DISSERTATION

INTERACTION AND NAVIGATION IN CROSS-REALITY ANALYTICS

Submitted by

Xiaoyan Zhou

Department of Computer Science

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2024

Doctoral Committee:

Advisor: Francisco Ortega

Indrakshi Ray
Marcia Moraes
Anil Ufuk Batmaz
Laura Malinin

ABSTRACT

INTERACTION AND NAVIGATION IN CROSS-REALITY ANALYTICS

Along with immersive display technology's fast evolution, augmented reality (AR) and virtual reality (VR) are increasingly being researched to facilitate data analytics, known as Immersive Analytics. The ability to interact with data visualization in the space around users not only builds the foundation of ubiquitous analytics but also assists users in the sensemaking of the data. However, interaction and navigation while making sense of 3D data visualization in different realities still need to be better understood and explored. For example, what are the differences between users interacting in augmented and virtual reality, and how can we utilize them in the best way during analysis tasks? Moreover, based on the existing work and our preliminary studies, improving the interaction efficiency with immersive displays still needs to be solved. Therefore, this thesis focuses on understanding interaction and navigation in augmented reality and virtual reality for immersive analytics.

First, we explored how users interact with multiple objects in augmented reality by using the "Wizard of Oz" study approach. We elicited multimodal interactions involving hand gestures and speech, with text prompts shown on the head-mounted display. Then, we compared the results with previous work in a single-object scenario, which helped us better understand how users prefer to interact in a more complex AR environment.

Second, we built an immersive analytics platform in both AR and VR environments to simulate a realistic scenario and conducted a controlled study to evaluate user performance with designed analysis tools and 3D data visualization. Based on the results, interaction and navigation patterns were observed and analyzed for a better understanding of user preferences during the sensemaking process.

Lastly, by considering the findings and insights from prior studies, we developed a hybrid user interface in simulated cross-reality for situated analytics. An exploratory study was conducted with a smart home setting to understand user interaction and navigation in a more familiar scenario with practical tasks. With the results, we did a thorough qualitative analysis of feedback and video recording to disclose user preferences with interaction and visualization in situated analytics in the everyday decision-making scenario.

In conclusion, this thesis uncovered user-designed multimodal interaction including mid-air hand gestures and speech for AR, users' interaction and navigation strategies in immersive analytics in both AR and VR, and hybrid user interface usage in situated analytics for assisting decision-making. Our findings and insights in this thesis provide guidelines and inspiration for future research in interaction and navigation design and improving user experience with analytics in mixed-reality environments.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Francisco R. Ortega, for his support, encouragement, and mentorship. I also would like to thank my Ph.D. committee members, Dr. Indrakshi Ray, Dr. Marcia Moraes, Dr. Anil Ufuk Batmaz, and Dr. Laura Malinin for offering their time and advice.

I'd like to acknowledge all of the members of the Natural User Interaction Lab who have helped me throughout my Ph.D. journey.

We would also like to thank our sponsors for their support. This work was supported by the National Science Foundation (NSF) awards: NSF-2327569, NSF-2238313, NSF-2223432, NSF-2223459, NSF-2037417, and NSF-1948254.

DEDICATION

This dissertation is lovingly dedicated to my advisor, Francisco, and to my parents. Francisco's unwavering support and guidance have been indispensable throughout this journey, both professionally and personally. To my parents, your endless patience and steadfast belief in me have been my constant motivation. I am also profoundly grateful to my dear friends Adam and Sarah, whose support and caring for me were invaluable. Thank you all for being my pillars of strength.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF WORKS RELEVANT TO THIS DEGREE

Papers included in this thesis:

Xiaoyan Zhou, Adam Sinclair Williams, and Francisco Raul Ortega. 2022. Eliciting Multimodal Gesture+Speech Interactions in a Multi-Object Augmented Reality Environment. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology (VRST '22)*. Association for Computing Machinery, New York, NY, USA, Article 17, 1–10. https://doi.org/10.1145/3562939.3565637

Xiaoyan Zhou, Anil Ufuk Batmaz, Adam S. Williams, Dylan Schreiber, and Francisco Ortega. 2024. I Did Not Notice: A Comparison of Immersive Analytics with Augmented and Virtual Reality. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3613905.3651085

Xiaoyan Zhou, Adam S. Williams, Francisco Ortega (2022). Towards Establishing Consistent Proposal Binning Methods for Unimodal and Multimodal Interaction Elicitation Studies. In: *Kurosu, M. (eds) Human-Computer Interaction. Theoretical Approaches and Design Methods. HCII 2022*. Lecture Notes in Computer Science, vol 13302. Springer, Cham. https://doi.org/10.1007/978-3-031-05311-5_25

Other papers not included in this thesis:

Xiaoyan Zhou, Yalong Yang, Francisco Ortega, Anil Ufuk Batmaz and Benjamin Lee, "Data-driven Storytelling in Hybrid Immersive Display Environments," *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, Sydney, Australia, 2023, pp. 242-246, https://doi:10.1109/ISMAR-Adjunct60411.2023.00056.

Xiaoyan Zhou, "Designing Navigation Tool for Immersive Analytics in AR," *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, Shanghai, China, 2023, pp. 975-976, https://doi:10.1109/VRW58643.2023.00330.

Adam S. Williams, **Xiaoyan Zhou**, Anil Ufuk Batmaz, Michel Pahud , Francisco Ortega (2023). A Pilot Study Comparing User Interactions Between Augmented and Virtual Reality. In: *Bebis, G., et al. Advances in Visual Computing. ISVC 2023*. Lecture Notes in Computer Science, vol 14362. Springer, Cham. https://doi.org/10.1007/978-3-031-47966-3_1

Chapter 1

Introduction

The emergence of virtual reality (VR) technology and the lower cost of owning a VR headset offered many consumers an opportunity to include VR in their daily entertainment. However, as augmented reality (AR) technology matures, people start to see how AR can really enhance their real-life quality. While blending the virtual world with the real world provides many potential, how to design the interaction with the virtual world as natural and effortless as possible is crucial to the popularity of this technology. To achieve the goal of developing efficient and effective interaction techniques, we first need to understand what it means to users. What strategy do users prefer when communicating with the virtual world? Does it relate to different tasks and scenarios? What factors influence those choices? Does it relate to the present of the real-world environment? Can we leverage the existing interaction techniques in the real world to facilitate user interaction in the virtual world? This thesis aims to answer those questions through multiple human-centered studies, with applications from simple to complex to realistic design.

1.1 Interaction in Simple Augmented Reality Environment

More and more people believe augmented reality will soon be an essential part of their daily lives, as the role of phones plays in our current society. Nevertheless, how to interact with AR technology is still in question. At the time back in 2020, HoloLens 2 showed it is no longer an issue for people to augment virtual objects in their physical environment with a lightweight head-mounted display. However, the interaction with virtual objects was rather unintuitive and uneasy to be recognized by the system, which led to users' frustration and less desire to use the headsets. Unsurprisingly, those expert-designed interactions include specific gestures and require not only continued physical effort but also high precision from users. To facilitate the popularity of augmented reality technology, easy-to-learn with less effort-required interaction techniques need to be suggested. Therefore, we aim to explore and create better interaction techniques for user interaction in AR. Firstly, we must understand how users prefer to interact in AR. By conducting an elicitation study, we collected both hand gestures and speech proposed by users while presenting text prompts in the display during the experiment. After the video coding and gestures binning process, we presented the winning gestures and ranked the top three proposals. The results were compared to a previous multimodal elicitation study with a single object AR environment.

1.2 Interaction with Immersive Analytics in Augmented Reality and Virtual Reality

Considering the results from the elicitation study with multiple objects, we wonder how users will interact in a more complex and realistic AR environment by using hand gestures. Therefore, we conducted a preliminary study with an immersive analytics platform in AR with 3D data visualization, by providing existing interaction techniques with HoloLens 2. However, users showed frustration with the interactions, which negatively affected the task completion time in 3D data analysis and user experience in AR. On the other hand, with similar tasks implemented in VR and interacting through VR controllers, users showed more engagement and better performance. To this end, we decided to investigate if only the different displays influence user experience when the same interaction techniques are adopted in both realities. Previous work has only compared interaction and navigation with text-based data in AR and VR [3], and the results can be various due to different data types, which can lead to different spread of data. Therefore, using the same immersive analytics platform, a comparison study with AR and VR was conducted with a mixed-reality headset, Varjo XR-3, to ensure the same field of view. In this study, users were guided to use hand controllers in both conditions to avoid the frustration caused by immature hand gestures. Detailed results and discussion were presented to show various user experiences in both conditions, and the factors should be considered in future studies. Our findings provided valuable insights into future design for interaction and navigation in AR and VR.

1.3 Interaction with Situated Analytics in Mixed Reality

The comparison study results showed that users' preferences and task difficulty influence which reality should be deployed. Nevertheless, users reflected on the difficulty of interacting with 3D visualization (e.g., rotation and scale), indicating mature interaction techniques or easy-to-navigate visualization can improve user experience. To explore the solution, our latest project adopted hybrid user interfaces to facilitate interaction by placing complicated visualization on a 2D screen and keeping the simple visualization in the 3D space attached to the related objects. It offloads the efforts for interacting all visualization in mixed reality with existing unnatural interaction techniques. Moreover, we want to understand how users utilize different interfaces and when they choose to use them. Also, knowing factors that influence their decision-making can help to improve the future design of the hybrid user interface, which leads to better user experiences and higher task performance efficiency with immersive analytics platforms. We used VR to simulate an apartment with situated AR visualization due to practical reasons and also environment consistency, with fewer confounding variables. A situated analysis scenario was adopted so the tasks are more realistic and as real-life decisions that users make in the real world. An exploratory study approach offered users the flexibility to set up goals based on personal needs and preferences. Both opportunities and challenges with using hybrid user interfaces in situated analytics were thoroughly discussed, and future research directions were presented.

1.4 Thesis Statement

The aim of this thesis is to improve users' experience and performance in interaction and navigation with augmented and virtual reality. We do so through three steps:

- As augmented reality (AR) technology emerges, expert-designed interaction techniques are implemented in consumers' AR head-mounted displays. However, user feedback with those interaction techniques is less than satisfactory. We seek to understand users' preferences for interaction in terms of hand gestures and speech. We do so by conducting an elicitation study with a Wizard of Oz approach to collect user-proposed gestures and speech with various referents (Chapter 3).

- With the insights of user preferences in interaction with a simple AR environment, we wonder if the same findings will stand with a more complex AR environment that requires sensemaking, and also in the VR environment. We developed an immersive analytics platform in both AR and VR and conducted a controlled study with a mixed-reality headset (Chapter 4).

- The findings from the controlled study indicate improving interaction efficiency is imperative, and one approach is to involve hybrid user interfaces (HUIs) which can provide user familiarity with 2D user interfaces and also offload the needs of using immature mid-air interaction techniques in 3D space. Moreover, we seek to understand if adopting HUIs can facilitate the user decision-making process in the context of situated analytics. We developed a situated analytics prototype with HUI in VR and conducted an exploratory study to identify the opportunities and challenges (Chapter 5).

1.5 Contribution

The following high-level contributions are made by this thesis:

- The design and development of a Wizard of Oz prototype in AR that aims to understand user-designed gestures and speech for interaction in augmented reality (Chapter 3).

- The design and development of an immersive analytics prototype in both AR and VR that aims to explore the differences in interaction and navigation during sensemaking with AR and VR (Chapter 4).

- The design and development of a situated analytics (SA) prototype with hybrid user interfaces (HUI) in VR that aims to identify opportunities and challenges with using HUI in SA (Chapter 5).

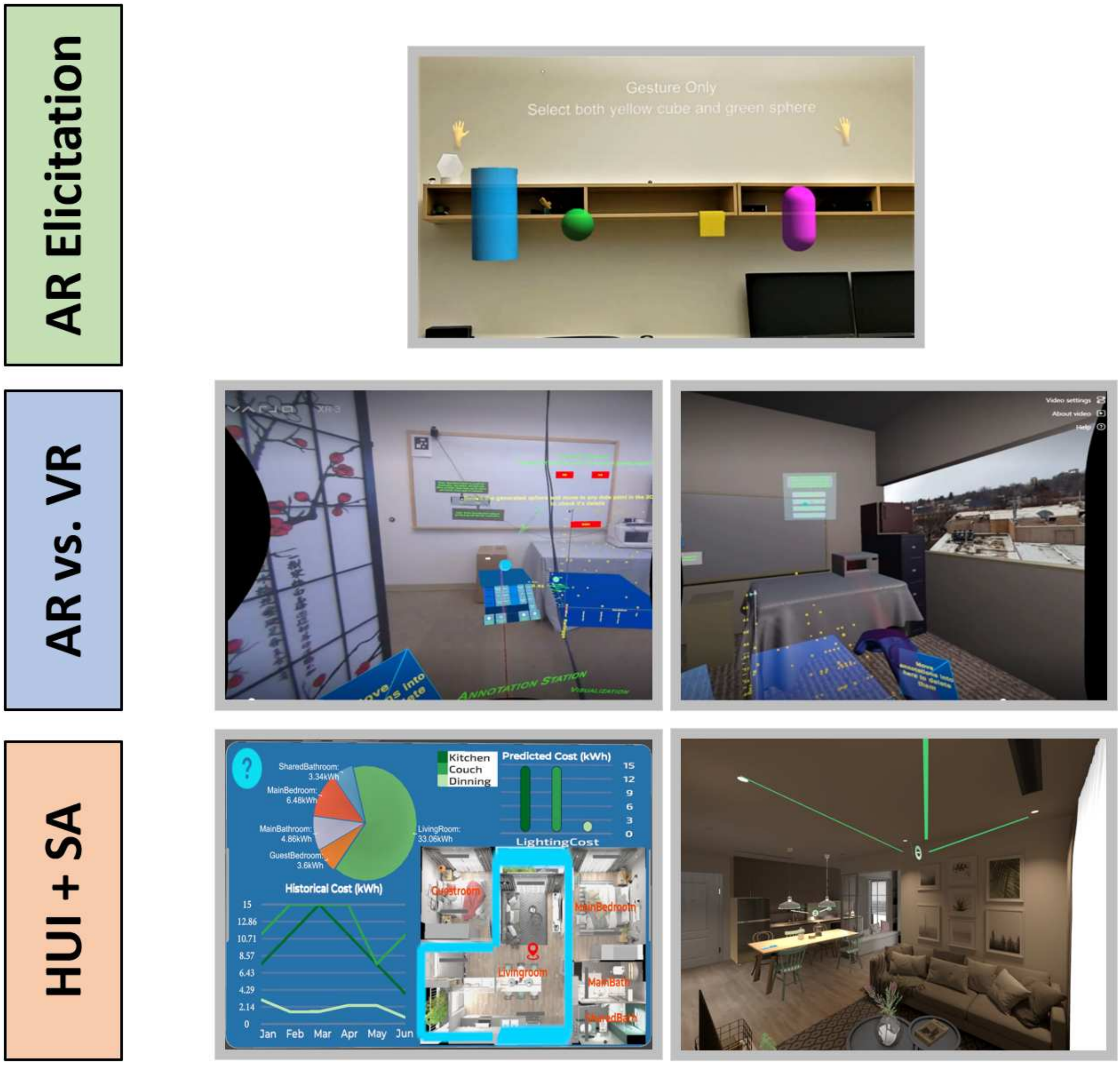


Figure 1.1: Summary overview of this thesis. (1) We investigate user-proposed gestures and speech in AR (Chapter 3). (2) We investigate user interaction and navigation in both AR and VR during sensemaking in immersive analytics (Chapter 4). (3) We investigate opportunities and challenges with using hybrid user interfaces during decision-making in situated analytics (Chapter 5).

1.6 Dissertation Layout

The rest of this thesis is structured as follows (Figure 1.1):

Chapter 2 firstly reviews the state-of-the-art across interaction and navigation with immersive technologies. It then provides the background for the research field of Immersive Analytics. Lastly, it provides the motivation for adopting cross-reality to facilitate all research topics in this thesis.

Chapter 3 presents an elicitation study aimed at understanding what intuitive hand gestures and speech are when users can propose them freely. With the Wizard of Oz approach, users were able to perform any hand gestures or speech on-site and get the response immediately, which was controlled by the experimenter. 24 participants were involved in this study and a within-subject were used across three conditions. Both quantitative and qualitative analyses were performed after data collection. This work was published and presented at the ACM Virtual Reality Software and Technology Conference in 2022 [4]. The related paper came from this study was published and presented at the HCI International (HCII) Conference in 2022 [5].

