrounding the movement of users within the interaction space.

7.2.2 Movement

World-anchored content remains fixed relative the physical environment, with users often free to adjust their position and achieve optimal viewing angles for content aquisition and interaction. This means that interface adaptions require strong consideration for how users choose to interact, capturing how they navigate AR environments and to what extent they employ locomotion (Bhowmick, Kalita, and Sorathia, 2020; Ballendat,

Marquardt, and Greenberg, 2010).

Movement can be defined as *the changes in an entity's distance and orientation over time*, and was explored in this work based on 3 key measures: 1) Locomotion Approaches, assessing whether users chose to walk or not; 2) Walking Trajectories, which illustrate how users navigated the interaction space over time; and 3) Locomotion Speeds, capturing how quickly users moved. The results presented in Chapter 6 demonstrate that by understanding and responding to these aspects of user movement, systems can leverage the strengths of different input techniques to provide more seamless and impactful interaction flows.

Results highlight that, in the context explored, users tended to tailor their interaction strategies to match the advantages and constraints of each technique. In line with results presented in Chapter 5, Head notably demonstrated high accuracy at farther distances and proved most effective when employed whilst stationary. This meant participants often opted to select the target from their starting position as opposed to to walking towards it.

In parallel, as Eye emerged as an effective technique to employ in motion, users increasingly took advantage of their ability to select objects whilst walking. This aligns with previous work, which shows Eye input to be particularly advantageous in mobile scenarios. Manakhov et al. (2024) found that users could efficiently interact with virtual objects without needing to pause movement, especially when targets were stable relative to the environment, along the axes perpendicular to the direction of locomotion, as is often the case with world-anchored content. Akin to results in Chapter 6, this results in a more integrated experience, where navigation and selection tasks can be performed concurrently.

With Eye, users are able to implicitly scan their surroundings and attend to points of interest, whilst at the same time explicitly triggering interactions as targets enter their field of view (Blattgerste, Renner, and Pfeiffer, 2018). The benefits of being able to

maintain motion during interaction without sacrificing performance has been appreciated by users, with such capabilities being shown to reduce the interruption of task flow in dynamic environments (F. Lu, Davari, and D. Bowman, 2021).

Although the visual angle (i.e. the apparent size of the target as seen by the user) was found to influence the appropriateness of Eye, Chapter 6 also suggests that for sufficiently large objects, many users could interact as effectively as Head, even at farther distances. This finding points towards object size having a greater impact on interaction efficiency than depth in room-scale environments. This is supported by Kapp et al. (2021), who revealed that bigger distances even resulted in better tracking fixation trueness. Despite this, discrepancies have been found to arise between the depth cues offered by virtual content and its perceived location relative to the real world, potentially resulting in vergence-accommodation conflicts and visual fatigue (Manakhov et al., 2024). Having large targets, with a larger visual angle, mitigates the demands of precise depth perception and reduces the need for subtle convergence adjustments. This means that, with larger objects, users can quickly employ gaze targeting, improving selection times and overall accuracy when engaged in linear locomotion (Kapp et al., 2021).

Consequently, results suggest that if virtual content is large enough and does not exceed the user's comfortable viewing distance, Eye interaction, based on both direction and/or depth information (F. Lu, Davari, and D. Bowman, 2021), could serve as a powerful method for engaging with world-anchored content, even whilst walking.

As freehand techniques, particularly Airtap, demanded greater precision, users felt they had to get closer to objects, often employing slow walking speeds and coming to a complete stop to interact effectively. Although freehand techniques remain valuable in many contexts, particularly for direct manipulation of near-field objects (Adam S. Williams and Francisco R. Ortega, 2020), more recent work suggests that they are less effective for far-field interactions (Danyluk et al., 2021; Kang, J.-h. Shin, and Ponto, 2020), with results from Chapter 6 also showing it to be difficult to employ when users are in motion.

F. Lu, Davari, and D. Bowman (2021) explored the interplay between locomotion and interaction, reporting that movement tends to increase physical and cognitive load and diminish overall task performance, with freehand techniques proving less accessible when walking. With freehand interaction, stable user posture is again critical for maintaining accuracy during selection tasks, meaning performance suffers whilst moving. This is further emphasised by users needing to divide their attention between navigating the environment with their head/eyes and performing explicit inputs with their hands (F. Lu, Pavanatto, and Doug A Bowman, 2023).

Research has begun to explore how continuous locomotion affects the appropriateness of interaction techniques (Kapp et al., 2021; Manakhov et al., 2024; F. Lu, Davari, and D. Bowman, 2021). Body-locked content, which remains anchored relative to the user, e.g. their torso or head (Q. Zhou et al., 2020), may also be commonplace in future AR interactions. Such interfaces provide improved access to virtual information, supporting interaction on the go. By maintaining a consistent reference frame that moves in tandem with the user, body-locked interfaces can support more streamlined, albeit less explorative, engagement with AR content (Frutos-Pascual, Gale, et al., 2021).

Since Eye has been shown to be the most effective technique when in motion, using it as the default pointing method in walking scenarios could better serve users and enable more efficient interaction flows in dynamic mobile contexts (F. Lu, Davari, and D. Bowman, 2021; Manakhov et al., 2024). While Press demonstrated the best performance for near-field interaction in Chapter 4, it was not considered to be the preferred technique. Future studies should therefore examine whether this performance advantage remains when users are in motion and how it compares to Eye. For instance, in contrast to previous findings, J. Sun and Liao (2024) recently reported no significant differences between hand gestures and eye gaze regarding task completion or cognitive load, even in more challenging interaction scenarios.

Although participants were found to adjust their approaches to accommodate the affordances of each technique in Chapter 6, it was clear that they were generally reluctant

to walk towards an object. This observation aligns with previous research, which found that users tend to engage with virtual objects from a greater distance than real-world objects (R. Li et al., 2019). Bhowmick, Kalita, and Sorathia (2020) also revealed that most users preferred not to physically approach content but instead use their abilities to interact with objects from a distance, for example, by employing methods to bring objects closer to them when using freehand interaction techniques. This reinforces that it is more appropriate to provide interaction techniques that align with users context, allowing for more straightforward input capabilities in real time, as opposed to expecting users to conform to the affordances of a single modality.

7.3 Orientation, Identity and Location

In addition to Distance and Movement, Proxemic Interaction also encompasses Orientation, Identity, and Location (as defined in Chapter 2). The following sections examine these additional dimensions, discussing how they can complement the contributions provided by the studies and findings presented. Prior research on Orientation, Identity, and Location is also reviewed to highlight their potential contributions to context-aware AR interactions.

7.3.1 Orientation

Although Orientation is a key consideration of pose and movement, this dimension of proxemic interaction was not considered in standalone througout the user studies. Orientation refers to *the relative positioning and facing direction of entities within the interaction space relative to other entities* (Marquardt, Diaz-Marino, et al., 2011), and could be a key factor in determining attention and engagement levels. For instance, a system can monitor changes in orientation to understand the exact angle between two entities and/or determine whether two entities are facing each other (Ballendat, Marquardt, and

Greenberg, 2010). Orientation therefore acts as a straightforward way to capture a user's focus and gauge what they are/are not intending to interact with (F. Lu, Davari, Lisle, et al., 2020; Pfeuffer, Abdrabou, et al., 2021), allowing for interactions in 3D environments to be appropriately adapted or constrained.

This benefit is partly supported by the studies conducted. For example, despite results in Chapter 4 revealing no significant differences, 62.5% of errors were produced with Hover, which was due to users accidentally positioning the cursor over an object during the observation stage (i.e. leaving their hand facing the object array in a resting pose whilst viewing the target object to their right). Hover also produced more errors in the Personal zone than any other technique, suggesting the midas touch problem was more likely when interacting with larger targets. This highlights the importance of interface layout and interaction context in defining the affordances of techniques and AR content (McGill, Kehoe, et al., 2020; Pfeuffer, Abdrabou, et al., 2021).

Although "always in front" interfaces (where content follows the orientation of the user) have been found impactful for certain interaction scenarios, such as within an AR typing study with freehand gesture (Frutos-Pascual, Gale, et al., 2021), in the "observe and interact" context explored, having interactive content move with the user would have cluttered the interface and occluded objects of interest. Similarly, if "always in front" content was provided during gaze-based interactions, it would likely become more difficult to avoid the Midas touch problem (Bhowmick, Kalita, and Sorathia, 2020) and errors would have been higher. This suggests that it could be more appropriate to adjust the affordances of interactive elements when the user is not focused on or attempting to interact with them (Pfeuffer, Abdrabou, et al., 2021).

Furthermore, Chapter 6 highlights the importance of considering Orientation during Movement. Here, some users were relatively comfortable walking with Head whilst selecting Medium and Large world-locked objects, results showing that a higher percentage of Active trials resulted in selections whilst in motion when compared to Airtap and Hover. This is not in line with previous research (F. Lu, Davari, and D. Bowman, 2021),

where users were unable to interact whilst walking with Head. However, this scenario involved a different task, where users were required to walk continuously at a consistent speed, and reference content positioned either side of them. This suggests that the angle of approach could have an impact on the users ability to perform selections whilst walking.

Orientation is not only important for understanding single-user attention but also serves as a strong social cue in multi-user scenarios (E. T. Hall, 1966; Ballendat, Marquardt, and Greenberg, 2010). In face-to-face interaction (whether virtual or in the real world), people often use orientation (e.g. torso positioning, head direction) to signal cues like availability, willingness to engage, or their desire to maintain privacy (Pfeuffer, Abdrabou, et al., 2021; Panda, Jane Nicholas, et al., 2023). This becomes critical in shared or collaborative XR environments, where the body orientation of others can help users and/or a system to gauge collective focus, turn-taking or determine who is leading a task (Spittle, Panda, et al., 2024; Panda, Tankelevitch, et al., 2024). This information could then be used by a system to make interface adaptations accordingly (Pfeuffer, Abdrabou, et al., 2021).

Although Orientation has been considered an effective way to gauge user attention, Eye and Head gaze do not usually align during the visual exploration of potential targets (Mathias N. Lystbæk et al., 2022b). When interacting within the real world, gaze shifts are achieved through a combination of eye, head and body movements: the eyes move seperately from the head, the head relative to the torso, and the torso relative to the world (Sidenmark and Gellersen, 2020). Hands could also be oriented towards objects without reorienting the torso, head or eyes as would be expected for natural interaction scenarios (Mathias N. Lystbæk et al., 2022b), such as for eyes-free applications (Q. Zhou et al., 2020). This again suggests that is important to consider context, which will have an impact on what body part (e.g. head, eye, hand, torso) should be referenced for capturing user orientation information. This can be determined by assigning descriptors to entities (in this case body parts of users) within the environment, and understanding

their distance, orientation and movement relative to other entities.

7.3.2 Identity

By assigning specific descriptors, entities are effectively transformed into identities, enriching the interaction space with additional contextual information. Identity is the dimension which *describes the entities within an interaction space* (Marquardt, Diaz-Marino, et al., 2011). This can be 1) broadly defined based on categories, such as real-world objects, real-world walls/surfaces, virtual objects, and people; 2) based on different instances of the same type of entity, such as person 1 and person 2, or; 3) based on the detailed characteristics for each element within an environment (Marquardt and Greenberg, 2012; Ballendat, Marquardt, and Greenberg, 2010). The latter describes more fine grained properties, for example, explaining a virtual object through factors like its dimensions, colour, shape, level of interactivity, or a user (i.e. John) by factors like their demographics, preferences, interaction history, and any permanent or situational impairments (Khan et al., 2022; Davari, Stover, et al., 2024). By cross-referencing the characteristics of entities within the space, as well as their spatial distribution, orientation and movement relative to eachother, a system could calculate the most appropriate technique to afford interactions in changing contexts (X. B. Liu et al., 2024).

For example, McGill, Williamson, et al. (2019) stress the importance of matching the semantics of virtual and physical elements in AR environments, suggesting that techniques should capitalise on the capabilities and affordances of instrumented, connected, interactive settings. This includes exploiting built-in physical features of the real world, such as a desk/table, car dashboard, or the backs of seats in a train/plane cabin when interacting on the go (Ng et al., 2021). Cheng, Gebhardt, and Holz (2023) also highlight the benefits of positioning AR interfaces relative to nearby features such as vertical walls, horizontal surfaces, or in line with physical objects and displays. They considered how this impacts the appropriateness of interaction techniques, with their system providing

direct touch (Press), pinching (Airtap), or remote cursor control interchangeably based on the interaction context. This exemplifies how AR technologies allow for a real-time understanding of the user's distance, movement and orientation over time, relative to physical/virtual environmental features (IDs), which could be harnessed to adapt interfaces and interaction techniques accordingly. This can be achieved by measuring factors such as the user's hand and head positions, the physical environment, and the type of interface elements (Cheng, Gebhardt, and Holz, 2023), all of which could be explained by metadata associated with individual IDs

When considering the results from Chapter 4, task load could have been reduced and user experience improved if interaction techniques and virtual content were adapted to the environment. This can be achieved by measuring factors such as the user's hand and head positions, the physical environment, and the type of interface elements (Cheng, Gebhardt, and Holz, 2023), all of which could be explained by metadata associated with individual IDs.