Chapter 4 presents a comparison study during the sensemaking process with an immersive analytics platform between AR and VR. The immersive analytics platform was developed based on toolkit IATK and implemented in both AR and VR environments. A digital twin of the AR physical world was simulated in VR. A controlled study with 40 participants was conducted with the mixed reality headset Varjo XR-3. Both quantitative and qualitative analyses were performed after data collection. This work was published and presented at the ACM CHI Late-Breaking Work Conference in 2024 [6].

Chapter 5 presents an exploratory study of a hybrid user interface with situated analytics AR scenarios simulated in VR. The goal is to understand how users use different visualizations in 2D surface and 3D space, how they navigate through the space and to the referents, what factors influenced their choices, and their experiences with using different user interfaces during the sensemaking process with situated analytics scenarios. 19 participants were included in this study, and a thorough qualitative analysis was performed after data collection. This work is submitted to IEEE VIS 2024.

Chapter 6 concludes this thesis by discussing its findings, contributions, limitations, and future research directions.

Chapter 2

Background

Along with the popularity of mixed reality technologies, various definitions show up regarding this new concept. The most common ones are mixed reality (MR), augmented reality (AR), and virtual reality (VR). Milgram and Kishino provide a taxonomy to illustrate how the real and virtual worlds can be blended with their famous Reality-Virtuality continuum [7]. To avoid ambiguity, we revisit the definitions with this continuum:

- **Virtual Reality.** VR is at one end of the Milgram and Kishino continuum, another end is the real environment. With VR, users are isolated from physical surroundings and fully immersed in a three-dimensional, computer-generated environment. The typical examples of this concept are CAVE Automatic Virtual Environment (CAVE) [8,9], and VR HMD such as Oculus Quest [10].
- **Augmented Reality.** With AR, users are able to experience both physical and virtual environments by fusing computer-generated objects with the real world. Some examples of this concept are mobile-based AR applications, and AR HMD such as HoloLens 2 [11].
- **Mixed Reality.** Based on the definition by Milgram and Kishino [7], anywhere between two ends of the continuum is MR, in which both the real and virtual world can be seen with a single display. In other words, MR is a superset that includes both AR and augmented virtuality (AV). With AV, the virtual environment is enhanced by real-world objects, which is the opposite of AR.

2.1 Interaction in 3D Space

In 1986, Hutchins et al. identified two gulfs at the interface that needed to be bridged [12] to reduce users' cognitive effort while performing a task through the interface. They are the gulf of execution and the gulf of evaluation. They represent the conceptual distance between the real system and the user's mental model of the system [13, 14]. The gulf of execution is caused by the mismatch between the actions required by the system and the users' intentions to achieve the goal. In other words, the true workings of the system and its requirements to the users fail to align with the users' mental model and expectations. On the other hand, the gulf of evaluation is caused by the mismatch between the actual state of the system and what users have perceived about the system's state. It could be due to the feedback that the system provides cannot be perceived or easy to interpret by users. Recently, works in the visualization domain have mostly focused on bridging the gulf of evaluation [15], which improved the understandability of the visualizations. However, less work has been done to explore intuitive interaction techniques to assist in immersive interaction, and also a lack of design guidelines for interaction in an IA environment [15].

The studies that have been done for interaction in 3D spaces in recent years mostly included free-hand gestures or hand-held controllers. The free-hand gestures are intuitive and easy to learn due to it being natural for users in real-world communication, which is a great implementation for getting users to accept the new technology quickly. However, the existing free-hand gestures are mainly mid-air gestures, which means the user needs to keep their arm in the middle of the air while performing those gestures. This can lead to muscle fatigue and arm tiredness after a short time which limits the usage and acceptance of the free-hand gesture for immersive interaction [15]. Previous work has proposed methods to measure the perceived exertion and arm fatigue [16, 17] and gave suggestions for designing fatigue-efficient gestures [17]. Based on those guidelines, some studies designed interaction techniques and interfaces that caused less physical exertion [18–20]. However, inputs from other body parts, such as the feet or head, are limited when compared to hands based on the illustration of Penfield's motor homunculus [21], and also not as natural as hands-based interaction. Several works have explored microinteraction with augmented objects, and they showed microgestures have the potential to be used for interaction in immersive environments with less physical effort [22, 23]. As Ens et al. suggested, to allow users to perform data analytics with continuous usage of the IA system, designers should consider users' physical exertion caused by all physical demands from interaction through gestures, interactions in the spatial realm, and navigation with physical movements [24].

To compensate for each modality's weaknesses and also to provide interactions closer to natural interaction between humans, multimodal interactions have been studied more often recently in augmented reality (AR) and virtual reality (VR) environments. While an interface supports multimodal interactions, it means the system provides the user with a richer set of input options [25]. For example, users can use complementary or equivalent input modalities to perform the task [26]. The complementary inputs enable a faster interaction since multiple modalities can be used simultaneously. The equivalent inputs can be used alternatively which enables adaption to the user's preference. Chandler et al. defined IA as a multimodal experience [27], however, few studies have looked into multiple modality inputs for interaction in IA [28]. Supporting multimodal interactions in immersive environments can help immersive analytics systems to be more prevalent, moreover, improving the usability of 3D visualization interfaces.

2.1.1 Multimodal Interaction with Free-Hand Gestures

Combining interactions from multiple input modalities has been studied more often in recent years. Multimodal interaction is a more natural approach to interacting with immersive environments since it mimics the communication between humans, which involves multiple channels. Therefore, supporting multimodal interaction in systems makes the interaction techniques more intuitive and discoverable. Also, some modalities have their limits to be used as an efficient input alone, and additional modalities can be used to circumvent the limitations. For example, eye-gaze has the "Midas Touch" issue that might unintentionally trigger an action [15]. Adding other modalities' input as a confirmation or refinement for gaze input can eliminate false positive detection [29, 30]. Cohen et al. proposed the complementarity principle for multimodal system design in 1989, which suggested picking modalities that can overcome each other's weaknesses based on their strengths [31]. Due to the complementary nature of multimodal techniques, one channel's input can be used to disambiguate another channel's input [32, 33]. The recognition rate can also be improved by combining inputs from multiple channels [34, 35]. Moreover, enabling different input modalities to complete the same task might increase the acceptance of new technology. The chance of discovering the functionality might increase due to multiple input modalities being supported by the new interface [14]. Before Piumsomboon et al.'s study [36], most prior multimodal interaction studies were conducted with 2D devices [37, 38]. Opposite to multimodal interaction, the unimodal interaction was provided by a single input channel and was previously often adopted for the interaction in a virtual environment. Using free-hand gestures for interaction with virtual objects has been proved as an intuitive unimodal interaction technique and was able to enhance user experience [39–41]. However, before Piumsomboon et al.'s study, most free-hand gesture researches in AR were conducted in a small space with very limited hand gestures.

To understand more about indirect and direct interaction techniques and how to support more intuitive and precise interaction in AR, Piumsomboon et al. designed and evaluated free-hand gestures and multimodal technique speech plus gestures for exploring natural interaction in augmented reality. The author proposed to offer users options for performing separate translations or rotations or combined manipulation. With the multimodal interaction, users' feedback indicated it provided better control and more precision, and it was more effective at both single and multiple-object manipulation. Also, Piumsomboon et al. pointed out that more command variants for the same interaction should be provided to users to increase the usability of speech. Furthermore, user feedback showed different preferences from different users. For example, some people preferred to use speech rather than gestures, and others liked the opposite. Therefore, providing complementary modalities for the same interaction could enhance the usability and user experience. The later book talked about multimodal interfaces and mentioned that while the task becomes more difficult, users will self-maintain limitations on working memory by using multiple modalities to distribute communicative information for decreasing the cognitive load [33]. However, while the task is simple or does not involve complex spatial content, users mostly will prefer to interact unimodally [33]. As a result, the multimodal system design should consider accepting either unimodal or multimodal input.

One of the free-hand gesture issues in Piumsomboon et al.'s study was limited control over small-size objects and objects in cluttered scenes. It is similar to issues with using ray-casting to select remote targets and occlusion in the scene. As mentioned in the paper [36], one solution could be to provide the zoom-in function with the whole AR scenario. Not surprisingly, zoom functionality is prevalent in today's 3D applications. Moreover, the ray-casting technique has evolved constantly to support the better acquisition of remote targets and acquisition in an occlusion scenario [42, 43]. Future work should consider effective ray-casting techniques for interaction with small targets or in cluttered scenarios in AR. It is worth noting that Piumsomboon et al. did not use the error rate as a measurement in their experiment. Although the paper mentioned that there were limitations with the speech and gesture recognition software in the experiment, the study eliminated the recognition issues by performing calibration with hand and speech for every participant. A possible solution could be adopting deep learning to improve the accuracy of speech and gesture recognition. Due to the increasing corpora of gesture and speech datasets, recent studies were adopting deep learning for gesture recognition and automatic speech recognition, which can be more beneficial than using single-layer hidden Markov models (HMMs) [33].

Many scenarios in an immersive analytics environment could benefit from using multimodal interaction techniques. For example, in the scenario that the user is trying to physically navigate to the target while getting assistance from the overview of data visualization, in order to maintain the focus on the main task, it could be very helpful if the user has the option to use speech plus simple microgestures instead of using only mid-air gestures or controller to manipulate the overview since it is difficult for most people to perform mid-air gestures while walking to the target. In another scenario, if the user needs to explore the data visualization in an immersive virtual environment while having an online meeting with the team, he or she can only use gestures to perform the navigation and interaction. In the opposite situation, while the user is typing on the phone or computer, he or she can only use speech to interact with the visualization. In those scenarios, providing multiple channels for performing the same interaction can support the user to perform the interaction while different modalities are available, and the user can still use other channels to complete the task while one channel is blocked at the moment [14].

2.1.2 Ray-casting with Controllers

Ray-casting is widely used for selecting targets in virtual reality or augmented reality environments. The virtual ray could be attached to a hand, a controller, or a head in the case of "gaze-based ray-casting" [14]. Prior work demonstrated that using ray-casting for remote selection was faster [44] than using other techniques such as GO-GO [45]. However, ray-casting is a powerful yet unstable technique when high precision is required for the target acquisition task. For example, users often suffer from unintentional hand jitter issues while using ray-casting for distant and small target selection. Moreover, the high density of the environment can increase the difficulty of target acquisition with ray-casting. Furthermore, occlusion happens when there are objects displayed in front of the target, which can cause ambiguity in the depths during selection. Many studies have been done to address these problems, such as using progressive refinement for selection [46–48]. It aims to progressively reduce the undesired targets through multiple steps, and no high-precision interaction is required from users [14].

Lu et al. conducted a study that investigated ray-casting techniques with bubble mechanisms and compared the new techniques with other 3D target acquisition techniques [43]. With the bubble cursor, the effective width of a target is decided by the surrounding Voronoi region, which means the location of the surrounding targets influences the size of the effective width [42]. This implied that the bubble cursor might not perform well in a density environment due to the small effective width. Nevertheless, it has been proved that using the bubble cursor can improve the speed of the selection task compared to other techniques [42, 49]. In Vanacken et al.'s study, participants were standing during the experiment, and changing the ray's position was sometimes beneficial for selecting targets with depth ray technique by avoiding certain surrounding targets to get larger effective width [42]. Moreover, based on previous studies about 3D space, using body movements for navigation can improve spatial memory [50] and maintain spatial orientation [51].

Therefore, body movement should be considered while designing effective target acquisition techniques. Also, combining bubble rays with other acquisition techniques could make use of different techniques' strengths for different scenarios. For example, by using the depth ray technique for selecting targets in scenes with high density or occlusion, the effective width could be less affected by the location of surrounding targets [42, 52].

Ray-casting is suitable for target acquisition in complex data visualization. While a data visualization contains a large number of data points, selecting the target data point could be troublesome and inefficient. Having effective target acquisition techniques is essential for helping 3D data visualization become prevalent since the dataset could have high density. It is unavoidable that users need to specify a certain data point before interacting with it. Based on previous research, remote target selection and close target selection could benefit from different acquisition techniques. Also, different navigation approaches in immersive environments could affect which technique is more suitable for target acquisition. For example, suppose an ego-centric navigation approach is adopted for the task. In this case, the user can move him or herself to the region of interest, and the target will mostly be close to the user, which means the direct grab or touch interaction is a better choice for target acquisition to increase engagement. On the other hand, if an exo-centric navigation approach is used for the 3D data analysis task, the user cannot get close to every target, in which case the ray-casting technique is mostly preferred for target acquisition.

2.1.3 Hybrid User Interface

Feiner and Shamash brought the concept of HUI in 1991 [53], as merging the "visual and interaction spaces [of] heterogeneous display and device technologies [to] take advantage of the strong points of each [device]." Specifically, complementary devices such as AR head-mounted displays (HMDs) and smartphones can be used together to offset each other's disadvantages during interaction. Researchers have included a wide range of mobile devices to facilitate the interaction with AR and VR HMDs, for example, smartphones [54–58], smartwatches [59], tablets [60–67], computers [68, 69], walls [70, 71], and other interactive surfaces [72–75]. While most studies investigate the usability and user experiences of using various HUIs simultaneously, several works explore using hybrid user interfaces asynchronously [68,69]. With the option to use heterogeneous devices sequentially, users are able to choose the appropriate device for certain tasks.

Hybrid User Interface (HUI) has been explored more often in recent AR and VR HMD studies [63, 68, 76] to investigate how to compensate for individual shortcomings while exploiting the strength of each device. Compared to designing the 2D user interface, designers have more to consider when designing a 3D user interface, including unlimited 3D space and various interaction modalities, such as hand gestures, controllers, speech, eye gaze, etc. This variety of possibilities not only provides more potential but also requires designers to have more considerations before making a suitable choice. The opportunities and challenges of adopting HUI in mixed reality are discussed in the recent paper of Zaky et al. [77]. Three optimization objectives which aim to improve user experience and usability were illustrated:

- **Reachability.** The unreachable issue is caused not only by outside-of-range distance but also by the dynamic physical environment, which leads to different surroundings that restrict their reachability. Adopting HUI to integrate various devices as an alternative interaction approach in order to offer user devices that are available in the current environment or devices that are more convenient to access for interaction. In addition, those easy-to-locate devices are normally more familiar for users to interact with, compared to HMDs.

- **Ergonomics.** Implementing HUI offers solutions to mitigate the arm fatigue issue, which is famous in 3D user interaction due to the mid-air arm posture. Replacing the mid-air interaction entirely is not necessarily beneficial to optimizing ergonomics for every body part, such as while the comfortable position for the arm is waist-level and for the neck is eye-level. With HUI, it is plausible to consider ergonomics quality for the particular UI element without compromising others.

- **Visibility.** The visibility issue categories to problems that can be improved by better design, such as cluttered environments and far distances, and problems that are out of designers' control, such as narrow field of view of HMDs and intensive background lighting. Adopting HUI provides options to offload certain content to different devices. To deal with specific causes, choose devices that can improve resolution and contrast, a bigger display area for detailed information, or just move overlapping content to a separate device to avoid occlusion.

In addition, three challenges need to be addressed while deploying hybrid user interfaces:

- **Availability of the devices.** The basic knowledge of the available devices, such as their capabilities and affordances they have are crucial for HUI implementation. Moreover, it is important to know the ownership of the devices to ensure the accessibility for HUI deployment.

- **Interaction cost calculation.** To choose the optimal interaction, designers need to know about the interaction cost to achieve individual optimization objectives. Furthermore, having multiple objectives simultaneously in realistic scenarios is unavoidable, which increases the complexity of mitigating interaction costs. Nevertheless, designers must minimize the interaction cost by addressing the conflict between different objectives.

- **Cost of switching between devices.** The divided UI causes the attention switch which increases cognitive load and interaction time. Additionally, including more devices for interaction, affects the discoverability and learnability of the system with different input modalities, interface locations, and various feedback.

Hybrid user interfaces show potential for improving user experience with augmented and mixed reality HMDs by offering opportunities to optimize common concerns in 3D user interface design, such as reachability, ergonomics, and visibility. Also, to fully benefit from using HUI, designers and creators need to consider the issues associated with including multiple devices for user interaction in 3D space.