Overlaying content on a surface (ID) has also been shown to minimise issues experienced from the lack of depth perception with freehand techniques and provide passive feedback to reduce fatigue (Cheng, Gebhardt, and Holz, 2023). This demonstrates the power of not only understanding spatial factors such as distance, movement and orientation of people and objects within the interaction space, but also understanding what the affordances and properties of these entities are.

IDs could be defined based on information that is (1) specified, for example, where users manually provide data such as usernames, demographic details, and preferences; (2) sensed, which refers to information captured though the built-in hardware of a device to understand what exists in the environment (e.g. via computer vision technologies); and (3) extracted, referring to data that may have been specified or sensed in the past and is subsequently stored in private or public databases (Davari, Stover, et al., 2024). This could include metadata associated with physical/virtual objects, or interaction logs of individual users that may be retrieved and referenced to better personalise future

interactions.

7.3.3 Location

The final dimension of Proxemic Interaction, Location, is defined as *the physical context surrounding the entities that define a particular interaction space and its characteristics.* Whereas Identity provides information for individual entities, Location provides a description of the interaction environment itself. This encompasses the overarching attributes of an environment, providing a full picture of the fixed and semifixed features, people and objects within the space, which can be used to describe factors such as room size, layout, furniture position, and the social practices and context associated with the interaction (Marquardt and Greenberg, 2012; Marquardt, Diaz-Marino, et al., 2011). This dimension underscores the importance of the environment's scale, the arrangement of objects relative to eachother, and the nature of the space (e.g. private vs. public, formal vs. relaxed, indoors vs. outdoors) (Davari, Stover, et al., 2024).

Location information can be inferred not only through gps data (Singh et al., 2020) but also semantically through a collection of identities. For example, Khan et al. (2022) demonstrates how detailed characteristics of an environment can be referenced using vision-based technologies to infer a user's location. By analyzing spatial features, detecting objects and surfaces, and embedding semantic information into content within the environment, a system can classify the user's location based on the collective understanding of features and descriptors defined (Flotyński, 2020).

Location also accounts for the sensed objective measures of an environment, described by information such as the volume and brightness of an interaction space, which can be used to understand the practicality of different inputs and outputs in any given context (Grubert et al., 2017; X. B. Liu et al., 2024; Davari, Stover, et al., 2024). This understanding allows for a system to adapt interactions to suit user capabilities and preferences in a particular setting. For example, a system could consider the social acceptance

of interaction techniques, modifying what is employed for a formal public meeting room versus a casual private lounge area, based on an overview of the IDs within the interaction space (X. B. Liu et al., 2024).

Having a holistic understaning of what is present in the interaction space, and there associated IDs, helps a system infer the wider context surrounding the interaction scenario. By tailoring context targets (inputs, outputs and system configuration) based on what is understood through context sources (the stimuli that a system can sense and/or store on databases and then respond to) (Grubert et al., 2017), systems can offer highly personalised experiences that respond to user needs, abilities, preferences, and behaviours (X. B. Liu et al., 2024; Davari, Stover, et al., 2024). This discussion builds on previous work (see Chapter 2) to suggest that context sources can arguably be understood through the distance, orientation and movement of identities within a location.

7.4 Limitations

This thesis has focused on exploring how Proxemic Dimensions, notably Distance and Movement, impact the appropriateness of commonplace, peripheral-free techniques for fundamental selection tasks. Although the research has highlighted the potential of referencing proxemic factors for providing adaptive AR interactions, many promising research topics and directions, including those provided in the recommendations of Chapter 3, remain under-explored. The three studies conducted focus on single-user object selection, and therefore provide limited insights when compared to the broader research landscape for context-awareness in XR. Although the work presented provides reasonable grounding, results are limited by factors such as the users, techniques, technologies, task and environment considered. This section addresses the main limitations of the thesis and their implications.

7.4.1 Users

Participants were limited to 22-44 year-old adults with knowledge and interest in computing. As a result, results can not be generalised to wider user populations, including children, older adults, people with disabilities, or different cultures (Munsinger, White, and Quarles, 2019; Che Dalim et al., 2020). Future studies should therefore adopt more inclusive and varied sampling, as recommended in Chapter 3, to explore the appropriateness of techniques with a wider range of users, all who will have different backgrounds, preferences, mental models and abilities.

7.4.2 Techniques

Only unimodal freehand and gaze-based techniques (Press, Airtap, Hover, Head, Eye) were explored, with time-based techniques having the same fixed dwell-times across all three studies. These were based on device-specific recommendations, which again limits its generalisability. This is because a wide range of techniques and settings exist for interacting with virtual content in AR - all with their own advantages and limitations (Hertel et al., 2021; Spittle, Frutos-Pascual, et al., 2022). For example, although speech and hardware-based input were not considered, this does not mean they are not valuable for AR interactions (W. Xu, Liang, He, et al., 2019; Adam S. Williams, Garcia, and F. Ortega, 2020), and should continue to be explored. Further, in line with previous work (Z. Wang et al., 2020; Wolf et al., 2019), findings across studies suggest that multimodal input (e.g. one as a pointing method and a distinct selection mechanism) would likely improve the performance and usability of interactions. Results presented in the thesis therefore act as a solid starting point to explore this potential further.

7.4.3 Technology

User performance and experience was assessed using the Microsoft HoloLens 2 throughout all three studies. Although this device offers state-of-the-art input methods, aspects such as its field of view, weight, and optical see-through display type differ from those of more recent commercial pass-through headsets such as the Apple Vision Pro and Meta Quest 3. It is therefore expected that results would differ when using these HWDs (Frutos-Pascual, Creed, and I. Williams, 2019), and it would be valuable to conduct similar studies with a range of devices. This would help define to what extent a common set of interaction guidelines could be mapped and adopted for the broad spectrum of XR technologies, and help to further define what inputs could be harnessed when not all modalities are available (e.g. there is no eye tracking technology implemented in the Meta Quest 3).

7.4.4 Task

After considering a range of tasks in Chapter 3 (pointing, selection, translation, rotation, scale, viewport, menu-based and abstract) empirical studies focused on exploring fundamental selection within an "observe and interact" scenario. Although this type of interaction is often experienced in AR applications, e.g. to refer to one interface component to gather information (e.g. a text document/set of instructions, specification, video, or design concept), before interacting with another virtual component to perform input (e.g. selecting a virtual menu panel/button or an object to interact with) (Cheng, Gebhardt, and Holz, 2023; Lischke et al., 2016), results are unlikely to directly translate to the wide range of AR applications, tasks and use cases. Notably, selections were performed on cubes only, however different types of objects will likely impact the affordances of techniques, and the approaches taken by users (Piumsomboon, Clark, et al., 2013). Tasks were also completed during singlular and relatively short interaction sessions, without considering the impacts of extended interaction durations. The distances explored were also limited, focusing on only one distance within each proxemic zone in Chapters

4 and 5, with Chapter 6 only considering movement approaches from a standing position (where objects were in the public zone) without exploring the impact of continuous movement (F. Lu, Davari, and D. Bowman, 2021). This means that more research is needed to better understand the relationship between depth and size for an array of tasks, activities and virtual content.

7.4.5 Environment

All studies were conducted in the same controlled lab environment. This means that results are unaffected by factors such as changing lighting, noise, and different levels of distraction in the physcial world (Tung et al., 2015). In real-world scenarios, the interaction space would often be more cluttered, dynamic and involve safety-critical contexts (e.g. work places, home environments, public spaces and outdoor environments) which would likely impact performance and usability (X. B. Liu et al., 2024). By understanding how different variables related to social and environmental considerations impact interaction, we can design adaptive input techniques that are more appropriate for realistic AR use cases and conditions. As highlighted in Chapter 3, it is often difficult to control for such factors in scientific research, however, it is important to explore how these variables can be measured and how different input modalities can be leveraged to harmonise the usability and flow of interactions based on the users environment.

7.4.6 Adaptation Possibilities

The approach to exploring adaptation strategies has been centred around Proxemic Dimensions; notably Distance and Movement. However, the research does not empirically explore the practical implementation of input adaption and has not explicitly outlined the range of human, environmental and system factors needed to infer how such adaptations occur. It also fails to consider how alternative adaptation approaches, such as AI/LLM driven prediction (X. B. Liu et al., 2024; Davari, Stover, et al., 2024), intent-based interac-

tions (Pfeuffer, Abdrabou, et al., 2021), and input channel-availability assessment (X. B. Liu et al., 2024), could support and be integrated into XR context awareness - which, dependent on context, could be employed alongside, or instead of, referencing Proxemic Dimensions. Output was also consistent for each technique across the studies conducted, even though output has been shown as an essential factor for interaction performance and usability in different use cases (Davari, Stover, et al., 2024; Grubert et al., 2017). Therefore, understanding how to synergise input, output, and system adaptations to ensure they are practical, and are in line with user expectations and capabilities, remains a significant challenge.

7.4.7 Qualitative Insights

Lastly, studies conducted primarily focused on quantitative metrics (e.g. time, error, NASA-TLX, UEQ) over qualitative considerations. While post-study interviews were carried out to capture high-level feedback and support quantitative results, there was no exploration of other important factors for AR interactions, such as users' sense of presence, immersion, learnability, social acceptability or their pain-points and motivations (H. Bai, Sasikumar, et al., 2020; Flavián, Ibáñez-Sánchez, and Orús, 2018; Waldow et al., 2018; Tung et al., 2015; Whitlock et al., 2018). Therefore, richer qualitative and subjective assessments would be needed to complement the findings presented, revealing more nuanced information that the three studies failed to address.

7.5 Summary

This chapter has explored how the five dimensions of proxemic interaction (Distance, Movement, Orientation, Identity, and Location) can be leveraged to provide more adaptive, transferable, and context-aware AR experiences. Notably, recommedations for adapting freehand and gaze-based techniques are presented to help guide the facilita-

tion of more flexible AR experiences based on distance and movement. Despite the limitations that have been highlighted, the work conducted suggests that rather than relying on one-size-fits-all solutions, or arbitrarily adapting system inputs, outputs and configurations, user experiences could be enhanced by referencing Proxemic Dimensions.

Framing context through Proxemic Dimensions could provide a more straightforward approach for designing adaptive AR experiences personalised to individual users. Contextual cues, such as how far a user is from a virtual object, if/how they move, what they are focused on, who or what is present in their interaction space, preferences, and how the environment is arranged, can be interpreted by a system to tailor AR interactions in real time (Davari, Stover, et al., 2024). This, in turn, would make it possible to capitalise on the advantages and disadvantages of different interaction techniques (X. B. Liu et al., 2024), adapt to physical affordances of the environment and a users cognitive demands (Cheng, Gebhardt, and Holz, 2023; Lindlbauer, Feit, and Hilliges, 2019), and ultimately better align AR interactions with user activities and expectations across a range of applications and contexts (Grubert et al., 2017). The next chapter provides final conclusions and future work that builds on the research and concepts presented and discussed throughout the thesis.

Chapter Eight

Conclusions and Future Work

Contents

8.1 Findings and Contributions **195**

8.2 Future Work **197**

8.3 Closing Statement **201**

8.1 Findings and Contributions

After exploring opportunities and research gaps in Chapters 2 and 3, three studies were reported and discussed to establish how Distance and Movement could act as key sources of context, providing a basis for adapting commonplace interaction techniques for fundamental AR selection tasks. Recommendations for employing techniques based on Distance, Movement and performance/usability are reported, highlighting their appropriateness and affordances when users are 1) seated and 2) interacting with objects from a standing position and have the freedom to approach them.

Results of Study 1 (Impact of Technique on AR Interaction in Chapter 4) show that most users had a preference for gaze techniques over freehand in the "Observe and Interact" scenario explored. Despite Press providing the best performance, Eye was most often preferred. Head was deemed to require higher taskload in the intimate proxemic zone, with Hover having worst performance and being least preferred overall.

This highlights the importance of considering how to balance performance with user experience and provide the most appropriate techniques, not only based on distance (when comparing these results to Studies 2 and 3), but also factors such as the tasks being conducted, interaction duration, and a users primary motivations and goals (Spittle, Frutos-Pascual, et al., 2022).

Results of Study 2 (Impact of Distance on AR Interaction in Chapter 5) reveal that user distance from virtual content significantly impacted the suitability of interaction techniques during seated far-field selection tasks. While Head offered the most consistent performance and user preference across Social and Public zones, Eye was shown to be more appropriate for selections within the Personal zone. Freehand techniques, particularly Airtap, were found to be physically demanding and less practical. However, Hover performed better than Airtap and was preffered over Eye in the Public zone. These findings reinforce the importance of considering not just the input technique but also the spatial context, suggesting that switching between methods like Head and Eye based on proximity could help balance usability and performance, improving user experience across varying distances in seated environments.

Results of Study 3 (User-Defined Locomotion in Chapter 6) build on the previous findings by considering interaction in a room-scale environment, exploring how users choose to move and position themselves when selecting world-anchored virtual content. Head was again the most reliable technique for interacting with small, distant content, where participants primarily decided to remain stationary and leverage its strengths for distant selection. Despite Head affording users to interact from a distance, the technique also presented issues with maintaining cursor position to select objects when in motion. When walking, users were able to maintain or increase their speed during selections more easily with Eye, however, did not like feeling forced to walk towards content (especially small objects). Airtap again ranked lowest in terms of preference, physical demand, and performance, with Hover providing better overall usability for freehand interaction. This study highlights the need to consider not only user-object distance but also user

locomotion and intent when designing adaptive AR interfaces.

Building on findings from the three studies, the thesis extends the discussion to incorporate Orientation, Identity, and Location in Chapter 7, demonstrating how these Proxemic Dimensions (alongside Distance and Movement) could enhance the approaches taken to research, design and implement interactions in XR. By framing context through Proxemic Dimensions (Ballendat, Marquardt, and Greenberg, 2010; Greenberg, Marquardt, et al., 2011), the discussion aims to spark ideas around how a more cohesive and holistic approach to providing impactful adaptive XR experiences could be achieved.

These insights highlight the importance of providing appropriate interaction techniques based on spatial and semantic factors, paving the way for the development of more intuitive, AI-driven, and context-aware XR experiences (Davari, Stover, et al., 2024; X. B. Liu et al., 2024). As we move towards an era of human-computer collaboration, where intelligent systems will need to be capable of adapting to individual users, their preferences and a range of dynamic environments, integrating contextual understanding into XR interaction design becomes increasingly important (Grubert et al., 2017; Xiaoan Liu et al., 2025; X. B. Liu et al., 2024). To end the thesis and build on the research presented, key areas recommended for future work are provided below.

8.2 Future Work

Further consider Distance thresholds for prompting system adaptations Although adopting theories that are grounded in human-human interaction, such as Proxemics (E. T. Hall, 1966), is a reasonable starting point for exploring interaction design in XR, as research suggests users behave differently when interacting with virtual objects when compared to the real-world (Huang et al., 2022), a challenge is to define these differences and adjust the theories accordingly for XR interaction. Therefore, to ensure the most effective XR experiences, distance thresholds will need to be accurately redefined for

different users and entities. This will require building on the work presented to provide more understanding around the relationship between depth and size for a range of real and virtual people, objects and surfaces in different XR use cases (Ballendat, Marquardt, and Greenberg, 2010; Greenberg, Marquardt, et al., 2011).

Consider how to provide adaptive systems without overwhelming the user Despite the possibiltiies presented, designing systems that adapt techniques based on Proxemic Dimensions could inadvertently increase users' cognitive load. By switching between multiple inputs (e.g. Press, Eye and Head), users may soon become confused if adaptations are not clear, and it could become easy to lose track of which input method is currently active. Consequently, future research should consider the role of output and how to effectively communicate that adaptations have occurred. If the thresholds for switching techniques are poorly defined or vary significantly for different tasks, users may feel overwhelmed by having to mentally track these changes (Venkatakrishnan et al., 2023). This would likely hinder the usability and flow of interactions, requiring additional training and/or system cues to help users understand why and/or when different techniques are triggered (Davari, Stover, et al., 2024; X. B. Liu et al., 2024).

Explore the most appropriate feedback on the Interacting Layer for different techniques Building on the recommendation above, as well as acting as a cue that an adaptation has occurred, different forms of system feedback will impact user confidence and efficiency when performing explicit interactions. Research suggests that the value of feedback (such as a cursor) is highly dependent on factors like the task being conducted and interaction modalities employed (Frutos-Pascual, Gale, et al., 2021; Venkatakrishnan et al., 2023). This indicates a need for iterative testing and refinement of feedback mechanisms to ensure alignment with user expectations for employing different interaction paradigms in various contexts, something which was not considered in the scope of the thesis.

Consider how to balance unimodal and multimodal interactions Although this research focuses on unimodal selection techniques, separating pointing and selections between two modalities could simplify freehand interactions with distant content and improve efficiency and usability (i.e. as highlighted when comparing Airtap and Hover in Chapter 5). Leveraging multimodal interactions; such as combining gaze for pointing with voice, button presses, or gestures for selection, has also shown potential to simplify interaction workflows (F. Lu, Pavanatto, and Doug A Bowman, 2023). Consequently, future research should prioritise understanding how unimodal and multimodal interactions can be designed to ensure seamless integration and adaptability across varying use cases and contexts. This includes exploring how to balance simplicity and functionality, ensuring that interactions remain intuitive, easy to navigate, and do not overwhelm users (Adam S. Williams, Garcia, and F. Ortega, 2020).

Explore opportunities for harnessing Dimensions of Proxemic Interaction for context-aware XR As distance and movement have been shown to act as a key source of information for adapting interaction techniques, research should also consider the influence of the remaining Dimensions of Proxemic Interaction (Orientation, Identity and Location (Ballendat, Marquardt, and Greenberg, 2010; Greenberg, Marquardt, et al., 2011)). In AR, users are free to adjust their position and achieve optimal viewing angles for content acquisition and interaction (Bhowmick, Kalita, and Sorathia, 2020). Therefore, by sensing, specifying, or extracting data about a user's distance to virtual objects, movement patterns, focus of attention, the people or items within their interaction space, personal preferences, and the spatial arrangement of the environment(Davari, Stover, et al., 2024; Grubert et al., 2017), a system could interpret situational context and tailor AR interactions in real time(X. B. Liu et al., 2024). This capability would enable XR technologies to capitalise on the respective strengths and mitigate the weaknesses of different input techniques(Spittle, Frutos-Pascual, et al., 2022; Hertel et al., 2021), accommodate environmental affordances and changes in cognitive load(Cheng, Gebhardt, and Holz, 2023; Lindlbauer, Feit, and Hilliges, 2019), and ultimately better align interaction be-

haviours with users tasks and expectations across diverse interaction scenarios(Grubert et al., 2017; X. B. Liu et al., 2024; Davari, Stover, et al., 2024).

Explore how to communicate complex information for inferring and providing adaptations in XR Although assigning attributes and characteristics to the people, objects and features within the environment provides a way to customise interactions, capturing and defining complex identity descriptors (e.g. user interpretations, mental states, physical attributes, social roles, or task motivation) is challenging (X. B. Liu et al., 2024; Davari, Stover, et al., 2024). Making a system capable of understanding, storing, and updating a detailed semantic map of each environment can also demand significant computing resources (Strecker et al., 2023), where in dynamic or crowded spaces, the system could struggle to maintain an accurate model of the environment, especially if objects and features are similar or move unpredictably (Zhang et al., 2022; Flotyński, 2020). Consequently, while this research proposes that providing a system with a holistic understanding of an environment provides enhanced opportunities for enabling context-aware experiences, ensuring that this data is correct, current and ethical remains a key barrier (Mcgill, 2021).

Consider how to navigate security and privacy concerns for inferring context in XR Collecting and storing detailed personal and sensitive information can raise serious questions about user control over their data, who has access to it, and whether it is used in ways they find acceptable (Zahid Iqbal et al., 2023). Although capturing distance, orientation, and movement of IDs via methods like computer vision (Khan et al., 2022; Strecker et al., 2023) or biometric data (Bhalla et al., 2021) will often be crucial for providing adaptive systems, this may simultaneously pose serious privacy risks if information is mishandled (Mcgill, 2021; Zahid Iqbal et al., 2023).

Because of this, context-aware systems often rely on context controllers, which determine how context sources are monitored and to what extent users have autonomy

to decide how adaptations occur (Grubert et al., 2017). Despite this, Davari, Stover, et al. (2024) note that providing users with expansive control may result in a diminished experience due to the likeliness of data being incorrectly specified. Overly complex preferences and privacy settings can also introduce confusion, decision fatigue, or cause users to unintentionally share sensitive information (Strauss et al., 2024).

Accordingly, widespread XR adoption depends on thoughtful interface design, customisation, transparency, and user education about how and why system adaptations occur (Xuhai Xu et al., 2023; Strauss et al., 2024). XR creators should therefore explore how to strike a balance between providing adaptive systems (which autonomously react to user or environmental data) and adaptable systems (which afford users the freedom to control and modify system behaviour) and consider which party (e.g. users, establishments, systems) should have precedence in defining how adaptations occur in different contexts (Grubert et al., 2017).

8.3 Closing Statement

This thesis has demonstrated the potential and value of leveraging Proxemic Dimensions as contextual cues for informing adaptive interaction techniques in AR. By conducting three empirical studies, the work demonstrates how distance and movement significantly influence the performance, usability, and user preference of commonplace freehand and gaze-based input methods. This has produced practical recommendations for designing more flexible, intuitive, and context-aware systems across seated and room-scale AR interactions.

Beyond presenting findings around the role of Distance and Movement for explicit interaction, the thesis sparks broader conversations around harnessing Orientation, Identity, and Location, and how they could be referenced for future XR experiences. These dimensions hold substantial potential for developing systems that are not only adaptive,

but also continuously responsive to user activities, intentions, and environments. As XR technologies move towards real ubiquity, the insights and recommendations presented aim to support the development of interactions that are dynamic, inclusive, and in line with real-world behaviours.

Although the theoretical approach presented moves towards a potential framework for context-awareness in XR, there are many challenges; spanning technical, ethical, and design, that must first be addressed. Despite this, due to recent technical advances in AI and peripheral-free interaction, the opportunity to create intelligent systems that can infer and respond to a range of contextual factors is now within reach. By continuing to explore the potential of referencing Proxemic Dimensions in XR, and the semantics of virtual and real-world environments, researchers and developers can work towards providing enhanced personalisation and usability. This work has aimed to lay a foundation for that end goal, and invites the research community to build on the work presented to help deliver the next generation of user-centered, context-aware XR technologies.

References

Abdullah, Fadzidah, Mohd Faredzuan Mohd Noor, and Mohd Raziff Abd Razak (Oct. 2023). "Future Public Parks: Integrating Facilities for Locative Augmented Reality Games". In: *Journal of Advanced Research in Applied Sciences and Engineering Technology* 33, pp. 197–207. DOI: 10.37934/araset.33.1.197207.

Alallah, Fouad et al. (Nov. 2018). "Performer vs. observer". In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. DOI: 10.1145/3281505.3281541.

Aliprantis, John et al. (2019). "Natural Interaction in Augmented Reality Context". In: *VIPERCIRCDL*.

Andersson, Jonathan and Yan Hu (Nov. 2023). "Exploring the Impact of Menu Systems, Interaction Methods, and Sitting or Standing Posture on User Experience in Virtual Reality". In: *2023 IEEE Gaming, Entertainment, and Media Conference (GEM)*. DOI: 10.1109/gem59776.2023.10390210.

Ariansyah, Dedy et al. (Jan. 2022). "A head mounted augmented reality design practice for maintenance assembly: Toward meeting perceptual and cognitive needs of AR users". In: *Applied Ergonomics* 98, p. 103597. DOI: 10.1016/j.apergo.2021.103597.

Arora, Rahul et al. (Oct. 2019). "MagicalHands: Mid-Air Hand Gestures for Animating in VR". In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. DOI: 10.1145/3332165.3347942.

Augstein, Mirjam, Thomas Neumayr, and Sebastian Pimminger (2019). "WeldVUI: Establishing Speech-Based Interfaces in Industrial Applications". In: *Human-Computer*

Interaction – INTERACT 2019 11748 LNCS, pp. 679–698. DOI: 10.1007/978-3-030-29387-1_40.

Bach, Benjamin et al. (Jan. 2018). "The Hologram in My Hand: How Effective is Interactive Exploration of 3D Visualizations in Immersive Tangible Augmented Reality?" In: *IEEE Transactions on Visualization and Computer Graphics* 24, pp. 457–467. DOI: 10.1109/tvcg.2017.2745941.

Bai, Huidong, Gun A. Lee, et al. (Nov. 2014). "3D gesture interaction for handheld augmented reality". In: *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications on - SA '14*, pp. 1–6. DOI: 10.1145/2669062.2669073.

Bai, Huidong, Prasanth Sasikumar, et al. (Apr. 2020). "A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* CHI 2020, April 25–30, 2020, Honolulu, HI, USA. DOI: 10.1145/3313831.3376550.

Bai, Zhen and Alan F. Blackwell (Nov. 2012). "Analytic review of usability evaluation in ISMAR". In: *Interacting with Computers* 24, pp. 450–460. DOI: 10.1016/j.intcom.2012.07.004.

Bailly, Charles, François Leitner, and Laurence Nigay (2019). "Head-Controlled Menu in Mixed Reality with a HMD". In: *Human-Computer Interaction – INTERACT 2019*, pp. 395–415. DOI: 10.1007/978-3-030-29390-1_22.

Ballendat, Till, Nicolai Marquardt, and Saul Greenberg (2010). "Proxemic interaction: designing for a proximity and orientation-aware environment". In: *ACM International Conference on Interactive Tabletops and Surfaces - ITS '10*, pp. 121–130. DOI: 10.1145/1936652.1936676.