2.2 Immersive Analytics

Immersive analytics (IA) was introduced in 2015 [27], which aims to explore means to use the technologies of immersive interaction and display to facilitate decision-making and analytical reasoning, as Marriott et al. defined the term in the Immersive Analytics book "the use of engaging, embodied analysis tools to support data understanding and decision making." [78]. Immersive analytics integrates multiple research areas, including data visualization, visual analytics, augmented/virtual reality, computer graphics, and human-computer interaction [78].

Many research efforts have been made to explore IA in terms of developing toolkits and frameworks, such as DXR [79] and IATK [80], evaluation of visualization techniques in IA [81–85], and immersive analytics in AR regarding the spatial perception [86], spatial interaction [87], and 3D exploration [88]. Nevertheless, Ens et al. disclosed 19 challenges in immersive analytics in 2021 [24], in which mentioned that even one of the advantages of using IA is providing users with embodied direct manipulation; however, it also leads to arm fatigue which is famous with mid-air interaction. Also, the paper pointed out that many interaction techniques used in IA were difficult for users to discover due to their novelty and unintuitive, which induce high entry barriers to interaction in IA and increase the learning curve for users. The author presented a common analytic scenario with IA that involves transitions between conventional desktops to immersive environments, in other words, using HUI for immersive analytics. With inconsistent interaction techniques between 2D and 3D user interfaces, interaction fluidity could be a big barrier for users to choose to adopt IA. To this end, the authors suggested future researchers bridge the gap between immersive techniques and traditional interfaces.

2.2.1 Situated Analytics

Situated Analytics (SA) was introduced by ElSayed et al. [89]. Thomas et al. later described SA as “the use of data representations organized in relation to relevant objects, places, and persons in the physical world for the purpose of understanding, sensemaking, and decision-making.” [90] In other words, situated analytics aims to support analytical reasoning by embedding the visualizations in a physical location using AR [91]. In a recent survey of situated analytics, a couple of criteria are generated to definite SA [92]:

- **Presents data.**
- **Uses interactive visualization.**
- **Deployed using Augmented Reality techniques.**
- **Utilizes the user's physical location.**
- **Integrates analytical reasoning.**

According to the survey, while interaction is a key component of situated analytics, most of the work focuses on improving data presentation rather than user interaction experience. Nevertheless, they found most research used innovative interaction techniques such as touch, gestures and controllers. However, evaluation and validation were lacking for those interaction techniques. Future work is needed to develop interaction facilitating the sensemaking process and also measure their effectiveness.

2.3 Navigation in 3D Space

Navigation includes two components: travel and wayfinding [14]. Wayfinding is a cognitive process of making decisions about the route to the destination. Travel is the motor component that allows the user to control the position and orientation of the viewpoint. Wayfinding requires the user to have spatial knowledge, including knowledge of their surrounding environment. Travel can be achieved through simple physical motions such as walking or other travel techniques such as teleportation. Compared to physical movement during exploration, teleportation assists users in translating to a specified destination instantly. However, it can cause spatial disorientation and discontinuity, which negatively influences the user's mental model. Previous work explored using "minimap" for teleportation in VR so that users were able to be teleported to a location outside of their field of view (FOV) [93]. Yet the teleportation in "minimap" could lead to greater context-switching costs due to a lack of spatial context during teleportation and decreased knowledge of the surrounding environment. Meanwhile, a prior study showed that users have more physical motion in AR, possibly because of increased situational awareness [94].

Researchers found that a large fully-immersed environment has drawbacks such as lacking overview and blind spots [84]. The survey results of IA showed that context-aware navigation techniques such as World In Miniature (WIM) were not exploited enough, even though context awareness is critical for the user while the user is exploring large-scale data visualizations in an immersive environment [28]. Also, based on prior works [94, 95], a single view can lead to inaccuracy depth estimates due to the insensitivity of the human eye to the perception of depth along the direction of the version. For mitigating the error, studies showed that providing multiple views can benefit the correct perceptive projection development [95]. Or users would spend more time and energy examining the visualization from different viewpoints on navigating around the visualization to ensure the accuracy of position estimation [94]. Therefore, offering users WIM with multiple views of the 3D visualization can help users achieve efficient and accurate navigation in the AR environment.

2.3.1 Navigation in IA

Based on recent literature on IA exploration and navigation in VR [93, 96–100] and AR [63, 101, 102], the research topics mainly focused on users' view (egocentric vs. exocentric) [98, 99], navigation modes (locomotion vs. manipulation) [93, 99], navigation cues (frame of references, landmarks) [99, 103], navigation interaction (mid-air gesture vs. controllers) [96, 102], navigation techniques (overview vs. zoom) [93, 101]. However, every implementation has its drawback, and no study has combined options and taken advantage of different methods while they can compensate for each other. For example, several studies showed that providing an overview in a room-scale 3D visualization can prevent disorientation and blind spots while users navigate by walking inside the visualization [96, 104]. But an overview can cause clutter in the field of view (FOV) and distract users while performing wayfinding which can end up with a high cognitive load [93, 105]. Also, several studies have looked at navigation cues for assisting navigation in the fully immersed environment, and it either placed in the environment ahead by the researchers [99, 106] or created by participants later while they are inside the 3D visualization [103], which can be either less effective or increase users' workload. Moreover, previous work has shown that both egocentric and exocentric can be beneficial to different types of tasks. No studies have utilized both views during the navigation process for IA. While studies have looked at panning, zooming in 3D, and improving interaction techniques to assist exploration in IA [101, 107, 108], no work has looked at microgesture navigation. In addition to navigation and exploration, interaction techniques for virtual objects in an IA environment are important. Studies have suggested different interactions for different tasks [44,109–112], making it difficult to generalize findings. For example, raycasting works better for distant targets, while hand manipulation was recommended for near targets [44,110,112]. Also, it often requires high-precision raycasting from the fingertips or palm and can be more difficult to accomplish in a limited time when the scene is cluttered [113]. Therefore, another obvious advantage of microgestures is that users can interact much faster than mid-air gestures.

Several studies have shown that enough training can improve the task performance of users with the low spatial ability with visualization manipulation to navigate the virtual dataset [114]. Moreover, enough training can also help users become familiar with the new interaction techniques and slowly lower the error rates [106]. Prior studies on cross-virtuality application design suggested that a unified interaction technique should be implemented across different reality environments [115]. This can reduce the efforts that users need to put into interacting with various environments and also lower their cognitive load.

2.4 Cross-Reality

With the concept of the metaverse being wildly spread and the technology of mixed reality getting more mature and accessible, utilizing the complementary display modalities with a mixed reality environment has became more feasible. Including multiple modalities allows the system to overcome the trade-offs of different displays while bringing and proving the flexibility for users to switch between display modalities as needed while tasks are performed. Based on the continuum of Mixed Reality (MR) stated by Milgram and Kishino in 1994 [7, 116], it is possible to adjust the degree of virtuality and provide users with diverse experiences since the physical and digital worlds are not a dichotomy. The continuum includes Augmented Reality, Virtual Reality, and Augmented Virtuality (AV). The system that makes use of multiple locations from the continuum is known as Cross-Reality (CR) [117] or Cross-Virtuality (XV) [118].

Different display modalities have diverse, unique advantages and disadvantages when they are compared to each other. Specifically, in an AR environment, receiving instructions from the real world (e.g., paper documents) or collaborating with others in the shared space (e.g., coworkers) will be feasible, which could be beneficial in scenarios such as training or collaboration. On the contrary, users can be more focused in VR, an ideal condition for exploring or learning, or performing tasks requiring concentration and personal space.

When both AR and VR are used for 3D navigation procedures, researchers should consider which environment should be adopted for better task performance based on each environment's strengths and weaknesses. Below are some elements that influence navigation efficiency in AR and VR.

Strength of AR:

- **Accessible of real-world references** – the real objects in the psychical world are visible in AR, which means they can assist the user in navigating the 3D visualization by using physical objects as frames of reference or landmarks.

- **Receive spatial knowledge** – the visibility of the real world increases users' situation awareness, which facilitates navigation efficiency and ensures users' safety.

- **Collaboration** – the visibility of the psychical world enables the user to receive instructions from the real world (e.g., paper documents) or collaborate with others in the shared space (e.g., gestures or body language from coworkers during the navigation tasks.

Weaknesses of AR:

- **Color/context** - the visibility of color and/or context in the user's space can influence participants' perceptions of the environment [119], which could increase the difficulty of navigation in AR.

- **Cluttered backgrounds** - the visibility of backgrounds in AR can cause distraction and affect users' navigation performance [120].

Strength of VR:

- **Unlimited space** – isolated from the real world provide the option to create an unlimited world, and users are able to travel inside with teleportation techniques.

- **Less distraction from the real world** - isolated from the real world with head-mounted displays prevent users from being distracted during navigation tasks.

- **Create illusion** - isolated from the real world, provide the chance to manipulate the travel route and speed rate while the user walks to the destination in a limited physical space.

Weaknesses of VR:

- **Depth Estimation Error** – distance compression is a known effect in VR [121, 122], which can cause inaccurate destination estimation.

- **No spatial knowledge** – isolated from real-world cues, which can lead users to feel disoriented easily during the navigation procedure [123].

2.4.1 Cross-Virtuality Analytics

Cross-Reality and Cross-Virtuality have gained much attention in the field since different environments are complementary, and designers or developers can adopt different display modalities for the most suitable scenario. In 2020, Riegler et al. defined "Cross-Virtuality Analytics" as a new approach to interacting with data visualization due to the possibility of seamless transit across the spectrum of the MR continuum [118]. Choosing the term "Cross-Virtuality" instead of "Cross-Reality" was based on the early technology that was used for immersive analytics leaning to the virtuality side of the MR continuum. They believed that the involvement of the reality side of the continuum could facilitate the acceptance of immersive analytics and allow users to collaborate with team members working in the same psychical space. Nevertheless, only a few XV applications are known in the HCI field [122, 124–128]. Recently, Anthes identified technical and conceptual considerations related to XV application design [115]. Regarding conceptual design, the paper suggested that appropriate visual translation techniques between different continuum stages are important, and the interaction techniques should be consistent throughout different continuum stages.

2.4.2 Transitional Interfaces

One of the early works of XV application is the Magic Book demonstrated by Billinghurst et al. in 2001 [124, 125]. In the study, users could smoothly transition from physical 2D book to augmented reality and then to virtual reality scene with a handheld video see-through device with an inertial tracker for enabling AR and VR. Along with the switching between stages of the MR continuum, users' viewpoint are also changing from exocentric in AR to egocentric in VR. The Magic Book project also defined the Transitional Interface (TI), which aims to enable users to move freely along the reality-virtual continuum of Milgram and Kishino. After the concept of TI was formalized by Grasset et al. in 2006, several works were conducted to explore different TIs [129–132]. By switching between different displays, interaction modalities, data representations, and system functionalities, users can choose the best-fit stage for performing tasks at hand [133, 134]. Riegler et al. claimed that a good design of TIs could help achieve "a seamless integration and transition between conventional 2D visualization, augmented reality, and virtual reality in order to provide users with optimal visual and algorithmic support with maximum cognitive and perceptual suitability, depending on their current tasks and needs." [118]. Recently, Friedl et al. identified four elements for evaluating TIs: Presence, Physical Discomfort, Spatial Orientation, and Cognitive Load [135].

Adopting hybrid user interfaces in cross-reality or cross-virtuality scenarios, especially when HUI includes different devices, can result in significant overhead due to switching between various device classes (e.g., tablet and VR HMD). In this context, a well-designed transitional interface is crucial to the hybrid user interfaces. Hubenschmid et al. identified the aspects that should be considered in the above scenarios:

- **Maintain context.** When there are heterogeneous devices and differing data representations, how do we maintain the context and spatial memory of users when switching between devices and different visualizations?

- **Liking content.** When there are heterogeneous devices and differing data representations, how do we help users establish a mental connection between different visualizations with the same content?

- **Switching cost between devices.** With heterogeneous devices, how can we facilitate users to switch between different hardware and accomplish transition between environments?

- **Consistent user experience.** Heterogeneous devices can lead to inconsistent interaction (e.g., WIMP interfaces versus 3D user interfaces), how can we mitigate the mental demands caused by the inconsistency during transformations between realities?

Chapter 3

Multimodal Elicitation Study

Interaction is a key component of AR technology adoption and popularity. Previous studies found that individual preference influences the choice of modality while users interact with AR. To develop and provide more intuitive and natural interaction techniques than the existing hand gestures for AR HMDs, it is necessary to explore user-designed gestures and speech. Section 3.1 reports on the elicitation study results by using AR HMD HoloLens 2 with the Wizard of Oz study approach.

During the analysis of elicited gestures, we found there is barely any literature discussing the detailed binning procedure for the elicitation study results. As a crucial step towards calculating the agreement rate and winning gestures, a guideline for binning users' proposed gestures is more than necessary. As a result, Section 3.2 reports on the suggested binning methods based on our gesture elicitation study results, which aims to lead to more future work in formalizing a guideline for consistent proposal binding methods.

Both studies are presented in this chapter as they were published [4, 5]. These articles are included with copyright holder permission.

3.1 Eliciting Multimodal Gesture+Speech Interactions in a Multi-Object Augmented Reality Environment[1]

3.1.1 Overview

As augmented reality (AR) technology and hardware become more mature and affordable, researchers have been exploring more intuitive and discoverable interaction techniques for immer-

[1] Verbatim From Xiaoyan Zhou, Adam Sinclair Williams, and Francisco Raul Ortega. 2022. "Eliciting Multimodal Gesture+Speech Interactions in a Multi-Object Augmented Reality Environment". *In Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology (VRST '22). Association for Computing Machinery, New York, NY, USA, Article 17, 1–10. https://doi.org/10.1145/3562939.3565637*

sive environments. This paper investigates multimodal interaction for 3D object manipulation in a multi-object AR environment. To identify the user-defined gestures, we conducted an elicitation study involving 24 participants and 22 referents using an augmented reality headset. It yielded 528 proposals and generated a winning gesture set with 25 gestures after binning and ranking all gesture proposals. We found that for the same task, the same gesture was preferred for both one and two-object manipulation, although both hands were used in the two-object scenario. We present the gestures and speech results, and the differences compared to similar studies in a single object AR environment. The study also explored the association between speech expressions and gesture stroke during object manipulation, which could improve the recognizer efficiency in augmented reality headsets.

3.1.2 Introduction

Easy-to-remember gestures produce high usability interfaces [136]. A gesture set that does not align with users' expectations or mental models often leads to frustrating user experiences [137]. Wobbrock et al. introduced an elicitation methodology to collect proposed gestures from users [138], facilitating intuitive gesture design without implementing perfect interaction recognizers in advance. Prior findings indicated that users prefer to choose input modalities based on their needs during the interaction [139–141]. Previous elicitation studies mostly focused on gesture set design for different devices and interfaces [138, 141–144], and several studies have explored multimodal interactions with speech and hand gestures [145–147]. However, few to no researches have involved multimodal interactions in a multi-object augmented reality (AR) or virtual reality (VR) environment.

These prior works raised multiple questions that were explored during this study: Does multimodal interaction look different in multi-object AR environments? How does a multi-object AR environment impact an elicitation study's gesture and speech proposals? What gestures do users prefer for multiple object manipulation, and are there any differences from single object manipulation? What speech commands do users choose for multiple object manipulation, and what are the differences compared to single object manipulation? The raised questions drive our motivation to understand if previous single object studies can transfer to more realistic environments. For this work, an elicitation study was conducted for multimodal interaction in AR with a Wizard of Oz (WoZ) experiment design (i.e., a researcher emulating a live system) [138, 148]. It involved 24 participants completing 22 referents (i.e., command) each in an AR head-mounted display (HMD). It yielded 528 proposals, which were used to generate a winning gesture set with 25 gestures after binning and ranking. We compared our single virtual object manipulation proposals to the findings from prior studies in a single object AR environment [145, 147, 149]. For multiple object manipulation proposals, we compared them with the proposed gestures of single virtual object manipulation in our study. To the best of our knowledge, this is the first study to conduct multi-object mid-air interaction using optical-see through augmented reality headsets.

3.1.3 Related Work

Elicitation methodology has been widely used in the HCI field to collect user-defined gestures. Wobbrock et al. popularized an elicitation methodology to collect proposed gestures from users [138], which aims to assist in designing more intuitive [138], guessable [150], learnable, and memorable [151] interaction techniques. Morris et al. found that people prefer gestures proposed by end-users, which were less complex than ones designed by human-computer interaction (HCI) experts [152]. Based on recent literature review results [153], over two hundred studies have adopted the use of an elicitation methodology in their work. Elicitation studies can provide valuable insights for developing new interaction techniques.

Prior findings proved that users prefer to choose input modalities based on their needs during the interaction, such as choosing gestures over speech in a quiet environment [139–141]. Wobbrock et al. [138] discovered that having synonyms in a user-defined gesture set can increase the guessability of proposed gestures. A multimodal elicitation study provides the opportunity to create multimodal synonyms [141], which can offer users different modalities to achieve the same effect and increase the acceptance of new technology.

Nevertheless, although the realistic scenarios in AR interaction include more than one virtual object, most elicitation studies involving mid-air gestures in AR only considered single object manipulation in a single object AR environment [145, 147, 149, 154]. Pham et al. conducted an elicitation study with an AR headset that included a scenario of single building manipulation among multiple buildings [149]. However, the whole model was attached to a physical surface so that the elicited gestures in the study were not mid-air gestures. Moreover, as far as we know, no research has been done on multimodal interactions with multiple object manipulations in AR. Piumsomboon et al. implemented an elicitation study in AR (video-see through) that asked participants to select multiple objects, yet the elicited gestures were surface gestures [154]. Wittorf et al. adopted an elicitation methodology for exploring mid-air gestures, yet the participants were interacting with a wall display [155]. Danielescu and Piorkowski conducted an elicitation study to explore free-space gestures with a projector display that included multiple target selections among a set of photos [156]. However, the referent showed that photos were selected one by one, which could bias the participants' gesture proposals. Furthermore, Wobbrock et al. found that users preferred one-handed over two-handed interactions for tabletop interaction [138, 157]. As a result, we were interested in whether users preferred two hands for two-object manipulation in an AR environment. To fill the gap in multimodal interaction in a multi-object AR environment, this work conducted a study to elicit speech and mid-air gestures in an augmented reality environment that contains four virtual objects (Figure 3.1) and compared results to elicited proposals with a single-object AR environment.

3.1.4 Study Design

This study conducted the elicitation experiment using a similar process as previous work [138, 148, 158]. Twenty-two tasks (i.e., referents) were used for each modality in this work. Of those, 17 basic referents were selected based on their inclusion in prior works [145, 147], while the other five were developed to be multi-object versions of basic referents. Specifically, participants were required to manipulate two objects simultaneously when multi-object referents were presented.

Referents included six translations (along x, y, and z axes), six rotations (about x, y, and z axes), three abstract actions (create, destroy, and select), and two scaling actions (enlarge and shrink). For multiple object manipulation, only abstract and scale referents were included. There were three experiment blocks in this study, which included modality gesture only (G), speech only (S), and gesture plus speech (GS). Each block took approximately 10 minutes, plus two questionnaires and three surveys. The experiment lasted approximately 45 minutes.

Participants

The study involved 24 participants (12 female, 12 male). Due to the pandemic, it was difficult to recruit outside of the Computer Science (CS) department; therefore, 17 out of 24 participants came from CS. Their ages ranged from 18-34 years (Mean = 23.42, SD = 4.20). All participants had previously used multi-touch devices, nineteen had used motion sensing devices (e.g. Xbox Kinect or Nintendo Wii Motion), sixteen had used virtual reality headsets, and three had used augmented reality headsets.

Setup

The experiment was conducted using Microsoft HoloLens 2 optical see-through AR head-mounted display (HMD). The system used for the experiment was developed in Unity Engine 2019.4.4f1. A GoPro Hero 7 Black was mounted on top of HoloLens 2 to record an ego-centric view of the interactions, as shown in Figure 3.2. A 4k camera was placed on the front left corner facing participants to record an exo-centric view of the interactions. Two hand-shape icons on the screen were used to indicate if the hand or hands were in the view of the headset [159], as shown in Figure 3.1. If either hand was out of view, the corresponding hand icon would disappear from the screen. Before starting the experiment, participants were requested to complete the informed consent and demographics questionnaire. Then participants were informed that there would be three experiment blocks with different modalities as input and they could use any interaction they felt was appropriate to execute the command based on the presented text referent and input modality. Participants were told to perform gestures inside the headset's view, which they

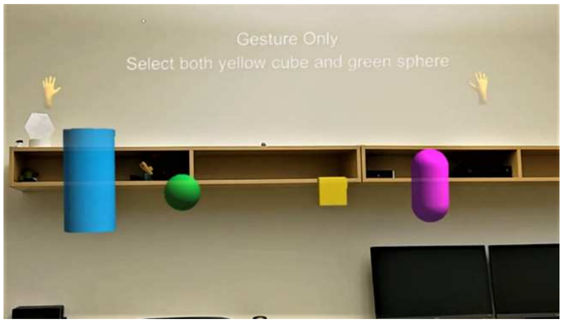


Figure 3.1: Participant view

Figure 3.2: Experiment setup

could tell by the hand icons displayed. The interaction modalities were presented to participants in a counter-balanced order. In each block, referents were presented in random order. The post-study questionnaire was filled out by each participant at the end of the experiment.

Hypotheses

Our hypotheses were grounded in previous observations in our lab and from previous work [145]: H_1) for the same single object manipulation referent, winning gestures in a multi-object AR environment will be different from ones in a single object AR environment; H_2) participants would prefer to use both hands for two object manipulation referents. Moreover, through elicited multimodal interaction, further connections between speech commands and hand gestures are expected to be found in this study.

3.1.5 Results

With the experiment, 528 proposals were collected from each modality. To eliminate the effect of the referent text biasing the speech proposal [145], prior to the analysis, speech proposals that were identical to the text displayed as part of the referent were removed. Resulting in 277 proposals from GS block and 261 proposals from S block.

The agreement rate ($\mathscr{AR}$), co-agreement rate ($\mathscr{CR}$) and (V_{rd}) significance test were used to determine consensus among gesture proposals [160]. $\mathscr{AR}$ is used to quantify the consensus of the binned proposals for interactions by referent [161], as shown in Eq. 3.1. $\mathscr{CR}$ is used to measure the amount of agreement shared between referents [160]. This study adopted Fliess's

Kappa coefficient (k_F) and the related chance agreement term (p_e) [158] when presenting the overall agreement rate of gesture proposals. The bootstrapped 95% confidence intervals were calculated to provide an interval estimate of each agreement score [158]. We used the AGATe 2.0 tool (AGreement Analysis Toolkit)[2] to assist our statistical analysis. The consensus-distinct ratio (CDR) was adopted to quantify the speech proposals [141]. For a complete treatment on elicitation studies and methods, see Williams et al. [161].

$$\mathscr{AR}_r = \frac{\sum_{P_i \subset P} \frac{1}{2}|P_i|(|P_i|-1)}{\frac{1}{2}|P|(|P|-1)} \tag{3.1}$$

The agreement rate $\mathscr{AR}$ for each referent r was calculated with Eq. 3.1. In Eq. 3.1, P is the set of all proposed gestures for referent r, and P_i are the subsets of identical proposed gestures from P.

The overall agreement rate for gestures from G and GS blocks was .190. Based on the interpretations proposed by Vatavu and Wobbrock [160], **our study achieved a medium agreement with 12 referents and a high agreement with 4 referents**. The individual agreement rate of gestures from G block and GS block alone were also calculated. The G block has .189 in agreement rate with k_F coefficient of .165. The chance agreement term p_e was .029, which indicates that the probability of agreement occurring by chance was minimal [158]. The GS block obtained .193 agreement rate with k_F coefficient of .151, and the chance agreement term p_e was .050, which shows evidence of agreement beyond chance. Compared to the previous elicitation study results in the single 3D object environment [145], we have lower agreement rates in general.

Unimodal Gesture and Unimodal Speech

Gesture Only We observed a significant effect of referent type on agreement rate in G block ($V_{rd(21,N=528)}$ = 639.363, p <.001). The study found there were 13 referents who obtained a medium to high agreement ($\mathscr{AR} > .10$), and they showed significant difference between agreement rates ($V_{rd(12,N=312)}$ = 191.492, p <.001). Accordingly, 9 referents have agreement rates below 0.10, which means they are in low agreement, and no significant difference in agreement

[2]Available at http://depts.washington.edu/acelab/proj/dollar/agate.html

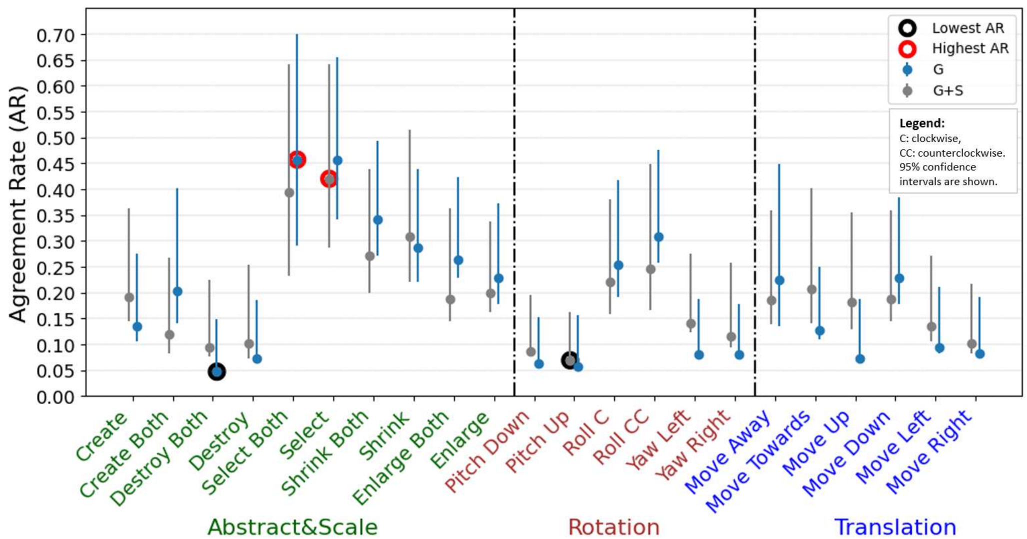


Figure 3.3: Agreement rates for the gesture (G) and gesture+speech (G+S) conditions grouped by referent category

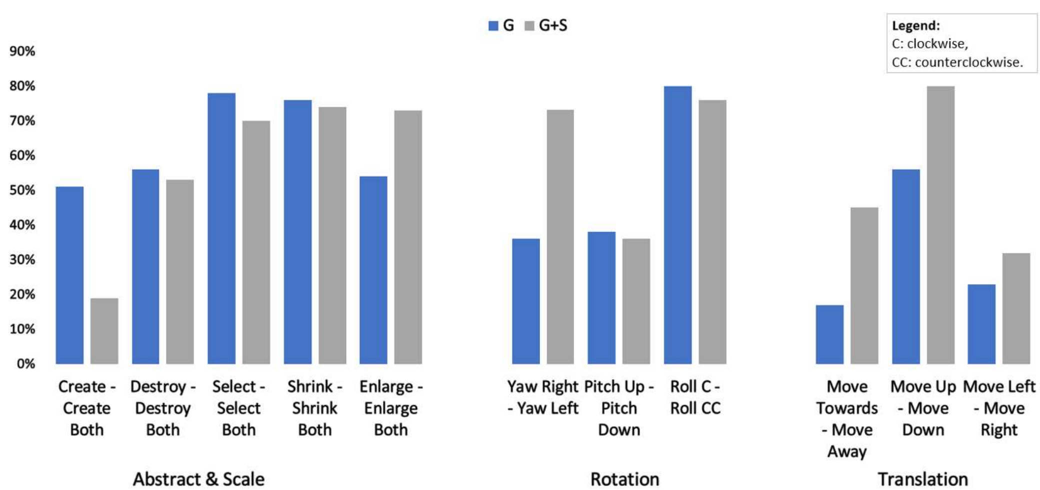


Figure 3.4: Gestures co-agreement between referents in gesture only (G) block and gesture+speech (G+S) block

Table 3.1: Consensus-distinct ratio (CDR) of speech only (S) and Gesture and Speech (GS) block by referent category

Referent Category	Speech Only	Gesture + Speech
Abstract	43.75%	35.21%
Rotation	24.56%	16.44%
Scale	57.14%	32.0%
Translation	42.65%	22.89%

rates was found ($V_{rd(8,N=216)} = 7.550$, $p < 1.000$). The highest agreement rates came from referents Select and Select Both (Figure 3.3). **The pointing gesture won the highest agreement rate for Select referent, mostly based on the natural interaction for specifying an object in the real world.** As shown in Figure 3.3, the referents Shrink Both and Roll Counter Clockwise (RCC) are also achieved high agreements. Among abstract referents, Destroy and Destroy Both got the two lowest agreement rates. In rotation referents, Pitch up and Pitch down exhibited the two lowest agreement rates. For the translation referent, referent Move Up has the lowest agreement rate ($\mathscr{AR}_{MoveUp} = .072$), yet Move Down has a much higher agreement rate ($\mathscr{AR}_{MoveDown} = .228$).

A co-agreement analysis for dichotomous referents and one object versus two object referents is shown in Figure 3.4. The co-agreement rates of one object versus two object referents were in general higher than in dichotomous referents. The referents Select and Select Both achieved a high co-agreement ($\mathscr{AR}_{Select} = .457$, $\mathscr{AR}_{SelectBoth} = .457$, $\mathscr{CR} = .355$), which indicates 78% of all pairs of participants have consistent gesture preference with both referents. Another high co-agreement rate came from Shrink and Shrink Both which showed 76% of all pairs of participants that were in agreement with referent Shrink were also in agreement with gestures for referent Shrink Both ($\mathscr{AR}_{Shrink} = .286$, $\mathscr{AR}_{ShrinkBoth} = .341$, $\mathscr{CR} = .217$).

Speech Only For speech data, we adopted the binning criterion wherein “enlarge yellow and green” and “enlarge cube and sphere” were equal to “enlarge yellow cube and green sphere”. However, “yellow cube pitch down” and “yellow cube rotate down” were counted as different proposals.