Bao, Yiwei et al. (Mar. 2023). "Exploring 3D Interaction with Gaze Guidance in Augmented Reality". In: *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. DOI: 10.1109/vr55154.2023.00018.

Bazzaza, Mhd Wael et al. (Dec. 2014). "iARBook: An Immersive Augmented Reality system for education". In: *2014 IEEE International Conference on Teaching, As-*

sessment and Learning for Engineering (TALE), pp. 495–498. DOI: 10.1109/tale.2014.7062576.

Becker, Vincent, Felix Rauchenstein, and Gábor Sörös (Mar. 2019). "Investigating Universal Appliance Control through Wearable Augmented Reality". In: *Proceedings of the 10th Augmented Human International Conference 2019*, pp. 1–9. DOI: 10.1145/3311823.3311853.

Belkacem, Ilyasse, Isabelle Pecci, and Benoıt Martin (May 2019). "Pointing task on smart glasses: Comparison of four interaction techniques". In: *arXiv:1905.05810 [cs]*.

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

Bhalla, Arman et al. (May 2021). "MoveAR: Continuous Biometric Authentication for Augmented Reality Headsets". In: *CPSS '21: Proceedings of the 7th ACM on Cyber-Physical System Security Workshop*. DOI: 10.1145/3457339.3457983.

Bhowmick, Shimmila, Pratul Kalita, and Keyur Sorathia (Nov. 2020). "A Gesture Elicitation Study for Selection of Nail Size Objects in a Dense and Occluded Dense HMD-VR." In: *IndiaHCI '20: Proceedings of the 11th Indian Conference on Human-Computer Interaction*, pp. 12–23. DOI: 10.1145/3429290.3429292.

Blaga, Andreea Dalia et al. (Apr. 2020). "Too Hot to Handle: An Evaluation of the Effect of Thermal Visual Representation on User Grasping Interaction in Virtual Reality". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. DOI: 10.1145/3313831.3376554.

— (Oct. 2021). "A Grasp on Reality: Understanding Grasping Patterns for Object Interaction in Real and Virtual Environments". In: *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* 28, pp. 391–396. DOI: 10.1109/ismar-adjunct54149.2021.00090.

Blattgerste, Jonas, Patrick Renner, and Thies Pfeiffer (June 2018). "Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying

field of views". In: *Proceedings of the Workshop on Communication by Gaze Interaction*. DOI: 10.1145/3206343.3206349.

Bothén, Simon, Jose Font, and Patrik Nilsson (Aug. 2018). "An analysis and comparative user study on interactions in mobile virtual reality games". In: *Proceedings of the 13th International Conference on the Foundations of Digital Games*, pp. 1–8. DOI: 10.1145/3235765.3235772.

Bowman, Doug A. and Larry F. Hodges (Feb. 1999). "Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments". In: *Journal of Visual Languages & Computing* 10, pp. 37–53. DOI: 10.1006/jvlc.1998.0111.

Brancati, Nadia et al. (Nov. 2018). "Experiencing touchless interaction with augmented content on wearable head-mounted displays in cultural heritage applications". In: *International Journal of Human–Computer Interaction* 35, pp. 203–217. DOI: 10.1007/s00779-016-0987-8.

Brooke, John (Nov. 1995). "SUS: A quick and dirty usability scale". In: *Usability Eval. Ind.* 189.

Buchta, Karolina et al. (Oct. 2022). "NUX Characters - interaction with voice assistants in Virtual Reality". In: *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 917–918. DOI: 10.1109/ismar-adjunct57072.2022.00204.

Caputo, Ariel, Riccardo Bartolomioli, and Andrea Giachetti (2023). "Remote and Deviceless Manipulation of Virtual Objects in Mixed Reality". In: *Smart Tools and Applications in Graphics - Eurographics Italian Chapter Conference*. Ed. by Francesco Banterle et al. The Eurographics Association. ISBN: 978-3-03868-235-6. DOI: 10.2312/stag.20231290.

Chan, Edwin et al. (May 2016). "User Elicitation on Single-hand Microgestures". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. DOI: 10.1145/2858036.2858589.

Che Dalim, Che Samihah et al. (Feb. 2020). "Using augmented reality with speech input for non-native children's language learning". In: *International Journal of Human-Computer Studies* 134, pp. 44–64. DOI: 10.1016/j.ijhcs.2019.10.002.

Cheema, Noshaba et al. (2020). "Predicting Mid-Air Interaction Movements and Fatigue Using Deep Reinforcement Learning". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. Honolulu, HI, USA: Association for Computing Machinery, pp. 1–13. ISBN: 9781450367080. DOI: 10.1145/3313831.3376701.

Chen, Di Laura, Ravin Balakrishnan, and Tovi Grossman (Mar. 2020). "Disambiguation Techniques for Freehand Object Manipulations in Virtual Reality". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 285–292. DOI: 10.1109/vr46266.2020.00048.

Chen, Timothy et al. (July 2023). "sPellorama: An Immersive Prototyping Tool using Generative Panorama and Voice-to-Prompts". In: *SIGGRAPH '23: ACM SIGGRAPH 2023 Posters*, pp. 1–2. DOI: 10.1145/3588028.3603667.

Chen, Zhaorui et al. (Oct. 2017). "Multimodal interaction in augmented reality". In: *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* Banff, AB, Canada,5-8 Oct. 2017, pp. 206–209. DOI: 10.1109/smc.2017.8122603.

Cheng, Yi Fei, Christoph Gebhardt, and Christian Holz (Oct. 2023). "InteractionAdapt: Interaction-driven Workspace Adaptation for Situated Virtual Reality Environments". In: *ACM Symposium on User Interface Software and Technology (UIST '23)*. DOI: 10.1145/3586183.3606717.

Chittaro, Luca and Riccardo Sioni (Nov. 2018). "Selecting Menu Items in Mobile Head-Mounted Displays: Effects of Selection Technique and Active Area". In: *International Journal of Human–Computer Interaction* 35, pp. 1501–1516. DOI: 10.1080/10447318.2018.1541546.

Cools, Robbe et al. (Oct. 2022). "Towards a Desktop-AR Prototyping Framework: Prototyping Cross-Reality Between Desktops and Augmented Reality". In: *2022 IEEE*

International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). DOI: 10.1109/ismar-adjunct57072.2022.00040.

Cottin, Tim et al. (Nov. 2016). "Gaze-Based Human-SmartHome-Interaction by Augmented Reality Controls". In: *Advances in intelligent systems and computing* 540, pp. 378–385. DOI: 10.1007/978-3-319-49058-8_41.

Dalim, Che Samihah Che et al. (Sept. 2016). "TeachAR: An Interactive Augmented Reality Tool for Teaching Basic English to Non-Native Children". In: *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 82–86. DOI: 10.1109/ismar-adjunct.2016.0046.

Danyluk, Kurtis et al. (May 2021). "A Design Space Exploration of Worlds in Miniature". In: *Libraries and Cultural Resources (University of Calgary)*. DOI: 10.1145/3411764.3445098.

Davari, Shakiba and Doug A Bowman (2024). *Towards Context-Aware Adaptation in Extended Reality: A Design Space for XR Interfaces and an Adaptive Placement Strategy*. arXiv.org. URL: https://arxiv.org/abs/2411.02607 (visited on 08/29/2025).

Davari, Shakiba, Daniel Stover, et al. (2024). *Towards Intelligent Augmented Reality (iAR): A Taxonomy of Context, an Architecture for iAR, and an Empirical Study*. arXiv.org.

Daza, Marcos et al. (Feb. 2021). "An Approach of Social Navigation Based on Proxemics for Crowded Environments of Humans and Robots". In: *Micromachines* 12, p. 193. DOI: 10.3390/mi12020193.

Dey, Arindam et al. (Sept. 2016). "A Systematic Review of Usability Studies in Augmented Reality between 2005 and 2014". In: *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 49–50. DOI: 10.1109/ismar-adjunct.2016.0036.

Diaz, Catherine et al. (Oct. 2017). "Designing for Depth Perceptions in Augmented Reality". In: *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. DOI: 10.1109/ismar.2017.28.

Divekar, Rahul R. et al. (2019). "You Talkin' to Me? A Practical Attention-Aware Embodied Agent". In: *Human-Computer Interaction – INTERACT 2019* 11748, pp. 760–780. DOI: 10.1007/978-3-030-29387-1_44.

Dong, Ze et al. (Apr. 2020). "A Comparison of Surface and Motion User-Defined Gestures for Mobile Augmented Reality". In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. DOI: 10.1145/3334480.3382883.

Dritsas, Elias et al. (Jan. 2025). "Multimodal Interaction, Interfaces, and Communication: A Survey". In: *Multimodal Technologies and Interaction* 9, pp. 6–6. DOI: 10.3390/mti9010006. URL: https://www.mdpi.com/2414-4088/9/1/6.

Esteves, Augusto, Yonghwan Shin, and Ian Oakley (July 2020). "Comparing selection mechanisms for gaze input techniques in head-mounted displays". In: *International Journal of Human-Computer Studies* 139, p. 102414. DOI: 10.1016/j.ijhcs.2020.102414.

Esteves, Augusto, David Verweij, et al. (Oct. 2017). "SmoothMoves: Smooth Pursuits Head Movements for Augmented Reality". In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. DOI: 10.1145/3126594.3126616.

Fadzli, Fazliaty Edora and Ajune Wanis Ismail (Dec. 2019). "VoxAR: 3D Modelling Editor Using Real Hands Gesture for Augmented Reality". In: *2019 IEEE 7th Conference on Systems, Process and Control (ICSPC)*, pp. 242–247. DOI: 10.1109/icspc47137.2019.9067992.

Farshidi, Siamak et al. (June 2024). "Understanding user intent modeling for conversational recommender systems: a systematic literature review". In: *User modeling and user-adapted interaction*. DOI: 10.1007/s11257-024-09398-x.

Flavián, Carlos, Sergio Ibáñez-Sánchez, and Carlos Orús (Nov. 2018). "The impact of virtual, augmented and mixed reality technologies on the customer experience". In: *Journal of Business Research* 100, pp. 547–560. DOI: 10.1016/j.jbusres.2018.10.050.

Flotyński, Jakub (Oct. 2020). "Creating explorable extended reality environments with semantic annotations". In: *Multimedia Tools and Applications*. DOI: 10.1007/s11042-020-09772-y.

Franco, Jéssica and Diogo Cabral (Nov. 2019). "Augmented object selection through smart glasses". In: *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*. DOI: 10.1145/3365610.3368416.

Frank, Jared A., Matthew Moorhead, and Vikram Kapila (Aug. 2016). *Realizing mixed-reality environments with tablets for intuitive human-robot collaboration for object manipulation tasks*. IEEE Xplore. DOI: 10.1109/ROMAN.2016.7745146.

Frutos-Pascual, Maite, Chris Creed, and Ian Williams (2019). "Head Mounted Display Interaction Evaluation: Manipulating Virtual Objects in Augmented Reality". In: *Human-Computer Interaction – INTERACT 2019* 11749, pp. 287–308. DOI: 10.1007/978-3-030-29390-1_16.

Frutos-Pascual, Maite, Clara Gale, et al. (Jan. 2021). "Character Input in Augmented Reality: An Evaluation of Keyboard Position and Interaction Visualisation for Head-Mounted Displays". In: *Lecture Notes in Computer Science*, pp. 480–501. DOI: 10.1007/978-3-030-85623-6_29.

Gagnon, Holly C. et al. (Apr. 2021). "Estimating Distances in Action Space in Augmented Reality". In: *ACM Transactions on Applied Perception* 18, pp. 1–16. DOI: 10.1145/3449067.

Gallardo, Andrea et al. (Oct. 2023). "Speculative Privacy Concerns about AR Glasses Data Collection". In: *Proceedings on Privacy Enhancing Technologies* 2023, pp. 416–435. DOI: 10.56553/popets-2023-0117.

Ganapathi, Priya and Keyur Sorathia (Sept. 2018). "Investigating controller less input methods for smartphone based virtual reality platforms". In: *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. DOI: 10.1145/3236112.3236136.

Genay, Adelaide, Anatole Lecuyer, and Martin Hachet (Dec. 2022). "Being an Avatar "for Real": A Survey on Virtual Embodiment in Augmented Reality". In: *IEEE*

Transactions on Visualization and Computer Graphics 28, pp. 5071–5090. DOI: 10.1109/tvcg.2021.3099290.

Ghaemi, Zeinab et al. (Jan. 2022). "Proxemic maps for immersive visualization". In: *Cartography and Geographic Information Science* 49, pp. 205–219. DOI: 10.1080/15230406.2021.2013946.

Ghosh, Debjyoti et al. (Apr. 2020). "EYEditor: Towards On-the-Go Heads-Up Text Editing Using Voice and Manual Input". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. DOI: 10.1145/3313831.3376173.