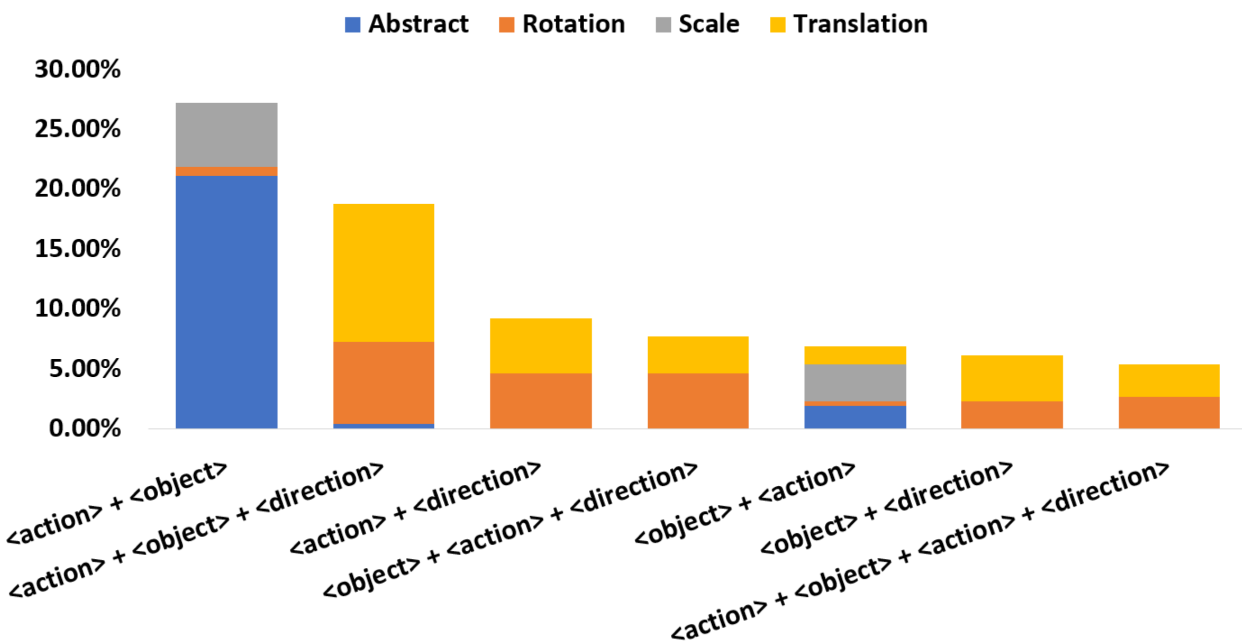


Figure 3.5: Usage of syntax format by referent type in the speech only (S) block

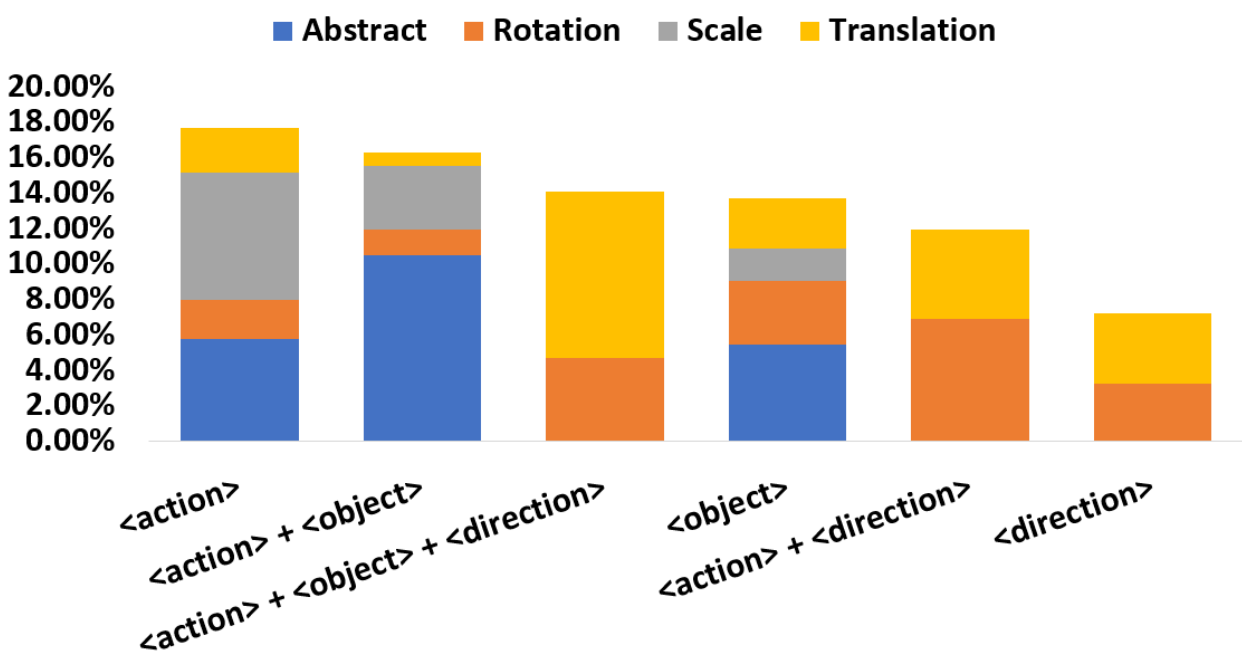


Figure 3.6: Usage of syntax format by referent type in the gesture and speech (GS) block

Table 3.1 shows the consensus-distinct ratio (*CDR*) of different categories of referents in the S block. The *CDR* is used to calculate the percent of distinct speech proposals by referent that achieved a consensus threshold of two [141]. The results demonstrated that scale referents have the highest *CDR*, in addition to abstract referents which present a *CDR* that are almost twice high when compared to 24.52% from the previous elicitation study with a single 3D object [145]. The rotation referents hold the lowest *CDR*. Based on the data, a low *CDR* could be caused by different expressions of rotation. For example, "spin" or "rotate" plus gesturing direction was proposed to achieve "roll", "yaw", or "pitch". A similar finding was presented in the previous elicitation study with a single 3D object [145]. There are few alternative phrases for "move up / down / left / right, " which could explain that translation referents have a higher *CDR* than rotation referents. Similarly, less options of replacement for action or status phrases such as "shrink" and "smaller" in scale proposals. Figure 3.5 presents the syntax formats that covered more than 80% of proposals in the S block. It is obvious that $\langle action \rangle$ $\langle object \rangle$ and $\langle action \rangle$ $\langle object \rangle$ $\langle direction \rangle$ are the most common formats for speech proposals. Moreover, rotation and translation referents elicited more variants of syntax, which means that various syntax should be considered while designing unimodal speech commands.

Despite bias from text referents, participants often preferred interaction from left to right with multiple 3D objects. For example, with the referents of "create two objects at the same time", two participants proposed "create green sphere and yellow cube", even though all text referents involving two objects started as "yellow cube and green sphere" in the experiment. **It shows that participants favored creating objects starting from the left since the green sphere was placed to the left side of the yellow cube in the scene.**

Multimodal interaction: Speech and Gesture

Gesture in GS – The results additionally demonstrate that the referent type has significant effect on gesture agreement rates in GS block ($V_{rd(21,N=528)}$ = 361.624, p <.001). There were 19 referents who achieved medium to high agreement ($\mathcal{AR} > 0.10$), and presented significant difference between agreement rates ($V_{rd(18,N=456)}$ = 262.325, p <.001). Only referents Destroy

Both, Pitch Up, and Pitch Down have low agreement rates that less than 0.10 (Figure 3.3), and further significant differences among those agreement rates were not found ($V_{rd(2,N=72)} = 1.368$, $p < 1.000$). As shown in Figure 3.3, the highest agreement rate in the GS block was from referent Select, and referent Select Both was not far behind in rank. The lowest agreement rates in the GS block came from referents Pitch Up, and Pitch Down was very close by. As in the G block, referent Shrink also obtained a high agreement rate while combining with speech ($\mathscr{AR}_{Shrink} = 0.308$). Moreover, referents Destroy and Destroy Both showed a similar low agreement as in the G block, compared to mostly other referents ($\mathscr{AR}_{Destroy} = 0.101$, $\mathscr{AR}_{DestroyBoth} = 0.094$).

Regarding co-agreement, for one object versus two object referents, the average co-agreement rate was 68% without including referent Create Both. **This finding indicates that a high number of participants kept the same preferences for both one object and two object manipulation.** The cause of a low co-agreement rate between referent Create and Create Both could be the low agreement rate for Create Both in the GS block ($\mathscr{AR}_{Create} = .192$, $\mathscr{AR}_{CreateBoth} = .120$, $\mathscr{CR} = .036$). Higher co-agreement rates were found for dichotomous referents compared to the G block. As shown in Figure 3.4, the co-agreement rates of translation referents were increased the most compared to the values in the G block, which indicates multimodal interaction assisted participants in achieving more agreement for dichotomous translation referents.

Speech in GS – Figure 3.6 shows syntax formats covered more than 80% of proposals in the GS block. Compared to the S block, participants have proposed a fair amount of single-word commands with compensation from gestures. These single-word proposals included $\langle action \rangle$ only, $\langle object \rangle$ only, $\langle direction \rangle$ only, and $\langle status \rangle$ only. All proposals with single-word commands account for 39.71% of the total proposals in the GS block. In contrast, the proportion of single-word proposals in the S block were merely 6.48%. The prior study mentioned that part of the $\langle action \rangle$ $\langle object \rangle$ syntax proposed in the S block turned into $\langle action \rangle$ plus gesture proposals in the GS block [145]. In the GS block, the most used syntax format was $\langle action \rangle$ only, shown in all four categories of referents. If the speech does not indicate the target, it can then be assumed that gestures were used for identifying the target object in a multi-object AR environment. As anticipated, based on the results, 87.5% of proposals with $\langle action \rangle$ only syntax format have involved gestures of "pointing", "tapping", or "grabbing". Furthermore, with $\langle direction \rangle$ only syntax proposals, 84.21% of gestures showed "pointing" or "tapping" to indicate the target object. In contrast, proposals consisting of the $\langle object \rangle$ only syntax format had merely 28.95% of the proposals involving the gestures "pointing" or "grabbing". This result proved the complementary feature of multimodal interaction. Due to the necessity of identifying the target object in a multi-object environment for manipulation and the flexibility of using speech that multimodal interaction gave participants, less agreement was shown with speech proposals in the GS block compared to in the S block (Table 3.1).

Gesture and Speech Association – The study looked into the association between the stroke of a gesture proposal and the corresponding speech proposal in the GS block. A stroke is considered the peak of effort for a specific gesture [162], which holds the meaningful content of the gesture. We classified the main speech content into three types of expressions (nominal, deictic, verb) based on prior work from Bourguet and Ando [163]. During the video annotation, recordings were made of the expressions in relation to speech content while the main stroke of the gesture occurred. The study found that strokes for abstract referents were mainly associated with nominal expressions, such as "the yellow cube" or "objects". The referents Destroy and Destroy Both were exceptions, which could be related to the low agreement on gesture proposals. All scale strokes were more synchronized with verb expressions, mostly "enlarge" and "shrink". It should be noted that there were far fewer deictic expressions used in the scale speech proposals, which indicates the limitation of associated expressions. In terms of the translation and rotation referents, 9 out of 12 showed a strong association between strokes and deictic expressions. For example, participants would execute the stoke of pitch up while saying "up". The Move Away and Yaw Right referents were slightly more synchronized with verb expressions, and the stroke of roll counterclockwise showed more association with nominal expressions.

Referent	Category	Winning Gesture	Second Place	Third Place	Coverage
Create	Abstract	bloom gesture (palm face forward) / point finger	tap finger	zoom out	77.09%
Create Both	Abstract	bloom gesture (palm face forward)	point finger	tap finger/bloom gesture (palm face up)	72.34%
Destroy Both	Abstract	squish with fist	point finger	zoom in	43.75%
Destroy	Abstract	point finger	squish with fist / toss hand forward	swipe finger to corner	56.25%
Select	Abstract	point finger	tap finger	open hand (palm face forward)	93.75%
Select Both	Abstract	point finger	tap finger	open hand (palm face forward)	85.41%

Figure 3.7: Top three proposed gesture variants for abstract referent

Referent	Category	Winning Gesture	Second Place	Third Place	Coverage
Enlarge Both	Scale	zoom out	bloom gesture (palm face forward)	bloom gesture (palm face forward) while move hand backward	79.16%
Enlarge	Scale	zoom out	bloom gesture (palm face forward)	extend hands distance diagonally (open hand or pinch gesture)	81.25%
Shrink Both	Scale	zoom in	gather all fingertips	reduce hands distance diagonally (open hand or pinch gesture)	89.59%
Shrink	Scale	zoom in	gather all fingertips	reduce hands distance diagonally (open hand or pinch gesture)	87.50%

Figure 3.8: Top three proposed gesture variants for scale referent

Referent	Category	Winning Gesture	Second Place	Third Place	Coverage
Move Towards	Translation	pull open hand	pull pinch gesture	pull finger(s)	58.33%
Move Away	Translation	push open hand	push finger(s)	point finger	64.59%
Move Down	Translation	swipe finger(s) down	push open hand down	grasp and move hand down	74.99%
Move Up	Translation	swipe finger(s) up	push open hand up	grasp and move hand up	58.33%
Move Left	Translation	pinch and move to top then left	point and move to top then left	grasp and move top then left	58.34%
Move Right	Translation	point and move to top then right	grasp and move top then right	pinch and move top then right	52.09%

Figure 3.9: Top three proposed gesture variants for translation referent

Referent	Category	Winning Gesture	Second Place	Third Place	Coverage
Yaw Left	Rotation	swipe finger or open hand to left / grasp and rotate hand around Z	flick finger to left	flick hand to left	70.83%
Yaw Right	Rotation	swipe open hand to right	grasp and rotate hand around Z / flick finger to right	swipe finger to right	58.33%
Pitch Down	Rotation	flick hand down	pinch and move hand down / flick finger down	swipe finger(s) down/ grasp and move down / pinch and rotate fingers around Y	72.93%
Pitch Up	Rotation	flick finger up	pinch and move hand up	grasping hand move up	41.67%
Roll Clockwise	Rotation	grasp and rotate hand around X	point and rotate fingers around X	rotating finger(s) around X	83.33%
Roll Counterclockwise	Rotation	grasp and rotate hand around X	point / pinch and rotate around X	rotating open hand around X	91.66%

Figure 3.10: Top three proposed gesture variants for rotation referent

3.1.6 Discussion

The results support our hypothesis H_2 that the consensus set of gestures indicates that participants preferred to use both hands for two virtual object manipulation. Chi-square analysis showed that the difference between one hand and two hands adoption in the two AR environments was statistically significant (X^2 = 255.33, p <.001). **This means that using a multi-object environment increased the usage of both hands.** The results support H_1 for some tasks, because there were 11 out of 17 single object manipulation referents resulting in different winning gestures, compared to ones from a single object AR environment. In general, there are similarities and dissimilarities in multimodal interaction within a multi-object AR environment and a single-object AR environment. **The multi-object environment has inspired more physical interaction which came from experience with the real world.**

Based on our resulting top three proposal variants for each referent, among 17 single object manipulation tasks, six referents have the same winning gestures as prior findings in a single object AR environment [145, 149]. Another five referents' second place proposals were identical to previous results in a single object AR environment [145, 149]. Within the six referents that have the same winning gestures in both AR environments, two of them are Shrink and Enlarge from scale referents, three translation referents are Move Toward, Move Away, and Move Left, and one rotation referent is Yaw Left. Legacy bias is an issue in elicitation study that uses' gesture proposals are biased due to the previous experience with existing interfaces [164]. The legacy bias from interaction with a multi-touch screen could contribute to the identical scaling proposals. The "screwing in a light bulb" gesture for rotating around the Z axis was also found in Williams et al., and Pham et al.'s works [145, 149]. Unsurprisingly, abstract referents have less similarity in their proposals. The winning gesture for creating an object in this work used gathered and then spread fingertips as the original blooming gesture from HoloLens 1 [165], but with the palm facing forward instead of facing up. Due to the difference, we do not consider that our blooming gesture came from legacy bias and more likely was a spontaneous proposal that could inspire future gesture designing for AR interaction.

Besides using the agreement rate to measure the consensus among gesture proposals and facilitate gesture sets generation, other criteria such as reversibility can also be considered validation after gesture sets are generated [166]. As shown in the gesture figures (Figure 3.7, Figure 3.8, Figure 3.9, Figure 3.10), all winning gestures for opposite scale referents and rotation referents met the reversibility criteria, then winning gestures for opposite translation referents were mostly reversible except for Move Left and Move Right. The winning gestures for opposite abstract referents were not fully reversible, which might be due to the difficulty of suggesting gestures for abstract interaction. Nevertheless, winning gestures for Create Both and Destroy Both perfectly met the reversibility criteria.

This paper has eliminated the presentations of variations of similar hand poses, which could be due to the physical features or object sizes or individual differences between users [167]. Through the binning procedure, most interchangeable variants of a hand pose with the same movement path are grouped into the same gesture. For example, pinching with two and three fingers is all grouped into the same pose. And if they all move clockwise, they are binned into the same gesture class.

In terms of speech proposals, our results showed more variety in syntax formats. We have two more single-word syntax formats in GS block compared to the ones found in Williams et al. [147], and they were $\langle object \rangle$ only and $\langle status \rangle$ only. For speech-only interaction, our study presented $\langle action \rangle$ + $\langle object \rangle$ as the top rank syntax format, compared to the $\langle action \rangle$ only syntax format which has a similar proportion in a single object environment [147]. We believe this result was due to the multiple object environment in our study, and participants tended to specify the target object for interaction.

The results of the study found that participants preferred symmetric bimanual versions of the single-handed gesture for two object manipulation. For example, the winning proposal for shrinking a single object was the zoom-in gesture, and the winning gesture for shrinking two objects side by side was to perform zoom-in with both hands simultaneously. This result of symmetric bimanual interaction is reasonable since both targets were inside the participant's field of view, which made symmetric actions easy to perform [168]. The exception of destroying proposals could be related to the low agreement rate for both destroy referents, which indicates people have less grounding for destroying from reality-based interaction [169]. According to the answers in the post-study questionnaire, 13 out of 24 participants expressed that it was fairly natural to think of using both hands for two object manipulation. Five participants indicated it was harder to develop the proposal for two object interaction compared to the single object manipulation. One participant said that the single-hand gesture could be used to replace two-hand interaction as needed. Our findings could be used to develop gesture recognizers for a multi-object AR environment by sensing the user's intent based on the hands involved.

Speech recognition with an AR headset is difficult due to the environment noise, unintended commands, and sometimes the accent of the user. With the knowledge of the association between speech expressions and gesture stroke, a more specific hypothesis can be implemented in the recognition system to improve speech detection efficiency and accuracy in AR. While previous work only focused on pointing gestures [163], our study discovered the association between common manipulation gestures and speech commands for interaction in a multi-object AR environment.

3.1.7 Design Guidelines

Based on the user-defined gesture sets from our study and literature, while some gestures and speech syntax formats remain similar, there were differences in multimodal interaction between a single object and a multi-object AR environments. Participants' proposals in our study showed more physical interactions such as pinching or grasping the target object and "turning a doorknob" for rotation tasks. Similar to prior findings suggested including aliasing for gestures and speech [138, 141, 145], we propose that including aliasing could significantly improve the performance of the recognizer. For example, using the commands "spin" or "rotate" plus a gesture that indicates direction should be equal to using commands "roll/yaw/pitch". With gestures, performing pinching or grasping and then moving the hand for virtual object translation should be equivalent to pointing at the target and then moving the finger. Our results indicate that implementing the top three proposed variants (Figure 3.7, Figure 3.8, Figure 3.9, Figure 3.10) of a gesture could increase the coverage of proposed gestures to 70% on average. The variety of syntax formats in the GS block indicates that various combinations of speech and gesture could be designed for interaction in an augmented reality environment. For example, to perform an interaction with a virtual object such as "move the clock to left", the system provides options that the user can either say "clock" plus move a finger to the left, or use a finger to point at the clock plus say "(go) left". Moreover, as Williams and Ortega mentioned in their work, legacy bias could be a benefit to new technology because it is memorable and discoverable [170]. We suggest that emerging technology such as AR-HMD should consider both legacy bias from the touchscreen and physical interaction based on body awareness and environmental skills [169].

3.1.8 Limitations and Future Work

The text referents could bias participants' speech proposals in our experiment. We also know that using animations as referents would bias the gesture proposal in the elicitation study [145, 171]. It is still a research question that how to eliminate the bias from referent presentation. Our experiment design requires participants to give both speech and gesture in GS block, which could end with unnatural speech proposals from participants. Therefore, we will use a more efficient but flexible way to elicit proposals from participants in our future elicitation studies. For example, we could adopt the "before" and "after" approach to present the desired effect of a referent for our future study [149, 172]. Reducing the fatigue caused by mid-air interactions is another necessary vein of future work. One way to mitigate this issue is to use other modalities such as eye-gazing combined with speech to replace mid-air gestures. Another option for reducing fatigue could be developing microgestures that require less psychical effort than mid-air gestures.

This study has focused on interaction in an AR environment, still, the elicited speech proposals can also be used for interaction in a VR environment due to the more occlusive VR environment. In terms of hand gesture proposals, further evaluation and comparison studies are necessary since gesture usability can be affected by the field of view of the HMD and the clutter of the scene.

This work used two object manipulation as multi-object interaction cases. Future works are expected to verify the current elicited proposals with referents that involve three or more object manipulation. Also, further studies can answer the questions of how more virtual object manipulation influences the speech and hand gesture proposals for interaction in AR.

3.1.9 Conclusion

This study investigated multimodal interaction in a multi-object AR environment. We chose 22 referents for the elicitation study that included canonical referents for scale, translation, and rotation tasks and three abstract referents. We generated a consensus set of gestures for interaction in a multi-object AR environment and found that participants used the same gesture for one and two objects but with both hands for two object manipulation. The results further demonstrated that participants tended to act on the target objects in a multi-object AR environment, indicating more physical interaction where preferred. Further, in the study, more speech syntax formats were proposed in multimodal interaction in a multi-object AR environment. We discovered the association between expressions and stroke, which can improve the accuracy and efficiency of the recognition system. We also provided design guidelines based on our findings and comparison with prior works in a simple AR environment.

3.2 Towards Establishing Consistent Proposal Binning Methods for Unimodal and Multimodal Interaction Elicitation Studies[3]

3.2.1 Overview

More than two hundred papers on elicitation studies have been published in the last ten years. These works are mainly focused on generating user-defined gesture sets and discovering natural feeling multimodal interaction techniques with virtual objects. Few papers have discussed binning the elicited interaction proposals after data collection. Binning is a process of grouping the entire set of user-generated interaction proposals based on similarity criteria. The binned set of proposals is then analyzed to produce a consensus set, which results in the user-defined interaction set. This paper presents a formula to use when deciding how to bin interaction proposals, thus helping to establish a more consistent binning procedure. This work can provide human-computer interaction (HCI) researchers with the guidance they need for interaction elicitation data processing, which is largely missing from current elicitation study literature. Using this approach will improve the efficiency and effectiveness of the binning process, increase the reliability of user-defined interaction sets, and most importantly, improve the replicability of elicitation studies.

3.2.2 Introduction

Elicitation is a form of participatory design where researchers ask novice users to produce input proposals that would execute various commands (i.e., referents) within a system (e.g., augmented reality displays, multi-touch surfaces) [138, 173]. During this process the participant is shown referents one at a time and the participant produces input proposals that they find appropriate for executing those referents. This might look like showing a participant a virtual object using

[3]Verbatim From Zhou, X., Williams, A.S., Ortega, F.R. (2022). "Towards Establishing Consistent Proposal Binning Methods for Unimodal and Multimodal Interaction Elicitation Studies". *In: Kurosu, M. (eds) Human-Computer Interaction. Theoretical Approaches and Design Methods. HCII 2022.. Lecture Notes in Computer Science, vol 13302. Springer, Cham. https://doi.org/10.1007/978-3-031-05311-5_25*

augmented reality technologies and asking them to produce a mid-air gesture that would cause the virtual object to move left [145, 154].

During an elicitation study researchers often identify the consensus set of input proposals by quantifying the agreement between participants on input proposals for a given referent [138, 160]. In between labeling the raw input proposals and measuring participant agreement, researchers often bin the input proposals into equivalence groups based on their similarity of execution. Without binning proposals, the consensus set will be biased by individual differences and agreement calculations will return pessimistic estimates of agreement. As an example, without binning, a swipe right with the index and middle fingers would be counted as a distinct gesture from swiping right with the index alone. In actuality these two gestures are similar in execution, intent, and at times may have been produced by the same user for the same referent [154, 155].

It is well known that the binning procedure is time-consuming due to the process of annotating the recorded videos [148, 174], and error-prone because of the subjective human judgments involved [148]. Moreover, adopting different similarity criteria can lead to underestimated or overestimated agreement rates as a result of more or less permissive binning protocols [148]. In this paper, we present a formula that can assist experimenters with binning gesture proposals from participants through a unified binning procedure. This work uses mid-air gesture elicitation proposals, which can be extracted from both unimodal and multimodal interactions. This formula could be used equally well to bin gesture proposals from elicitation studies with different systems, such as virtual reality displays or multi-touch devices.

It is critical that experimental science is reproducible, replicable, and reapeatable. Researchers in HCI have been calling for more work focusing on the replication or reproduction of prior works [148, 175]. In particular, for elicitation studies, adopting a standard data processing procedure and guidelines can accelerate the replication or reproduction process for future researchers. The necessity of a standard procedure for binning is increasing along with the rapid growth of elicitation studies. It is time to fill this gap and form a guideline for elicitation data processing.

3.2.3 Related Work

Based on recently published elicitation study literature reviews, over two hundred elicitation studies have been conducted in the last ten years [176]. Most of the previous works aimed to develop user-defined gesture sets [138, 141, 145, 154, 155, 177–180], and a few works have discussed the tools and criteria used for binning the interaction proposals in the studies [174, 181, 182].

Binning Tools

Several studies proposed new approaches to improve the efficiency of the binning process and reduce the effort required by researchers. Ali et al. proved that letting online crowd workers make the proposal's similarity comparison, then using automatic binning algorithms for sign classification, can produce similar results compared to using experts with less time spent on the task [174]. Nevertheless, the crowd-powered tool Crowdsensus proposed by Ali et al. only demonstrated the use of elicited voice commands in text string format. Vatavu introduced the dissimilarity-consensus method to replace the subjective binning criteria, and manual labeling with automated computation by a computer [181]. This method helps mitigate human error and subjective decisions on proposals similarity. However, using the dissimilarity consensus approach accurate skeletal and pose information is needed, which is not suitable for most elicitation studies' data captured in video format. Moreover, despite the computational method, subjectivity still exists during raw data preprocessing and extracting procedures. Anthony et al. developed a tool to identify similar gesture articulations and generate clusters automatically, and the agreement rates were reported after binning [182]. Nevertheless, the tool allows the researcher to edit and correct the cluster partition, making the binning process still subjective. While it is difficult to entirely avoid the subjective human judgments during the binning process, following the guidelines put forth here can help to mitigate it.

Binning Criteria

It is not uncommon that different elicitation studies adopt different criteria for evaluating the similarity of proposed interactions from participants [148, 181]. Few works have reported the criteria they used for proposal binning in the study, which led to irreproducible study results [148]. Most of the elicitation studies have described the classification process in a broad manner [138,150, 155]. Several works in augmented reality have mentioned the general binning approach or criteria was used in their studies [145, 149, 154]. Piumsomboon et al. defined "similar gestures" as the gestures that have identical static pose and path, or have different static poses but with consistent directionality [154]. Pham et al. adopted themes such as the analytic unit and binned gestures based on similar themes despite different executions (e.g. binned squishing with clapped hands, fingers, or fist) [149]. Williams et al. binned gestures according to the movement path and hand pose, and explained the definition of each hand pose presented in the study [145]. Many elicitation studies have used a taxonomy for gesture classification and binning [154, 155, 183–186].

Taxonomy based binning criteria are not always well suited for direct interaction elicitation study proposals. In most elicitation studies a large proportion of the referents revolve around direct manipulations (i.e., translations, rotations, and scaling). When eliciting direct manipulations, interaction proposals will often be impacted by the participants' notions of real world physics [169]. If the referent is *move left* the participants will most likely push the object to the left with their hand [149, 154, 155]. These proposals will often fall under a single category when using a taxonomy [187] (e.g., "direct manipulation"), making using taxonomy for binning difficult. This paper will not discuss the procedure of classifying mid-air gesture proposals based on taxonomy. Our study aims to provide researchers and designers with a guideline for executing the proposals binning procedure in an elicitation study, as shown in Figure 3.11. With the structured criteria and procedure, more elicitation studies can be replicated, and the results can be reproduced.

3.2.4 Data Collection

Often elicitation studies' data is captured via video recordings [141, 145, 147, 148]. An exo-centric camera is commonly used to record participants' proposals. If the study used head-mounted displays (HMDs), an ego-centric camera can also be used. For data completion and analysis convenience, recommend experimenters adopt both ego-centric and exo-centric cameras to cap-

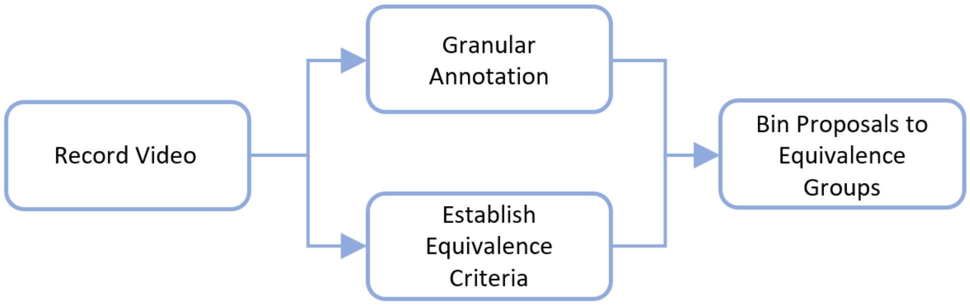


Figure 3.11: Procedure to Group Interaction Proposals Into Equivalence Groups in an Elicitation Study

ture different views of interaction. Based on observations from our prior experiments' data with HMDs, participants have various preferences on hand position while performing manipulation tasks [145, 147]. Some participants proposed physical interaction more than others, which means the gesture acted physically on the object, and their hands reached out to the objects. When participants proposed more metaphorical gestures for interaction, their hands were closer to their bodies during the manipulation. As a result, it is possible that the ego-centric camera is unable to capture the full view of the interaction due to the different positions of individual proposals. The exo-centric view could be very helpful, when the ego-centric camera does not capture the full hand movement. For example, while participants' hands are too close to the face and the movement is out of the view of the ego-centric camera, the exo-centric recording became the only source of participants' proposal. For exo-centric view, it is suggested to place the camera with a tripod on the front right or left corner that faces the participant. For the ego-centric view, the camera (e.g. GoPro) can be mounted on top of the head-mounted display (e.g. HoloLens 2). It should be noted that experimenters need to check the camera angle before starting the experiment. It is not unusual that the head-mounted display is unable to hold the camera on the top, or the camera mounting cause unbalance or discomfort to users. In that case, the experimenters could choose to use a head strap attaches to the camera and ask participants to wear it on their heads. The downside of this approach is the complexity of the camera angle set-up, since the head-mounted display could block the view of the ego-centric camera.

3.2.5 Video Annotation Procedure

Once the elicitation data has been collected a granular labeling pass should be done to convert the raw video to meaningful interaction information (Figure 3.11). These labels should capture all relevant information about the gestures. For each interaction elicited, the hand used, fingers used, direction of movement, start time, stop time, changes in hand shape or pose, and the corresponding referent name should be recorded. Some additional interaction features that can be recorded depending on the studies' aims are outlined in the following sections.

Stroke Segmentation

A gesture generally starts with a preparation phase, moves into a stroke, then returns to the resting position [188]. For example, the hand was usually placed on the lap of the participant or on the desk surface with a relaxed hand pose. When the participant is ready to perform the gesture, the hand will hold up and leave the lap or desk first, then reach out to the target object or perform the interaction immediately. This means that experimenters need to find the boundaries of a gesture, and it also can be called the stroke of a gesture. According to McNeil, a stroke is considered the peak of effort for a specific gesture [162], which holds the meaningful content of the gesture. Based on previous experiment data, when the stroke happens, the direction of movement will change, or the speed of movement will show a sudden increase [147]. This can be observed while the experimenters go through each recorded video frame.

Gesture Information Recording

After the boundaries of the gesture are found, the video of the stroke can be segmented, or the screenshot of the stroke can be captured for later reference. This reference could be very useful while experimenters need to review or double-check certain proposed gestures. The start and end times of both gestures and strokes should be noted during video annotation for later analysis of stroke time distribution, or the relationship between strokes and other modalities of interaction. In addition, the experimenters should also record which hand and finger have been used for the interaction, how many hands were used, and what fingers were involved. These elements contain important information for verifying results from previous studies with mid-air gestures in augmented reality environment [145, 147, 149, 154].

Stroke Coding

To code proposed gestures, experimenters need to define each hand shape before starting coding. For example, the "pointing" gesture could involve the index finger, index middle fingers, or all fingers from one hand. "Pinching" could include thumb and index finger, or thumb index and middle fingers. "Grasping" could be gathered five fingertips, and "grabbing" shows as a fist. During stroke coding, the hand pose and the path of hand movement are the main points that experimenters should pay attention to. In terms of hand pose, hand shape and palm and finger orientation should be noted. For example, moving an object up with the thumb and index finger holding the object should be coded as pinching with the palm facing forward and moving the hand along the Z-axis. In terms of movement path, experimenters should code proposed gestures based on which axis it moved along or around. For example, translations up and down should be both considered as movements along the Z-axis. The referents of translations left or right should be identified as movements along Y-axis. Moving an object towards or away should be recognized as movements along X-axis. The rotations clockwise and counterclockwise should be both coded as movements around X-axis. Rotating around Y-axis could describe the response to referents *pitch up* or *pitch down*, and rotating around Z-axis could represent proposals to referents *yaw left* or *yaw right*.