Goh, Eg Su, Mohd Shahrizal Sunar, and Ajune Wanis Ismail (2019). "3D Object Manipulation Techniques in Handheld Mobile Augmented Reality Interface: A Review". In: *IEEE Access* 7, pp. 40581–40601. DOI: 10.1109/access.2019.2906394.

Greenberg, Saul, Sebastian Boring, et al. (2014). "Dark patterns in proxemic interactions". In: *Proceedings of the 2014 conference on Designing interactive systems - DIS '14*. DOI: 10.1145/2598510.2598541.

Greenberg, Saul, Nicolai Marquardt, et al. (Jan. 2011). "Proxemic interactions". In: *interactions* 18, p. 42. DOI: 10.1145/1897239.1897250.

Grubert, Jens et al. (June 2017). "Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality". In: *IEEE Transactions on Visualization and Computer Graphics* 23, pp. 1706–1724. DOI: 10.1109/tvcg.2016.2543720.

Hall, Edward T (1966). *The hidden dimension*. Anchor Books.

Hardian, B., J. Indulska, and K. Henricksen (Mar. 2006). *Balancing autonomy and user control in context-aware systems - a survey*. IEEE Xplore. DOI: 10.1109/PERCOMW.2006.26.

Harrigan, Jinni A. (Mar. 2008). "Proxemics, Kinesics, and Gaze". In: *The New Handbook of Methods in Nonverbal Behavior Research*, pp. 136–198. DOI: 10.1093/acprof:oso/9780198529620.003.0004.

Harrison, Chris, Shilpa Ramamurthy, and Scott E. Hudson (Feb. 2012). "On-Body Interaction: Armed and Dangerous". In: *TEI '12: Proceedings of the Sixth Interna-*

tional Conference on Tangible, Embedded and Embodied Interaction, pp. 69–76. DOI: 10.1145/2148131.2148148.

Henderson, Jay, Jessy Ceha, and Edward Lank (Oct. 2020). "STAT: Subtle Typing Around the Thigh for Head-Mounted Displays". In: *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 1–11. DOI: 10.1145/3379503.3403549.

Henrikson, Rorik et al. (Apr. 2020). "Head-Coupled Kinematic Template Matching: A Prediction Model for Ray Pointing in VR". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1, 14. DOI: 10.1145/3313831.3376489.

Heo, Hwan et al. (Mar. 2020). "Gaze+Gesture Interface: Considering Social Acceptability". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. DOI: 10.1109/vrw50115.2020.00196.

Hertel, Julia et al. (Oct. 2021). "A Taxonomy of Interaction Techniques for Immersive Augmented Reality based on an Iterative Literature Review". In: *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. DOI: 10.1109/ismar52148.2021.00060.

Heydn, Katharina Anna Maria et al. (Sept. 2019). "The Golden Bullet: A Comparative Study for Target Acquisition, Pointing and Shooting". In: *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, pp. 1–8. DOI: 10.1109/vs-games.2019.8864589.

Hirschberg, Julia and Christopher D. Manning (July 2015). "Advances in natural language processing". In: *Science* 349, pp. 261–266. DOI: 10.1126/science.aaa8685.

Hochmair, Hartwig H., Levente Juhász, and Takoda Kemp (Aug. 2024). "Correctness Comparison of ChatGPT-4, Gemini, Claude-3, and Copilot for Spatial Tasks". In: *Transactions in GIS*. DOI: 10.1111/tgis.13233.

Hoffmann, Fabian et al. (Nov. 2019). "User-defined interaction for smart homes". In: *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*. DOI: 10.1145/3365610.3365624.

Hu, Jinghui, John J Dudley, and Per Ola Kristensson (Oct. 2022). "An Evaluation of Caret Navigation Methods for Text Editing in Augmented Reality". In: *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. DOI: 10.1109/ismar-adjunct57072.2022.00132.

Huang, Ann et al. (Apr. 2022). "Proxemics for Human-Agent Interaction in Augmented Reality". In: *CHI Conference on Human Factors in Computing Systems*, pp. 1–13. DOI: 10.1145/3491102.3517593.

Hussain, Muhammad, Jaehyun Park, and Hyun K Kim (Jan. 2023). "Effects of Interaction Method, Size, and Distance to Object on Augmented Reality Interfaces". In: *Interacting with computers* 35, pp. 1–11. DOI: 10.1093/iwc/iwad034.

Hwang, Alex D, Eli Peli, and Jae-Hyun Jung (Jan. 2023). "Development of virtual reality walking collision detection test on head-mounted display". In: *PubMed Central*, pp. 103–103. DOI: 10.1117/12.2647141. (Visited on 05/20/2025).

Iftikhar, Zainab et al. (2021). "Designing Parental Monitoring and Control Technology: A Systematic Review". In: *Human-Computer Interaction – INTERACT 2021*, pp. 676–700. DOI: 10.1007/978-3-030-85610-6_39.

Index, The NASA TLX Tool: Task Load (Dec. 2020). *TLX NASA Ames - Home*. Nasa.gov. URL: https://humansystems.arc.nasa.gov/groups/tlx/.

Jackson, Philip (2020). "Understanding understanding and ambiguity in natural language". In: *Procedia Computer Science* 169, pp. 209–225. DOI: 10.1016/j.procs.2020.02.138.

Jing, Allison, Gun Lee, and Mark Billinghurst (Mar. 2022). "Using Speech to Visualise Shared Gaze Cues in MR Remote Collaboration". In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. DOI: 10.1109/vr51125.2022.00044.

Kang, Hyo Jeong, Jung-hye Shin, and Kevin Ponto (Mar. 2020). *A Comparative Analysis of 3D User Interaction: How to Move Virtual Objects in Mixed Reality*. IEEE Xplore. DOI: 10.1109/VR46266.2020.00047.

Kapp, Sebastian et al. (Mar. 2021). "ARETT: Augmented Reality Eye Tracking Toolkit for Head Mounted Displays". In: *Sensors* 21, p. 2234. DOI: 10.3390/s21062234.

Khamis, Mohamed et al. (May 2018). "VRpursuits". In: *AVI '18: Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. DOI: 10.1145/3206505.3206522.

Khan, Dawar et al. (May 2022). "Recent advances in vision-based indoor navigation: A systematic literature review". In: *Computers Graphics* 104, pp. 24–45. DOI: 10.1016/j.cag.2022.03.005.

Kim, Kangsoo et al. (Nov. 2018). "Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008–2017)". In: *IEEE Transactions on Visualization and Computer Graphics* 24, pp. 2947–2962. DOI: 10.1109/tvcg.2018.2868591.

Kim, Minseok and Jae Yeol Lee (Feb. 2016). "Touch and hand gesture-based interactions for directly manipulating 3D virtual objects in mobile augmented reality". In: *Multimedia Tools and Applications* 75, pp. 16529–16550. DOI: 10.1007/s11042-016-3355-9.

Kim, Woojoo and Shuping Xiong (Nov. 2024). "TouchView: Mid-Air Touch on Zoomable 2D View for Distant Freehand Selection on a Virtual Reality User Interface". In: *Sensors* 24, pp. 7202–7202. DOI: 10.3390/s24227202. (Visited on 09/03/2025).

Koop, Mandy Miller et al. (Dec. 2020). "The HoloLens Augmented Reality System Provides Valid Measures of Gait Performance in Healthy Adults". In: *IEEE Transactions on Human-Machine Systems* 50, pp. 584–592. DOI: 10.1109/thms.2020.3016082.

Koutsabasis, Panayiotis and Chris K. Domouzis (June 2016). "Mid-Air Browsing and Selection in Image Collections". In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. DOI: 10.1145/2909132.2909248.

Koutsabasis, Panayiotis and Panagiotis Vogiatzidakis (Feb. 2019). "Empirical Research in Mid-Air Interaction: A Systematic Review". In: *International Journal of Human–Computer Interaction* 35, pp. 1747–1768. DOI: 10.1080/10447318.2019.1572352.

Krupke, Dennis et al. (Oct. 2018). "Comparison of Multimodal Heading and Pointing Gestures for Co-Located Mixed Reality Human-Robot Interaction". In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. DOI: 10.1109/iros.2018.8594043.

Kytö, Mikko et al. (Apr. 2018). "Pinpointing". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. DOI: 10.1145/3173574.3173655.

Lacoche, Jérémy et al. (Mar. 2014). "A survey of plasticity in 3D user interfaces". In: *7th Workshop on Software Engineering and Architectures for Realtime Interactive Systems, IEEE VR*. DOI: 10.1109/searis.2014.7152797.

Lages, Wallace S. and Doug A. Bowman (Mar. 2019). "Walking with adaptive augmented reality workspaces". In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*. DOI: 10.1145/3301275.3302278.

Lamberti, Fabrizio et al. (Feb. 2017). "Using Semantics to Automatically Generate Speech Interfaces for Wearable Virtual and Augmented Reality Applications". In: *IEEE Transactions on Human-Machine Systems* 47, pp. 152–164. DOI: 10.1109/thms.2016.2573830.

Laugwitz, Bettina, Theo Held, and Martin Schrepp (2008). "Construction and Evaluation of a User Experience Questionnaire". In: *Lecture Notes in Computer Science* 5298, pp. 63–76.

Laviola, Joseph J et al. (2017). *3D user interfaces : theory and practice*. Addison-Wesley.

Lee, Hock Siang et al. (May 2024). "Snap, Pursuit and Gain: Virtual Reality Viewport Control by Gaze". In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–14. DOI: 10.1145/3613904.3642838.

Lee, Jaewook et al. (Oct. 2023). "Towards Designing a Context-Aware Multimodal Voice Assistant for Pronoun Disambiguation: A Demonstration of GazePointAR". In: *UIST '23 Adjunct: Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–3. DOI: 10.1145/3586182.3615819.

Lee, Jihyeon, Jinwook Kim, and Jeongmi Lee (Oct. 2023). "Comparison of Virtual Reality Teleportation Targeting Method Performance depending on the Teleport Dis-

tance". In: *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 742–745. DOI: 10.1109/ismar-adjunct60411.2023.00160.

Lee, Minkyung et al. (Sept. 2013). "A usability study of multimodal input in an augmented reality environment". In: *Virtual Reality* 17, pp. 293–305. DOI: 10.1007/s10055-013-0230-0.

Li, Rui et al. (Mar. 2019). *Comparing Human-Robot Proxemics Between Virtual Reality and the Real World*. IEEE Xplore. DOI: 10.1109/HRI.2019.8673116.

Li, Xiangdong et al. (May 2022). *UOUU: User-Object and User-User Distance-Combined Method for Augmented Reality Collaborative Task*. IEEE Xplore. DOI: 10.1109/ICVR55215.2022.9848246.

Lilligreen, Gergana, Nico Henkel, and Alexander Wiebel (Jan. 2022). "Near and Far Interaction for Augmented Reality Tree Visualization Outdoors". In: *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. DOI: 10.5220/0010785700003124.

Lin, Tica et al. (Mar. 2023). "Labeling Out-of-View Objects in Immersive Analytics to Support Situated Visual Searching". In: *IEEE Transactions on Visualization and Computer Graphics* 29, pp. 1831–1844. DOI: 10.1109/TVCG.2021.3133511. URL: https://ieeexplore.ieee.org/abstract/document/9645242 (visited on 05/01/2023).

Lindlbauer, David, Anna Maria Feit, and Otmar Hilliges (Oct. 2019). "Context-Aware Online Adaptation of Mixed Reality Interfaces". In: *Repository for Publications and Research Data (ETH Zurich)*. DOI: 10.1145/3332165.3347945.

Lischke, Lars et al. (June 2016). "Screen arrangements and interaction areas for large display work places". In: *KOPS (University of Konstanz)*. DOI: 10.1145/2914920.2915027.

Liu, Chang, Alexander Plopski, and Jason Orlosky (June 2020). "OrthoGaze: Gaze-based three-dimensional object manipulation using orthogonal planes". In: *Computers & Graphics* 89, pp. 1–10. DOI: 10.1016/j.cag.2020.04.005.

Liu, Xiaoan et al. (2025). "Reality Proxy: Fluid Interactions with Real-World Objects in MR via Abstract Representations". In: *arXiv.org*. DOI: 10.1145/3746059.3747709. URL: https://arxiv.org/abs/2507.17248 (visited on 09/02/2025).

Liu, Xingyu Bruce et al. (May 2024). "Human I/O: Towards a Unified Approach to Detecting Situational Impairments". In: *arXiv (Cornell University)*. DOI: 10.1145/3613904.3642065.

Liu, Xinyi et al. (Dec. 2022). "Exploring Text Selection in Augmented Reality Systems". In: *Proceedings of the 18th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*. DOI: 10.1145/3574131.3574459.

Lo, Wei Hong, Stefanie Zollmann, and Holger Regenbrecht (Dec. 2021). "XRSpectator: Immersive, Augmented Sports Spectating". In: *VRST '21*. DOI: 10.1145/3489849.3489930.

Lu, Feiyu, Shakiba Davari, and Doug Bowman (Nov. 2021). "Exploration of Techniques for Rapid Activation of Glanceable Information in Head-Worn Augmented Reality". In: *Symposium on Spatial User Interaction*. DOI: 10.1145/3485279.3485286.

Lu, Feiyu, Shakiba Davari, Lee Lisle, et al. (Mar. 2020). "Glanceable AR: Evaluating Information Access Methods for Head-Worn Augmented Reality". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. DOI: 10.1109/vr46266.2020.00113.

Lu, Feiyu, Leonardo Pavanatto, and Doug A Bowman (Oct. 2023). "In-the-Wild Experiences with an Interactive Glanceable AR System for Everyday Use". In: *SUI '23: Proceedings of the 2023 ACM Symposium on Spatial User Interaction*. DOI: 10.1145/3607822.3614515.

Lu, Lu et al. (Mar. 2021). "Gaze-Pinch Menu: Performing Multiple Interactions Concurrently in Mixed Reality". In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. DOI: 10.1109/vrw52623.2021.00150.

Lu, Xueshi et al. (Mar. 2019). "DepthText: Leveraging Head Movements towards the Depth Dimension for Hands-free Text Entry in Mobile Virtual Reality Systems".

In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1060–1061. DOI: 10.1109/vr.2019.8797901.

Lystbæk, Mathias N et al. (May 2022a). "Exploring Gaze for Assisting Freehand Selection-based Text Entry in AR". In: *Proceedings of the ACM on human-computer interaction* 6, pp. 1–16. DOI: 10.1145/3530882.

Lystbæk, Mathias N. et al. (May 2022b). "Gaze-Hand Alignment: Combining Eye Gaze and Mid-Air Pointing for Interacting with Menus in Augmented Reality". In: *Proceedings of the ACM on Human-Computer Interaction* 6, pp. 1–18. DOI: 10.1145/3530886.

Manakhov, Pavel et al. (May 2024). "Gaze on the Go: Effect of Spatial Reference Frame on Visual Target Acquisition During Physical Locomotion in Extended Reality". In: *Scopus (Elsevier)*. DOI: 10.1145/3613904.3642915.

Manuri, Federico and Giovanni Piumatti (2015). "A Preliminary Study of a Hybrid User Interface for Augmented Reality Applications". In: *Proceedings of the 7th International Conference on Intelligent Technologies for Interactive Entertainment*, pp. 37–41. DOI: 10.4108/icst.intetain.2015.259629.

Marini, Marco Raoul et al. (June 2024). "A Natural Interaction System for Medical Training through VR Technology". In: *IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 237–242. DOI: 10.1109/cbms61543.2024.00047.

Marquardt, Nicolai, Robert Diaz-Marino, et al. (2011). "The proximity toolkit". In: *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*. DOI: 10.1145/2047196.2047238.

Marquardt, Nicolai and Saul Greenberg (Feb. 2012). "Informing the Design of Proxemic Interactions". In: *IEEE Pervasive Computing* 11, pp. 14–23. DOI: 10.1109/mprv.2012.15.

— (Feb. 2015). "Proxemic Interactions: From Theory to Practice". In: *Synthesis Lectures on Human-Centered Informatics* 8, pp. 1–199. DOI: 10.2200/s00619ed1v01y201502hci025.

Marques, Bernardo et al. (Nov. 2020). "Interaction with Virtual Content using Augmented Reality". In: *Proceedings of the ACM on Human-Computer Interaction* 4, pp. 1–17. DOI: 10.1145/3427324.

Mayer, Sven, Gierad Laput, and Chris Harrison (Apr. 2020). "Enhancing Mobile Voice Assistants with WorldGaze". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. DOI: 10.1145/3313831.3376479.

Mcgill, Mark (2021). *White Paper - The IEEE Global Initiative on Ethics of Extended Reality (XR) Report–Extended Reality (XR) and the Erosion of Anonymity and Privacy*. Ieee.org.

McGill, Mark, Aidan Kehoe, et al. (May 2020). "Expanding the Bounds of Seated Virtual Workspaces". In: *ACM Transactions on Computer-Human Interaction* 27, pp. 1–40. DOI: 10.1145/3380959.

McGill, Mark, Julie Williamson, et al. (Dec. 2019). "Challenges in passenger use of mixed reality headsets in cars and other transportation". In: *Virtual Reality*. DOI: 10.1007/s10055-019-00420-x.

Medeiros, Daniel, Rafael dos Anjos, et al. (Mar. 2021). "Promoting Reality Awareness in Virtual Reality through Proxemics". In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. DOI: 10.1109/vr50410.2021.00022.

Medeiros, Daniel, Mark McGill, et al. (Nov. 2022). "From Shielding to Avoidance: Passenger Augmented Reality and the Layout of Virtual Displays for Productivity in Shared Transit". In: *IEEE Transactions on Visualization and Computer Graphics* 28, pp. 3640–3650. DOI: 10.1109/TVCG.2022.3203002.

Medeiros, Daniel, Maurício Sousa, et al. (Nov. 2016). "Perceiving depth: Optical Versus Video See-through". In: *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. DOI: 10.1145/2993369.2993388.

Meng, Xuanru, Wenge Xu, and Hai-Ning Liang (Oct. 2022). "An Exploration of Hands-free Text Selection for Virtual Reality Head-Mounted Displays". In: *BCU Open Access Repository (Birmingham City University)*. DOI: 10.1109/ismar55827.2022.00021.

Meta (Feb. 2023). *Use Your Fingers (Not Controllers) to Swipe Through the VR Interface on Meta Quest*. Meta Newsroom. URL: https://about.fb.com/news/2023/02/meta-quest-direct-touch-use-your-fingers-in-vr/?utm_source=chatgpt.com (visited on 09/03/2025).

Microsoft (Oct. 2021a). *Comfort - Mixed Reality*. learn.microsoft.com. URL: https://learn.microsoft.com/en-us/windows/mixed-reality/design/comfort (visited on 04/16/2025).

— (Nov. 2021b). *HoloLens 2 gestures for authoring/navigating in Dynamics 365 Guides - Dynamics 365 Mixed Reality*. learn.microsoft.com. URL: https://learn.microsoft.com/en-us/dynamics365/mixed-reality/guides/authoring-gestures-hl2.

— (Mar. 2022a). *HoloLens environment considerations*. learn.microsoft.com. URL: https://learn.microsoft.com/en-us/hololens/hololens-environment-considerations (visited on 04/16/2025).

— (Dec. 2022b). *MRTK2-Unity Developer Documentation - MRTK 2*. learn.microsoft.com. URL: https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2 (visited on 04/16/2025).

— (Mar. 2023a). *Eye-gaze and dwell - Mixed Reality*. learn.microsoft.com. URL: https://learn.microsoft.com/en-us/windows/mixed-reality/design/gaze-and-dwell-eyes (visited on 04/16/2025).

— (Mar. 2023b). *Eye-gaze-based interaction - Mixed Reality*. learn.microsoft.com. URL: https://learn.microsoft.com/en-us/windows/mixed-reality/design/eye-gaze-interaction (visited on 04/16/2025).

— (Mar. 2023c). *Interactable object - Mixed Reality*. learn.microsoft.com. URL: https://learn.microsoft.com/en-us/windows/mixed-reality/design/interactable-object (visited on 04/16/2025).

Mifsud, Domenick et al. (Mar. 2022). "Augmented Reality Fitts' Law Input Comparison Between Touchpad, Pointing Gesture, and Raycast". In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. DOI: 10.1109/vrw55335.2022.00146.

Mirbagheri, Mahya and Tom Chau (Apr. 2024). "Optimising virtual object position for efficient eye-gaze interaction in Hololens2". In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging Visualization* 12. DOI: 10.1080/21681163.2024.2337765. (Visited on 09/03/2025).

Mohamed, Aya Khaled Youssef Sayed et al. (Aug. 2022). "A systematic literature review for authorization and access control: definitions, strategies and models". In: *International Journal of Web Information Systems* 18. DOI: 10.1108/ijwis-04-2022-0077.

Mohan, Pallavi, Wooi Boon Goh, et al. (Nov. 2019). "Head-Fingers-Arms: Physically-Coupled and Decoupled Multimodal Interaction Designs in Mobile VR". In: *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 1–9. DOI: 10.1145/3359997.3365697.

Mohan, Pallavi, Wooi Boon Goh, et al. (Oct. 2018). "DualGaze: Addressing the Midas Touch Problem in Gaze Mediated VR Interaction". In: *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. DOI: 10.1109/ismar-adjunct.2018.00039.

Mohd, Ekram Alhafis Hashim et al. (July 2024). "Revolutionizing Virtual Reality with Generative AI: An In-Depth Review". In: *Journal of Advanced Research in Computing and Applications* 30, pp. 19–30. DOI: https://doi.org/10.37934/arca.30.1.1930.

Monteiro, Pedro et al. (Jan. 2023). "Evaluation of Hands-Free VR Interaction Methods During a Fitts' Task: Efficiency and Effectiveness". In: *IEEE Access* 11, pp. 70898–70911. DOI: 10.1109/access.2023.3293057.

Morotti, Elena et al. (Nov. 2021). "Exploiting fashion x-commerce through the empowerment of voice in the fashion virtual reality arena". In: *Virtual Reality*. DOI: 10.1007/s10055-021-00602-6.

Morris, Meredith Ringel et al. (May 2014). "Reducing legacy bias in gesture elicitation studies". In: *interactions* 21, pp. 40–45. DOI: 10.1145/2591689.

Mossel, Annette, Benjamin Venditti, and Hannes Kaufmann (Mar. 2013). "3DTouch and HOMER-S". In: *Proceedings of the Virtual Reality International Conference: Laval Virtual*, pp. 1–10. DOI: 10.1145/2466816.2466829.

Mott, Martez et al. (Oct. 2019). "Accessible by Design: An Opportunity for Virtual Reality". In: *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 451–454. DOI: 10.1109/ismar-adjunct.2019.00122.

Muhammad Nizam, Siti Soleha et al. (Sept. 2018). "A Review of Multimodal Interaction Technique in Augmented Reality Environment". In: *International Journal on Advanced Science, Engineering and Information Technology* 8, p. 1460. DOI: 10.18517/ijaseit.8.4-2.6824.

Müller, Florian et al. (Apr. 2020). "Walk The Line: Leveraging Lateral Shifts of the Walking Path as an Input Modality for Head-Mounted Displays". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–15. DOI: 10.1145/3313831.3376852.

Munsinger, Brita, Greg White, and John Quarles (Sept. 2019). *The Usability of the Microsoft HoloLens for an Augmented Reality Game to Teach Elementary School Children*. IEEE Xplore. DOI: 10.1109/VS-Games.2019.8864548.

Munteanu, Cosmin et al. (2016). "Designing Speech and Multimodal Interactions for Mobile, Wearable, and Pervasive Applications". In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16* Association for Computing Machinery, New York, NY, USA, pp. 3612–3619. DOI: 10.1145/2851581.2856506.

Mutasim, Aunnoy K, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger (May 2021). "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*. DOI: 10.1145/3448018.3457998.

Nazri, Nur Intan Adhani Muhamad and Dayang Rohaya Awang Rambli (Dec. 2015). "The roles of input and output modalities on user interaction in mobile augmented

reality application". In: *Proceedings of the Asia Pacific HCI and UX Design Symposium*, pp. 46–49. DOI: 10.1145/2846439.2846449.

Ng, Alexander et al. (Oct. 2021). "The Passenger Experience of Mixed Reality Virtual Display Layouts in Airplane Environments". In: *Enlighten: Publications (The University of Glasgow)*. DOI: 10.1109/ismar52148.2021.00042.

Nijholt, Anton (Sept. 2021). "Experiencing Social Augmented Reality in Public Spaces". In: *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, pp. 570–574. DOI: 10.1145/3460418.3480157.

Nijholt, Anton, Job Zwiers, and Jan Peciva (Apr. 2007). "Mixed reality participants in smart meeting rooms and smart home environments". In: *Personal and Ubiquitous Computing* 13, pp. 85–94. DOI: 10.1007/s00779-007-0168-x.

Norouzi, N. et al. (Oct. 2019). *Walking Your Virtual Dog: Analysis of Awareness and Proxemics with Simulated Support Animals in Augmented Reality*. IEEE Xplore. DOI: 10.1109/ISMAR.2019.000-8.

Novick, David G and Aaron E Rodriguez (Jan. 2021). "A Comparative Study of Conversational Proxemics for Virtual Agents". In: *Lecture notes in computer science*, pp. 96–105. DOI: 10.1007/978-3-030-77599-5_8.

Obaid, Mohammad et al. (2012). "Cultural Behaviors of Virtual Agents in an Augmented Reality Environment". In: *Intelligent Virtual Agents* 7502, pp. 412–418. DOI: 10.1007/978-3-642-33197-8_42.

Özacar, Kasım et al. (Dec. 2016). "3D Selection Techniques for Mobile Augmented Reality Head-Mounted Displays". In: *Interacting with Computers*. DOI: 10.1093/iwc/iww035.

Palomino-Roldan, Geovanny, Roberto Rojas-Cessa, and Ernesto Suaste-Gomez (2023). "Eye Movements and Vestibulo-Ocular Reflex as User Response in Virtual Reality". In: *IEEE Access*, pp. 1–1. DOI: 10.1109/access.2023.3264637.