It is worth noting that the thumb position could be less meaningful compared to other fingers. We suggested experimenters consider the interaction theme while recording the thumb position for each proposal. The variants of thumb position in hand poses could be due to hand differences or individual habits. In general, the relaxed position for the thumb is to stick out instead of closing to the palm or gathering with other fingers. As a result, when a gesture does not require the thumb, it normally stays relaxed, which could look like the thumb is sticking out. For example, based on our previous data, some proposals with index finger pointing gestures also had the thumb pointed to the front [147]. In this case, we ignored the thumb pose and only recorded the pointing gesture executed by the index finger. However, with proposals involving pinching or grasping poses, which could be proposed for referents of translation or rotation, the experimenters should pay attention to every fingers' status including the thumb.

3.2.6 Interaction Proposal Binning

Once all of the collected video has been annotated the elicited interactions can be binned into equivalence groups based on pre-defined metrics of similarity. This reduced the space of elicited proposals to a smaller set of binned interaction proposals which can then be used during agreement rate and consensus calculations. Conceptually, the binning process smooths out interpersonal differences between interaction proposals. As an example, if binning processes were omitted, a three finger pinching gesture where the participants index, middle, and thumb come together to grab a virtual object would be considered different from a two finger pinching gesture where only the index and thumb are used. When considering how to implement the elicited interaction proposals into a interaction system, it makes more sense to say that those two gestures are the same. Their intent and motion trajectory's are similar. Outside of motion and intent, this step can be further justified when considering how the same user may use inconsistent numbers of fingers for the same gesture [154], making the difference between using the index and thumb or the index, middle, and thumb trivial.

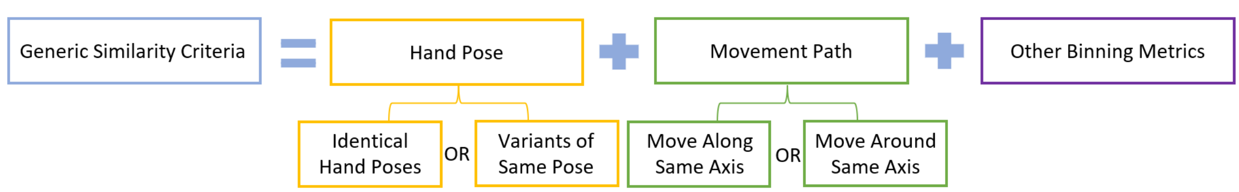


Figure 3.12: Formula for Binning Gesture Proposals

As shown in our formula (Figure 3.12), while binning mid-air gesture proposals, we suggest the experimenter mainly focuses on the similarity of static hand pose and hand movement path. Bins are created based on the pairing of binning criteria. This might look like making bins for pinching hand poses moving along the x-axis, or grasping gestures that move vertically. Some bins may not be predictable a priori. To account for that researchers can expect to add new binning criteria while analyzing their results. For example someone might pantomime usage of a steering wheel for rotations. This gesture might necessitate a new equivalence group to be established. The experimenters can set up a worksheet to assign a name or number and provide a brief description to each bin for later reference. It is possible that the same gesture was proposed for different referents, and we recommend sharing bins through the binning procedure for all referents in the study. If the same gesture resulted in the same agreement rate with different referents, the experimenters could use further analysis or use other criteria to decide which referent the gesture is best suited for.

Figure 3.13: Different Proposals of the Pinching Gesture

Figure 3.14: Interchangeable Variants for Pinching or Grasping Gestures

Hand Pose

The proposed gestures with identical or near identical hand poses and moving paths belong to the same bin (Figure 3.15). For example, both gestures pinch an object with thumb and index finger then move the hand along the Y-axis. However, some near identical hand poses could look a bit different from others due to individual preferences, as shown in Figure 3.13. For example, in prior work done by this lab, some participants proposed using blooming gestures with the palm face up to perform the create object task, and some participants opened their hands less wide than others [145, 147]. For referent *move away*, among participants who adopted using an index finger to point at the object and moving along the X-axis, few participants held their pointing finger with their palm facing left instead of facing frontwards.

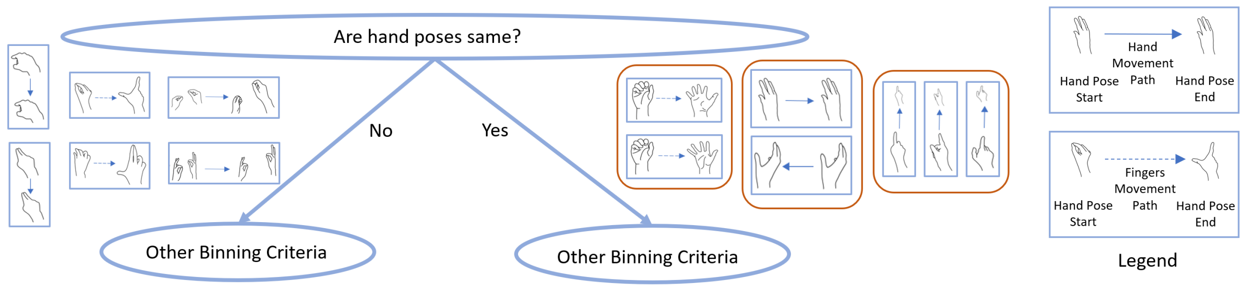


Figure 3.15: Binning Gesture Proposals with Identical Hand Poses Criteria

As previous works pointed out [147, 149, 154, 156], variations of similar hand poses were often elicited among participants. The variations could be due to the physical features or the size of the target object [149], and we are focusing on using the same objects for the entire elicitation study. According to prior studies, the variants of a single hand pose could come from the same participant or different participants [154], and the previous study has shown that participants preferred to perform the same gesture in a different manner [155]. In Piumsomboon et al.'s work on eliciting user gestures in augmented reality, variants of pinching and grasping gestures were similar and grouped into the same bin. The shapes or affordances of the objects used to elicit interactions can also impact the hand poses elicited. If the same object was used for all experiments, we suggested experimenters consider that variants of a single pose are often interchangeable, and that those variants should be grouped into the same bin if they have the same movement path (Figure 3.14).

The most common variations of hand poses that were mentioned in prior studies involved a different number of fingers when gesturing [145, 154, 155]. As an example, the proposals for referent *rotate clockwise* in previous elicitation studies were performed by pinching gestures with two or three fingers, or grasping gestures with five fingers [145, 147]. It was not due to different sizes or affordances of the object since the same object was presented to every participant. A similar case is that while participants performed *move away* with pointing fingers, some chose to use both their index and middle fingers pointing at the object compared to others who used

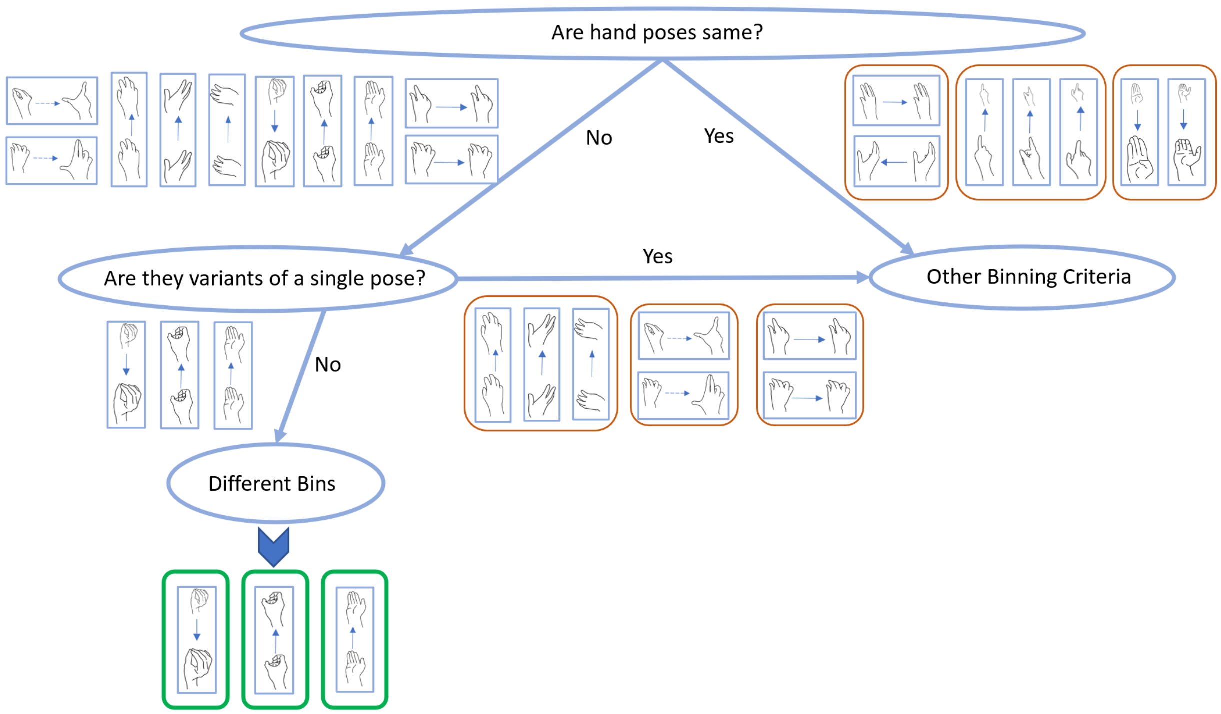


Figure 3.16: Binning Gesture Proposals with Identical Hand Poses and Pose Variants Criteria

only their index finger. Another example would be the same pinching gestures with thumb and index finger, but different positions of the other three fingers. Some participants kept their other three fingers spread instead of together while performing pinch gestures. Furthermore, based on our elicitation data, we found another type of variation included different hand orientations. One example is that for the referent *move up* in our study, most participants have proposed grasping the object and moving their hand straight up, and we found three variants of the same hand pose with different hand orientations. The participants were grasping from the front of the object with palm facing front and fingers pointing up, left, or right (Figure 3.16).

Movement Path

If the proposed gestures have identical or variant hand poses, experimenters can continue to check if the moving paths of proposals are the same before making the binning decision (Figure 3.17). Based on the stroke coding results, if the moving path is along or around the same axis, we can put those gestures in the same bin. For example, pinching with thumb and index finger

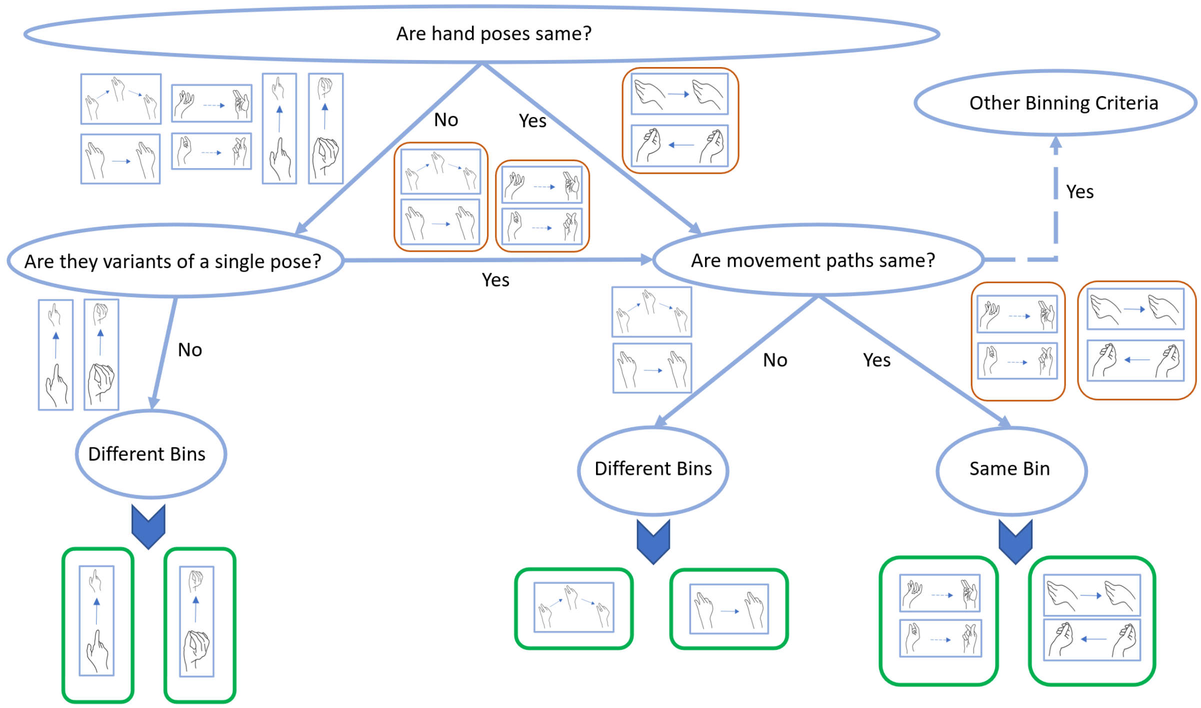


Figure 3.17: Binning Gesture Proposals with Identical Hand Poses, Pose Variants, and Movement Path Criteria

then moving hand along Y-axis can be grouped with grasping with five fingers then moving hand along Y-axis. However, pinching with a thumb and index finger then moving the hand along Y-axis should not be binned with grasping with five fingers and moving the hand along Z-axis then moving it along Y-axis. Not all gestures will move in a straight line or follow a directional axis. Some proposals for enlarging or shrinking objects, especially when the gestures involved both hands, were proposed with diagonal movement paths.

3.2.7 More than Mid-air Gestures

Our formula can be used equally well for elicitation studies outside augmented reality environments, because researchers can adopt our formula to bin gesture proposals from studies conducted with virtual reality headsets or multi-touch devices. For multi-touch devices, hand poses and movement path are still the crucial criteria for the binning procedure. The experimenters should reconsider the interchangeable variants of a single pose since the multi-touch devices could have a different definition for the number of fingers in a gesture counted as a single-point touch or a whole-hand touch. Based on Wobbrock et al.'s elicitation study on tabletop gestures [138], using 1-3 fingers should be considered as a single-point touch, and using four fingers or more could be considered as a whole-hand touch. It means that the experimenters can consider variants of touch gestures with 1-3 fingers as interchangeable during the binning process.

3.2.8 Limitation and Future Work

The formula presented in this paper focused on hand gesture proposals, and it is not suitable for other proposals generated by other body parts such as the head, foot, or arm. Since different body gestures have their features and limitations, it is difficult to use the same criteria for all elicitation studies' data processing. In future work, we are interested in perfecting the binning criteria for hand gesture proposals in elicitation studies. Another future direction regarding binning procedures is to explore further criteria for binning other body gestures and creating guidelines for researchers that can be used in different elicitation studies.

3.2.9 Conclusion

This paper proposed a formula that can help experimenters or researchers to bin gesture proposals in elicitation studies. In addition, we explained the necessity of a manual binning procedure and the formation of guidelines for elicitation data processing in HCI. To achieve this goal, this paper went through how experimenters might capture the elicitation data, annotate the recorded video, and what similarity criteria should be considered during the proposals binning process. We also presented real data from our previous elicitation study to demonstrate the possible variations of each gesture. We hope that this paper can be leveraged to create binning procedures for different categories of elicitation studies, such as studies involving gestures using different body parts or using different devices.

Chapter 4

Interaction and Navigation in AR vs. VR

Due to their characteristics, augmented reality and virtual reality provide different benefits and setbacks to user interaction and navigation. What are the factors that cause the differences? How do different realities influence the sensemaking process, and what strategies do users choose to interact and navigate in AR and VR? How can we utilize different realities to maximize users' efficiency during sensemaking? Most prior studies have explored sensemaking in AR or VR, and no comparison work has been done with data visualization sensemaking tasks with the AR and VR HMDs. This chapter reports on the design and implementation of a prototype immersive analytics platform that allows users to perform 3D data analysis with the provided analytic toolkit. With a controlled study with 40 participants, we present our observations and insights into how users interact and navigate in AR and VR and what factors future researchers should consider when choosing between AR and VR for immersive analytics.

The study is presented in this chapter as it was published [6]. This article is included with copyright holder permission.