Panda, Payod, Molly Jane Nicholas, et al. (July 2023). "Beyond Audio: Towards a Design Space of Headphones as a Site for Interaction and Sensing". In: DOI: 10.1145/3563657.3596022.

Panda, Payod, Lev Tankelevitch, et al. (Nov. 2024). "Hybridge: Bridging Spatiality for Inclusive and Equitable Hybrid Meetings". In: *Proceedings of the ACM on Human-Computer Interaction* 8, pp. 1–39. DOI: 10.1145/3687040.

Pathmanathan, Nelusa et al. (June 2020). "Eye vs. Head: Comparing Gaze Methods for Interaction in Augmented Reality". In: *ACM Symposium on Eye Tracking Research and Applications*. DOI: 10.1145/3379156.3391829.

Pedersen, Isabel et al. (Apr. 2017). "More than Meets the Eye". In: *Journal on Computing and Cultural Heritage* 10, pp. 1–15. DOI: 10.1145/3051480.

Perea, Patrick, Denis Morand, and Laurence Nigay (Sept. 2020). "Target Expansion in Context: the Case of Menu in Handheld Augmented Reality". In: *Proceedings of the International Conference on Advanced Visual Interfaces*, pp. 1–9. DOI: 10.1145/3399715.3399851.

Pereira, Andre et al. (Aug. 2017). "Augmented reality dialog interface for multimodal teleoperation". In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. DOI: 10.1109/roman.2017.8172389.

Pfeuffer, Ken, Yasmeen Abdrabou, et al. (Jan. 2021). "ARtention: A design space for gaze-adaptive user interfaces in augmented reality". In: *Computers & Graphics*. DOI: 10.1016/j.cag.2021.01.001.

Pfeuffer, Ken, Benedikt Mayer, et al. (Oct. 2017). "Gaze + pinch interaction in virtual reality". In: *Proceedings of the 5th Symposium on Spatial User Interaction*. DOI: 10.1145/3131277.3132180.

Pfeuffer, Ken, Lukas Mecke, et al. (Nov. 2020). "Empirical Evaluation of Gaze-enhanced Menus in Virtual Reality". In: *26th ACM Symposium on Virtual Reality Software and Technology*. DOI: 10.1145/3385956.3418962.

Pham, Tran et al. (2018). “Scale Impacts Elicited Gestures for Manipulating Holograms”. In: *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS ’18*. DOI: 10.1145/3196709.3196719.

Piening, Robin et al. (2021). “Looking for Info: Evaluation of Gaze Based Information Retrieval in Augmented Reality”. In: *Human-Computer Interaction – INTERACT 2021*, pp. 544–565. DOI: 10.1007/978-3-030-85623-6_32.

Pike, Shane (Dec. 2023). “THEATRE AND TECHNOLOGY”. In: *Arte da Cena* 9, pp. 195–208. DOI: 10.5216/ac.v9i1.78046.

Piumsomboon, Thammathip, David Altimira, et al. (Sept. 2014). “Grasp-Shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality”. In: *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. DOI: 10.1109/ismar.2014.6948411.

Piumsomboon, Thammathip, Adrian Clark, et al. (Apr. 2013). “User-Defined Gestures for Augmented Reality”. In: *Human-Computer Interaction – INTERACT 2013* 8118, pp. 282–299. DOI: 10.1007/978-3-642-40480-1_18.

Piumsomboon, Thammathip, Gun Lee, et al. (2017). “Exploring natural eye-gaze-based interaction for immersive virtual reality”. In: *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. DOI: 10.1109/3dui.2017.7893315.

Plasson, Carole et al. (Nov. 2019). “Tabletop AR with HMD and Tablet”. In: *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, pp. 409–414. DOI: 10.1145/3343055.3360760.

— (Sept. 2020). “3D Tabletop AR”. In: *Proceedings of the International Conference on Advanced Visual Interfaces*. DOI: 10.1145/3399715.3399836.

Ponto, Kevin et al. (Apr. 2013). “Perceptual Calibration for Immersive Display Environments”. In: *IEEE Transactions on Visualization and Computer Graphics* 19, pp. 691–700. DOI: 10.1109/tvcg.2013.36. (Visited on 04/21/2020).

Pourmemar, Majid and Charalambos Poullis (Nov. 2019). “Visualizing and Interacting with Hierarchical Menus in Immersive Augmented Reality”. In: *The 17th Interna-*

tional Conference on Virtual-Reality Continuum and its Applications in Industry, pp. 1–9. DOI: 10.1145/3359997.3365693.

Prilla, Michael, Marc Janßen, and Timo Kunzendorff (Sept. 2019). "How to Interact with Augmented Reality Head Mounted Devices in Care Work? A Study Comparing Handheld Touch (Hands-on) and Gesture (Hands-free) Interaction". In: *AIS Transactions on Human-Computer Interaction* 11, pp. 157–178. DOI: 10.17705/1thci.00118.

Pringle, Andrew et al. (June 2019). "Ethnographic study of a commercially available augmented reality HMD app for industry work instruction". In: *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, pp. 389–397. DOI: 10.1145/3316782.3322752.

Puig, J. et al. (July 2012). *Towards an efficient methodology for evaluation of quality of experience in Augmented Reality*. IEEE Xplore. DOI: 10.1109/QoMEX.2012.6263864.

Qi, Feng and Wenchuan Wu (June 2019). "Human-like machine thinking: Language guided imagination". In: *arXiv:1905.07562 [cs, q-bio]*.

Qian, Jing et al. (June 2020). "Modality and Depth in Touchless Smartphone Augmented Reality Interactions". In: *ACM International Conference on Interactive Media Experiences*, pp. 74–81. DOI: 10.1145/3391614.3393648.

Qian, Yuan Yuan and Robert J. Teather (Oct. 2017). "The Eyes Don't Have It: An Empirical Comparison of Head-Based and Eye-Based Selection in Virtual Reality". In: *Proceedings of the 5th Symposium on Spatial User Interaction*. DOI: 10.1145/3131277.3132182.

Raeburn, Gideon and Laurissa Tokarchuk (Oct. 2021). "Varying user agency and interaction opportunities in a home mobile augmented virtuality story". In: *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 347–356. DOI: 10.1109/ismar52148.2021.00051.

Rao, N. et al. (Mar. 2020). *Investigating the Necessity of Meaningful Context Anchoring in AR Smart Glasses Interaction for Everyday Learning*. IEEE Xplore. DOI: 10.1109/VRW50115.2020.00091.

Regenbrecht, Holger et al. (Jan. 2024). "To See and Be Seen–Perceived Ethics and Acceptability of Pervasive Augmented Reality". In: *IEEE access*, pp. 1–1. DOI: 10.1109/access.2024.3366228.

Rutten, Isa and David Geerts (2020). "Better Because It's New: The Impact of Perceived Novelty on the Added Value of Mid-Air Haptic Feedback". In: CHI '20. Honolulu, HI, USA: Association for Computing Machinery, pp. 1–13. ISBN: 9781450367080. DOI: 10.1145/3313831.3376668.

Sadri, Shirin et al. (Oct. 2019). "Manipulating 3D Anatomic Models in Augmented Reality: Comparing a Hands-Free Approach and a Manual Approach". In: *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 93–102. DOI: 10.1109/ismar.2019.00-21.

Samini, Ali and Karljohan Lundin Palmerius (Nov. 2016). "A study on improving close and distant device movement pose manipulation for hand-held augmented reality". In: *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pp. 121–128. DOI: 10.1145/2993369.2993380.

Sanz, Ferran Argelaguet et al. (Mar. 2015). *Virtual proxemics: Locomotion in the presence of obstacles in large immersive projection environments*. IEEE Xplore. DOI: 10.1109/VR.2015.7223327.

Satriadi, Kadek Ananta et al. (Mar. 2019). "Augmented Reality Map Navigation with Freehand Gestures". In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 593–603. DOI: 10.1109/vr.2019.8798340.

Sawan, Nedal et al. (Dec. 2020). "Mixed and Augmented Reality Applications in the Sport Industry". In: *2020 2nd International Conference on E-Business and E-commerce Engineering*. DOI: 10.1145/3446922.3446932.

Scholl, Catelyn and Susan McRoy (Feb. 2019). “Using Gestures to Resolve Lexical Ambiguity in Storytelling with Humanoid Robots”. In: *Dialogue Discourse* 10, pp. 20–33. DOI: 10.5087/dad.2019.102.

Schoonenboom, Judith and R. Burke Johnson (July 2017). “How to Construct a Mixed Methods Research Design”. In: *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 69, pp. 107–131. DOI: 10.1007/s11577-017-0454-1.

Schramm, Robin Connor et al. (Sept. 2023). “Assessing Augmented Reality Selection Techniques for Passengers in Moving Vehicles: A Real-World User Study”. In: *AutomotiveUI '23: Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. DOI: 10.1145/3580585.3607152.

Schrepp, Martin (Sept. 2015). *User Experience Questionnaire Handbook*. DOI: 10.13140/RG.2.1.2815.0245.

Schuchhardt, Matthew et al. (Oct. 2015). “Optimizing mobile display brightness by leveraging human visual perception”. In: *2015 International Conference on Compilers, Architecture and Synthesis for Embedded Systems (CASES)*, pp. 11–20. DOI: 10.1109/cases.2015.7324538.

Seeliger, Arne, Raphael P. Weibel, and Stefan Feuerriegel (Sept. 2022). “Context-Adaptive Visual Cues for Safe Navigation in Augmented Reality Using Machine Learning”. In: *International Journal of Human–Computer Interaction*, pp. 1–21. DOI: 10.1080/10447318.2022.2122114. (Visited on 08/29/2025).

Sharma, Adwait et al. (June 2024). “GraspUI: Seamlessly Integrating Object-Centric Gestures within the Seven Phases of Grasping”. In: *Designing Interactive Systems Conference*, pp. 1275–1289. DOI: 10.1145/3643834.3661551.

Siddhpuria, Shaishav et al. (June 2017). “Exploring At-Your-Side Gestural Interaction for Ubiquitous Environments”. In: *Proceedings of the 2017 Conference on Designing Interactive Systems*. DOI: 10.1145/3064663.3064695.

Sidenmark, Ludwig, Christopher Clarke, et al. (Apr. 2020). “Outline Pursuits: Gaze-assisted Selection of Occluded Objects in Virtual Reality”. In: *Proceedings of the*

2020 CHI Conference on Human Factors in Computing Systems. DOI: 10.1145/3313831.3376438.

Sidenmark, Ludwig and Hans Gellersen (Jan. 2020). "Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality". In: *ACM Transactions on Computer-Human Interaction* 27, pp. 1–40. DOI: 10.1145/3361218.

Sidenmark, Ludwig, Zibo Sun, and Hans Gellersen (Mar. 2024). "ConeBubble: Evaluating Combinations of Gaze, Head and Hand Pointing for Target Selection in Dense 3D Environments". In: *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 642–649. DOI: 10.1109/vrw62533.2024.00126.

Sidorakis, Nikolaos, George Alex Koulieris, and Katerina Mania (Mar. 2015). "Binocular eye-tracking for the control of a 3D immersive multimedia user interface". In: *IEEE 1st Workshop on Everyday Virtual Reality (WEVR)*, pp. 15–18. DOI: 10.1109/WEVR.2015.7151689.

Singh, Shubham et al. (2020). *Real-time Collaboration Between Mixed Reality Users in Geo-referenced Virtual Environment*. arXiv.org.

Sloth, Emil et al. (Apr. 2023). "Partially Blended Realities: Aligning Dissimilar Spaces for Distributed Mixed Reality Meetings". In: *Scopus (Elsevier)*, pp. 1–16. DOI: 10.1145/3544548.3581515. (Visited on 01/30/2025).

Špakov, Oleg and Päivi Majaranta (2012). "Enhanced gaze interaction using simple head gestures". In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*. DOI: 10.1145/2370216.2370369.

Speicher, Marco et al. (Apr. 2018). "Selection-based Text Entry in Virtual Reality". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. DOI: 10.1145/3173574.3174221.

Speicher, Maximilian, Brian D. Hall, and Michael Nebeling (May 2019). "What is Mixed Reality?" In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. DOI: 10.1145/3290605.3300767.

Spittle, Becky, Maite Frutos-Pascual, et al. (2022). "A Review of Interaction Techniques for Immersive Environments". In: *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1. DOI: 10.1109/TVCG.2022.3174805.

Spittle, Becky, Payod Panda, et al. (May 2024). "Comparing the Agency of Hybrid Meeting Remote Users in 2D and 3D Interfaces of the Hybridge System". In: pp. 1–12. DOI: 10.1145/3613905.3651103.

Strauss, Marvin et al. (Sept. 2024). "Designing and Evaluating Scalable Privacy Awareness and Control User Interfaces for Mixed Reality". In: *arXiv (Cornell University)*. DOI: 10.48550/arxiv.2409.00739.

Strecker, Jannis et al. (Sept. 2023). "MR Object Identification and Interaction". In: *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 7, pp. 1–26. DOI: 10.1145/3610879.

Stropnik, Vid et al. (July 2018). *A Look into the Future of Sports: A Study of the Actual State of the Art - the Microsoft HoloLens and Augmented Reality*. IEEE Xplore. DOI: 10.1109/COBCOM.2018.8443967.

Su, Goh Eg, Mohd Shahrizal Sunar, and Ajune Wanis Ismail (Sept. 2020). "Device-based manipulation technique with separated control structures for 3D object translation and rotation in handheld mobile AR". In: *International Journal of Human-Computer Studies* 141, p. 102433. DOI: 10.1016/j.ijhcs.2020.102433.

Sun, Jiacheng and Ting Liao (Sept. 2024). "MazeMind: Exploring the Effects of Hand Gestures and Eye Gazing on Cognitive Load and Task Efficiency in an Augmented Reality Environment". In: *Design Computing and Cognition'24*, pp. 105–120. DOI: 10.1007/978-3-031-71922-6_7.

Symes, Ed, Rob Ellis, and Mike Tucker (Feb. 2007). "Visual object affordances: Object orientation". In: *Acta Psychologica* 124, pp. 238–255. DOI: 10.1016/j.actpsy.2006.03.005.

Tanikawa, Tomohiro et al. (Nov. 2015). "Integrated view-input ar interaction for virtual object manipulation using tablets and smartphones". In: *Proceedings of the 12th*

International Conference on Advances in Computer Entertainment Technology, pp. 1–8. DOI: 10.1145/2832932.2832956.

Togwell, Henry et al. (Apr. 2022). "In-cAR Gaming: Exploring the use of AR headsets to Leverage Passenger Travel Environments for Mixed Reality Gameplay". In: *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. DOI: 10.1145/3491101.3519741.

The Future of Proxemic Interaction in Smart Factories (May 2021). Vol. 2905. Proceedings of the Workshop on Automation Experience at the Workplace, AutomationXP 2021, co-located with the ACM Conference on Human Factors in Computing Systems (CHI 2021). CEUR-WS.org.

Tung, Ying-Chao et al. (2015). "User-Defined Game Input for Smart Glasses in Public Space". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pp. 3327–3336. DOI: 10.1145/2702123.2702214.

Turk, Matthew (2014). "Multimodal interaction: A review". In: *Pattern Recognition Letters* 36, pp. 189–195. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2013.07.003.

Ugarte, Jesus et al. (Mar. 2022). "Distant Hand Interaction Framework in Augmented Reality". In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. DOI: 10.1109/vrw55335.2022.00332.

Union, International Telecommunication (May 2023). *Recommendation ITU-R BT.500-15: Methodologies for the subjective assessment of the quality of television images*. URL: https://www.itu.int/rec/R-REC-BT.500-15-202305-I/en (visited on 04/16/2025).

Uva, Antonio Emmanuele et al. (Oct. 2019). "A User-Centered Framework for Designing Midair Gesture Interfaces". In: *IEEE Transactions on Human-Machine Systems* 49, pp. 421–429. DOI: 10.1109/thms.2019.2919719.

Uzor, Stephen and Per Ola Kristensson (Oct. 2021). "An Exploration of Freehand Crossing Selection in Head-Mounted Augmented Reality". In: *ACM Transactions on Computer-Human Interaction* 28, pp. 1–27. DOI: 10.1145/3462546.

Väyrynen, Jani et al. (Nov. 2018). "Exploring Head Mounted Display based Augmented Reality for Factory Workers". In: *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, pp. 499–505. DOI: 10.1145/3282894.3289745.

Venkatakrishnan, Roshan et al. (May 2023). "Give Me a Hand: Improving the Effectiveness of Near-field Augmented Reality Interactions By Avatarizing Users' End Effectors". In: *IEEE Transactions on Visualization and Computer Graphics* 29, pp. 2412–2422. DOI: 10.1109/tvcg.2023.3247105.

Verdi, Di, Daniel Nurmi, and Tobias Höllerer (Mar. 2004). "ARWin - a Desktop Augmented Reality Window Manager". In: *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.* DOI: 10.1109/ismar.2003.1240729.

Vergari, Maurizio et al. (Sept. 2022). *Investigation of Personal Space perception in Augmented Reality.* IEEE Xplore. DOI: 10.1109/QoMEX55416.2022.9900887.

Villarreal-Narvaez, Santiago et al. (July 2020). "A Systematic Review of Gesture Elicitation Studies". In: *Proceedings of the 2020 ACM Designing Interactive Systems Conference.* DOI: 10.1145/3357236.3395511.

Vogiatzidakis, Panagiotis and Panayiotis Koutsabasis (Aug. 2020). "Mid-Air Gesture Control of Multiple Home Devices in Spatial Augmented Reality Prototype". In: *Multimodal Technologies and Interaction* 4, p. 61. DOI: 10.3390/mti4030061.

Waldow, Kristoffer et al. (Nov. 2018). "An evaluation of smartphone-based interaction in AR for constrained object manipulation". In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–2. DOI: 10.1145/3281505.3281608.

Wang, Gang et al. (Dec. 2022). "Freehand Gestural Selection with Haptic Feedback in Wearable Optical See-Through Augmented Reality". In: *Information* 13, p. 566. DOI: 10.3390/info13120566.

Wang, Nanjia, Daniel Zielasko, and Frank Maurer (June 2024). "User Preferences for Interactive 3D Object Transitions in Cross Reality - An Elicitation Study". In:

Proceedings of the 2024 International Conference on Advanced Visual Interfaces, pp. 1–9. DOI: 10.1145/3656650.3656698.

Wang, Suyuchen et al. (Apr. 2021). "Enquire One's Parent and Child Before Decision: Fully Exploit Hierarchical Structure for Self-Supervised Taxonomy Expansion". In: *arXiv (Cornell University)*. DOI: 10.1145/3442381.3449948. (Visited on 09/15/2025).

Wang, Tianyi et al. (Oct. 2021). "GesturAR: An Authoring System for Creating Freehand Interactive Augmented Reality Applications". In: *The 34th Annual ACM Symposium on User Interface Software and Technology*. DOI: 10.1145/3472749.3474769.

Wang, Zhimin et al. (Nov. 2020). "Comparing Single-modal and Multimodal Interaction in an Augmented Reality System". In: *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* 1, pp. 165–166. DOI: 10.1109/ismar-adjunct51615.2020.00052.

Wei, Shu, Desmond Bloemers, and Aitor Rovira (June 2023). "A Preliminary Study of the Eye Tracker in the Meta Quest Pro". In: *IMX '23: Proceedings of the 2023 ACM International Conference on Interactive Media Experiences*. DOI: 10.1145/3573381.3596467.

Wei, Yaguang, Jason Orlosky, and Tomohiro Mashita (Mar. 2021). "Visualization and Manipulation of Air Conditioner Flow via Touch Screen". In: *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 430–431. DOI: 10.1109/vrw52623.2021.00097.

Weiser, Mark (July 1999). "The computer for the 21st century". In: *ACM SIGMOBILE Mobile Computing and Communications Review* 3, pp. 3–11. DOI: 10.1145/329124.329126. URL: https://doi.acm.org/10.1145/329124.329126.

Weiser, Mark and John Seely Brown (1997). "The Coming Age of Calm Technology". In: *Beyond Calculation*, pp. 75–85. DOI: 10.1007/978-1-4612-0685-9_6.

Whitlock, Matt et al. (Mar. 2018). "Interacting with Distant Objects in Augmented Reality". In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 42–48. DOI: 10.1109/vr.2018.8446381.

Williams, Adam S., Jason Garcia, and Francisco Ortega (Dec. 2020). "Understanding Multimodal User Gesture and Speech Behavior for Object Manipulation in Augmented Reality Using Elicitation". In: *IEEE Transactions on Visualization and Computer Graphics* 26, pp. 3479–3489. DOI: 10.1109/tvcg.2020.3023566.

Williams, Adam S. and Francisco R. Ortega (Nov. 2020). "Understanding Gesture and Speech Multimodal Interactions for Manipulation Tasks in Augmented Reality Using Unconstrained Elicitation". In: *Proceedings of the ACM on Human-Computer Interaction* 4, pp. 1–21. DOI: 10.1145/3427330.

Williamson, Julie et al. (May 2021). "Proxemics and Social Interactions in an Instrumented Virtual Reality Workshop". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. DOI: 10.1145/3411764.3445729.

Wobbrock, Jacob O., Lisa A. Elkin, et al. (2024). *ARTool*. Washington.edu. URL: https://depts.washington.edu/acelab/proj/art/ (visited on 09/03/2025).

Wobbrock, Jacob O., Leah Findlater, et al. (May 2011). "The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures". In: *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 143–146. DOI: 10.1145/1978942.1978963.

Wobbrock, Jacob O., Meredith Ringel Morris, and Andrew D. Wilson (2009). "User-defined Gestures for Surface Computing". In: *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, pp. 1083–1092. DOI: 10.1145/1518701.1518866.

Wolf, Erik et al. (Oct. 2019). ""Paint that object yellow": Multimodal Interaction to Enhance Creativity During Design Tasks in VR". In: *2019 International Conference on Multimodal Interaction*, pp. 195–204. DOI: 10.1145/3340555.3353724.

Xu, Wenge, Hai-Ning Liang, Yuzheng Chen, et al. (Mar. 2020). *Exploring Visual Techniques for Boundary Awareness During Interaction in Augmented Reality Head-Mounted Displays*. IEEE Xplore. DOI: 10.1109/VR46266.2020.00039.

Xu, Wenge, Hai-Ning Liang, Anqi He, et al. (Oct. 2019). "Pointing and Selection Methods for Text Entry in Augmented Reality Head Mounted Displays". In: *2019 IEEE*

International Symposium on Mixed and Augmented Reality (ISMAR), pp. 279–288. DOI: 10.1109/ismar.2019.00026.

Xu, Xuanhui et al. (Oct. 2022). "Using HMD-based Hand Tracking Virtual Reality in Canine Anatomy Summative Assessment: a User Study". In: *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 287–296. DOI: 10.1109/ismar55827.2022.00044.

Xu, Xuhai et al. (Apr. 2023). "XAIR: A Framework of Explainable AI in Augmented Reality". In: *arXiv (Cornell University)*. DOI: 10.1145/3544548.3581500.

Ye, Hui et al. (July 2020). "ARAnimator: In-situ Character Animation in Mobile AR with User-defined Motion Gestures". In: *ACM Transactions on Graphics* 39. DOI: 10.1145/3386569.3392404.

Yin, Jibin et al. (2019). "Precise Target Selection Techniques in Handheld Augmented Reality Interfaces". In: *IEEE Access* 7, pp. 17663–17674. DOI: 10.1109/access.2019.2895219.

Yu, Difeng et al. (Oct. 2019). "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)*, pp. 103–114. DOI: 10.1109/ismar.2019.00-20.

Zahid Iqbal, Muhammad et al. (Jan. 2023). "Security, Ethics and Privacy Issues in the Remote Extended Reality for Education". In: *Gaming media and social effects*, pp. 355–380. DOI: 10.1007/978-981-99-4958-8_16.

Zhang, Bowen et al. (Jan. 2022). "Semantic Sensing and Communications for Ultimate Extended Reality". In: *arXiv (Cornell University)*. DOI: 10.48550/arxiv.2212.08533.

Zhao, Caijun, Kai Way Li, and Lu Peng (Jan. 2023). "Movement Time for Pointing Tasks in Real and Augmented Reality Environments". In: *Applied Sciences* 13, p. 788. DOI: 10.3390/app13020788.

Zhao, Junhong et al. (Nov. 2020). "Voice Interaction for Augmented Reality Navigation Interfaces with Natural Language Understanding". In: *2020 35th International*

Conference on Image and Vision Computing New Zealand (IVCNZ), pp. 1–6. DOI: 10.1109/ivcnz51579.2020.9290643.

Zhou, Qiushi et al. (Dec. 2020). "Eyes-free Target Acquisition During Walking in Immersive Mixed Reality". In: *IEEE Transactions on Visualization and Computer Graphics* 26, pp. 3423–3433. DOI: 10.1109/tvcg.2020.3023570.

Zhou, Xiaoyan, Adam Sinclair Williams, and Francisco Raul Ortega (Nov. 2022). "Eliciting Multimodal Gesture+Speech Interactions in a Multi-Object Augmented Reality Environment". In: *VRST '22: Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–10. DOI: 10.1145/3562939.3565637.

Zhu, Fengyuan and Tovi Grossman (Apr. 2020). "BISHARE: Exploring Bidirectional Interactions Between Smartphones and Head-Mounted Augmented Reality". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14. DOI: 10.1145/3313831.3376233.

Zollmann, Stefanie et al. (2019). "ARSpectator: Exploring Augmented Reality for Sport Events". In: *SIGGRAPH Asia 2019 Technical Briefs - SA '19*. DOI: 10.1145/3355088.3365162.