I Did Not Notice: A Comparison of Immersive Analytics with Augmented and Virtual Reality.[4]

[4]Verbatim From Xiaoyan Zhou, Anil Ufuk Batmaz, Adam S. Williams, Dylan Schreiber, and Francisco Ortega. 2024. "I Did Not Notice: A Comparison of Immersive Analytics with Augmented and Virtual Reality.". *In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24),. May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3613905.3651085*

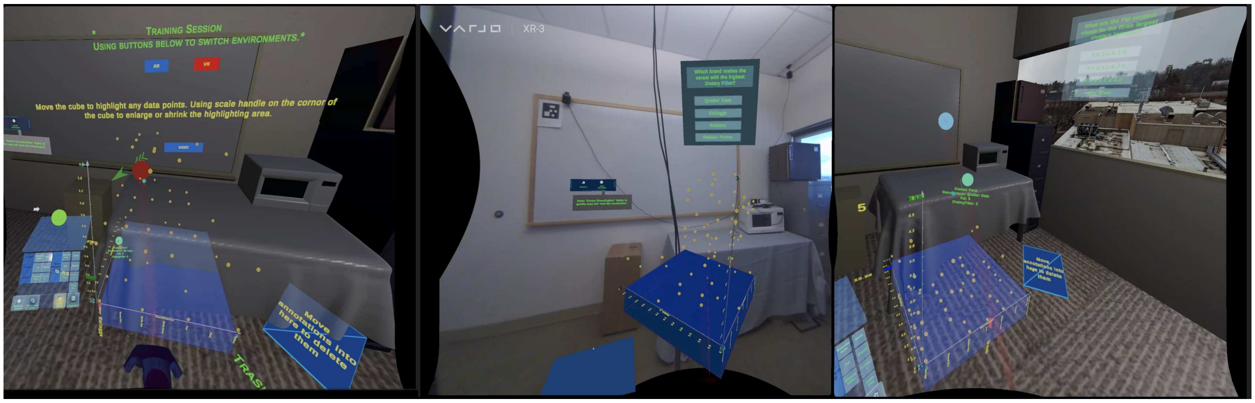


Figure 4.1: The interface displayed in the training session (left), AR session (middle), and VR session (right)

4.1 Overview

Immersive environments enable users to engage in embodied interaction, enhancing the sensemaking processes involved in completing tasks such as immersive analytics. Previous comparative studies on immersive analytics using augmented and virtual realities have revealed that users employ different strategies for data interpretation and text-based analytics depending on the environment. Our study seeks to investigate how augmented and virtual reality influences sensemaking processes in quantitative immersive analytics. Our results, derived from a diverse group of participants, indicate that users demonstrate comparable performance in both environments. However, it was observed that users exhibit a higher tolerance for cognitive load in VR and travel further in AR. Based on our findings, we recommend providing users with the option to switch between AR and VR, thereby enabling them to select an environment that aligns with their preferences and task requirements.

4.2 Introduction

Along with the rapid growth in the popularity of virtual reality (VR) and mixed reality (MR) headsets, many researchers are particularly interested in exploring the differences between augmented reality (AR) and VR Head Mounted Displays (HMDs) and how they can be leveraged to enhance user performance [3, 94, 189, 190]. While see-through AR provides users with an interactive and intuitive experience by seeing 3D objects and their own hands, VR offers users a fully immersive experience with clutter-free and infinite spaces. Previous studies have explored the effects of AR or VR environments on sensemaking strategies in non-quantitative immersive analytics [3] and their impact on the perception of 3D data visualization [94]. However, there is limited empirical evidence on how different realities influence 3D data visualization sensemaking, as well as their respective advantages and disadvantages. To address this gap, we developed a data analytics platform in both AR and VR with the same HMD, using identical objects and layouts. We then conducted a user study to measure performance and record feedback in each reality. To minimize confounding variables such as differing fields of view, we utilized a headset that supports displays for both AR and VR. Our results reveal that user performance, overall subjective workload, and presence experience are comparable between AR and VR. However, users reported distinct perceptions of frustration in AR versus VR; additionally, they displayed more varied postures and physical navigation in the VR environment. Based on our observations, we offer insights and guidelines for future researchers developing immersive analytics platforms to consider.

4.3 Related Work

4.3.1 Immersive Analysis

Immersive analytics extends the domains of data visualization, visual analytics, mixed reality, computer graphics, and human-computer interaction. Users engage in sensemaking within immersive environments using abstract data representations to facilitate a comprehensive understanding of data and decision-making processes for independent work as well as collaborative endeavors [27, 123, 191].

Recent studies have explored user performance in immersive analytics with different realities [88, 192] while also focusing on improving sense-making strategies to increase user performance and experience when engaging in immersive analysis tasks [193–196]. For example, several studies have shown that the use of spherical layouts for displaying multiple visualizations can enhance the sensemaking process and encourage greater user engagement [192, 194, 197, 198].

Although earlier work primarily focused on sensemaking with quantitative datasets [192,193], researchers suggested that non-quantitative data could also benefit from the significant space available in immersive environments for offloading cognition during a sensemaking process [199,200]. Our research focuses on immersive data visualization with a 3D scatter plot to explore variations in users' sensemaking strategies in AR and VR settings.

4.3.2 AR vs. VR

Several previous studies have compared AR and VR focused on specific aspects such as object manipulations [201,202], data visualization interpreting [94], depth estimations [203], eye-hand coordination [204], mode switching between display types [66], or different research methods [205]. Steffen et al. provided guidelines for developing specific affordances for better user experiences by comparing the affordances between AR, VR, and physical reality [189]. Williams et al. initiated discussions about how different levels of reality choices may impact user interactions while performing immersive analytics [190].

The most related study to this one came from Lisle et al. [3], who explored the sensemaking task in an immersive analytics context with AR and VR. They found that users can focus more on the task in VR than AR; however, with the ability to access real-world tools, user experience has improved during the task execution. Differently, our study focuses on immersive data visualization instead of text-based data. This could lead to different approaches to user navigation and sensemaking processes due to the spread of the data. Davidson et al. found that navigation patterns change according to the different stages of the sensemaking process [206].

Cross-virtuality analytics has become increasingly prominent in recent years, with the goal of enabling users to transition seamlessly between different levels of mixed reality. This approach aims to address the limitations of each technology and facilitate a combination of their advantages for enhanced analysis efficiency [118,134]. Our findings indicate that the choice of reality level depends on individual preferences and the complexity of the analysis task at hand. We offer insights and guidelines for future researchers looking to implement immersive analytics in cross-virtuality environments.

4.4 Study Design

We employed a within-design approach with two display types (DT_2 =VR, AR). Each participant completed two question sets (QS_2 = VR questions, AR Questions), 6 questions in AR and 6 in VR, with the same data set. We ensured an even distribution of questions across both conditions concerning question type and difficulty level. The order of display type plus question set was counterbalanced among participants to mitigate learning effects, as was the order of questions for each participant under the same condition. Besides objective metrics such as task completion time and accuracy of answers, we used subjective measures including VZ-2 paper folding test [207], NASA Task Load Index scale (NASA TLX) [208], Presence Questionnaire (PQ) [209], and System Usability Scale (SUS) [210]. Additionally, a semi-interview was conducted at the end of each experiment to gather more insights into user experience.

4.4.1 Experiment Setup

The study utilized a Varjo XR-3 headset [211], which is the best mixed reality headset that is available on the market. This headset operated on a desktop PC equipped with an Intel Core i9-11900F processor and an NVIDIA GeForce RTX 3090 graphics card. Additionally, participants used HTC Vive controllers [212] to manipulate virtual objects in both environments, with tracking performed by three HTC Vive Tracker 2.0 units [213]. Our immersive analytics platform was employed in both conditions, and a simulated office environment was presented in VR, as seen in Figure 5.1. To elicit natural interactions from participants, we included an office desk, corner chairs, a bookshelf, a microwave, and a mini refrigerator to replicate the real office setting. Furthermore, participants were provided with bar stools for easy body rotation as well as standing up or sitting down without safety concerns.

4.4.2 Procedure

Upon the participant's arrival, the experimenter provided a brief introduction and handed consent forms for signatures. Then, participants completed a demographic questionnaire and the VZ-2 paper folding test. Next, the experimenter introduced the study and experiment process before instructing participants to put on headsets and use controllers to interact with virtual objects. They then followed the instructions displayed to complete eye calibration, training sessions, as well as two experimental sessions. During the training session, the participants were guided through each function of the platform with detailed instructions shown on the HMD. During the experimental session, participants were asked to answer 6 questions in each environment with the same dataset. These questions included outlier detection requiring axis switching or identifying data at a fourth dimension, identification of summarization (mean or median) and data trend, along with cluster recognition. Following each experimental session, participants filled out NASA TLX, PQ, and SUS. A 3-minute break was implemented between conditions. Each experiment finished with a semi-structured interview. The entire study takes around 60 minutes.

4.4.3 Participants

40 participants (17 female, 20 male, 3 preferred not to specify their gender or identify as nonbinary) aged between 18 and 67 (average age of 29.45) volunteered for this experiment. Of the participants, 36 were right-handed while the rest were left-handed. Our participant pool comprised individuals from various fields: computer science, engineering, natural science, social science, business school, and others who were employees or did not specify a field of study (Figure 4.2).

4.4.4 Hypotheses

Based on the results from the preliminary study [214] and prior research comparing AR and VR [3, 94], we have developed three hypotheses for this study.

H1: Participants' performance in AR is higher than in VR. Our preliminary study showed that participants in VR performed better than those in AR. Apart from variables like different head

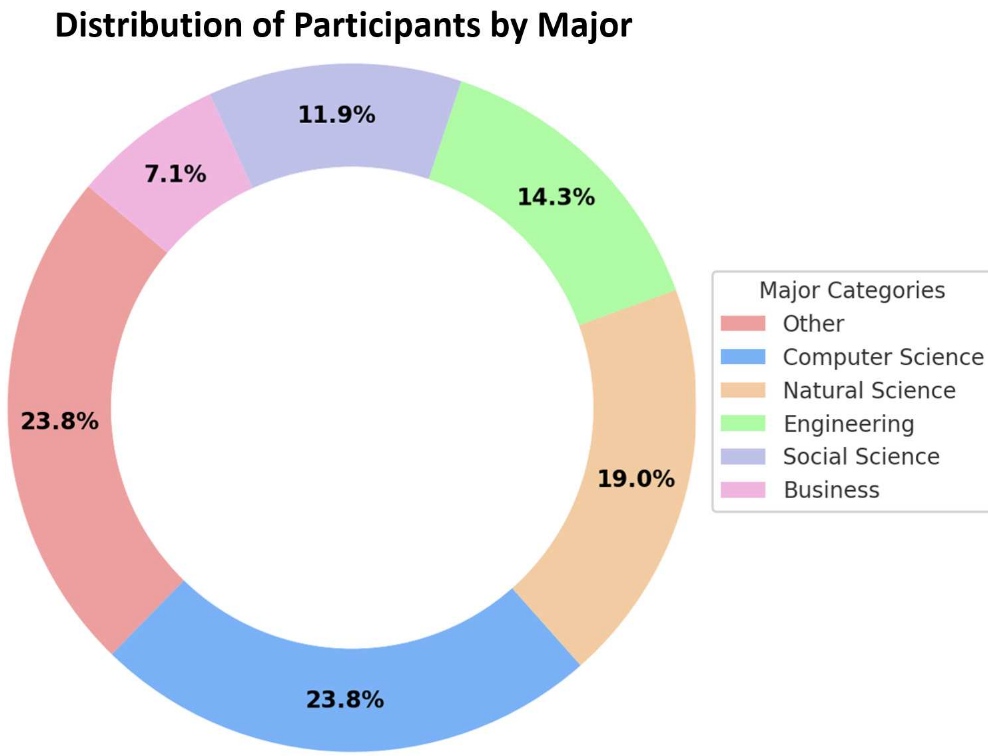


Figure 4.2: Participants Background in the Study

displays, we suspect that the presence of the experimenter and a cluttered room may influence performance, as suggested by a previous study [3,201]. Therefore, we eliminated unnecessary objects from the room and introduced a folding screen between the participants and the experimenter to see if these adjustments led to similar performances across both conditions.

H2: Participants spend more time in AR than in VR. The preliminary study's results indicated that some participants experienced frustration during the experiment. It could have been caused by the difficulty of using hand gestures of HoloLens 2, as noted in another comparison work between AR and VR head-mounted displays [215]. To address potential issues with hand gesture interactions, we provide controllers in both environments. Additionally, the visibility of the real world is anticipated to influence color perceptions among participants, possibly affecting their engagement with tasks in AR [94, 216]. It is also assumed that due to real-world distractions, participants will take longer to complete tasks in AR compared to VR.

H3: Participants in AR navigate through visualization physically more than VR. In the preliminary study, participants were given a chair equipped with a handle and asked to sit at a table before the experiment started. The results showed that only some VR participants exhibited upper body movement. Building on prior research, which suggests that with full environmental awareness, participants feel more comfortable moving around in AR than in VR [94]. By informing participants about the freedom of movement and implementing a bar stool setting in this study, we anticipated that greater physical movement in AR than in VR.

4.5 Results

After examining the correctness of each question in both conditions, we found that question 4 in the AR condition has a significantly lower correctness rate compared to other questions. It was also identified by the Interquartile Range method as an outlier, so we decided to remove question 4 from our dataset before proceeding further analysis. We performed a chi-square test to compare the proportion of correct responses between AR and VR conditions. It yielded a p-value of 0.405, suggesting no significant difference in correctness between the two display conditions. When comparing the time spent on questions between the two conditions, the Mann-Whitney U test indicated no significant difference in the time spent on answering questions across these two conditions (p-value = 0.745). These results indicate that hypothesis H1 is not supported. Moreover, there is no evidence supporting H2, which suggests participants spent more time in AR.

We conducted a correlation analysis to investigate the connection between scores from the VZ-2 paper folding test and the answers' accuracy in the experiment. The analysis uncovered a moderate positive correlation, with a correlation coefficient of 0.40, suggesting that participants who achieved higher scores on the folding test tended to have greater accuracy. However, this relationship was not strongly pronounced. After removing an outlier, further regression analysis supported these findings by showing a moderate positive linear relationship (Figure 4.3).

A non-parametric Wilcoxon signed-rank test was conducted to compare the mean SUS scores between the AR and VR environments after removing one outlier in the AR condition. The non-significant p-value of 0.119 suggests that there is no statistically significant difference in system usability across the AR and VR environments. We attribute the comparatively low SUS scores

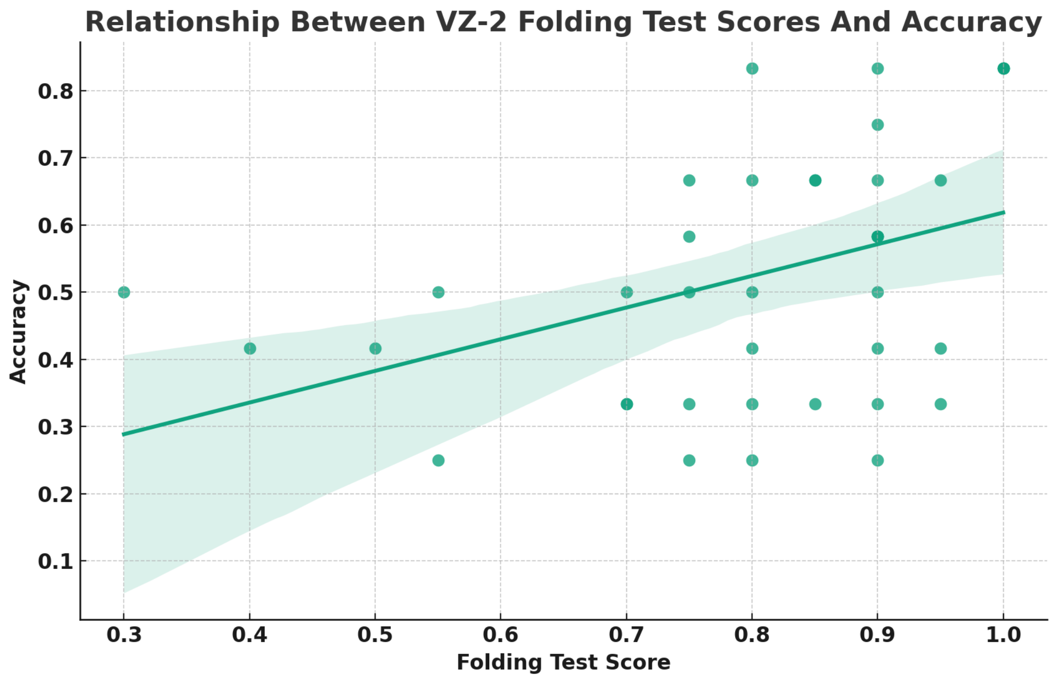


Figure 4.3: Regression Line of Folding Test and Accuracy

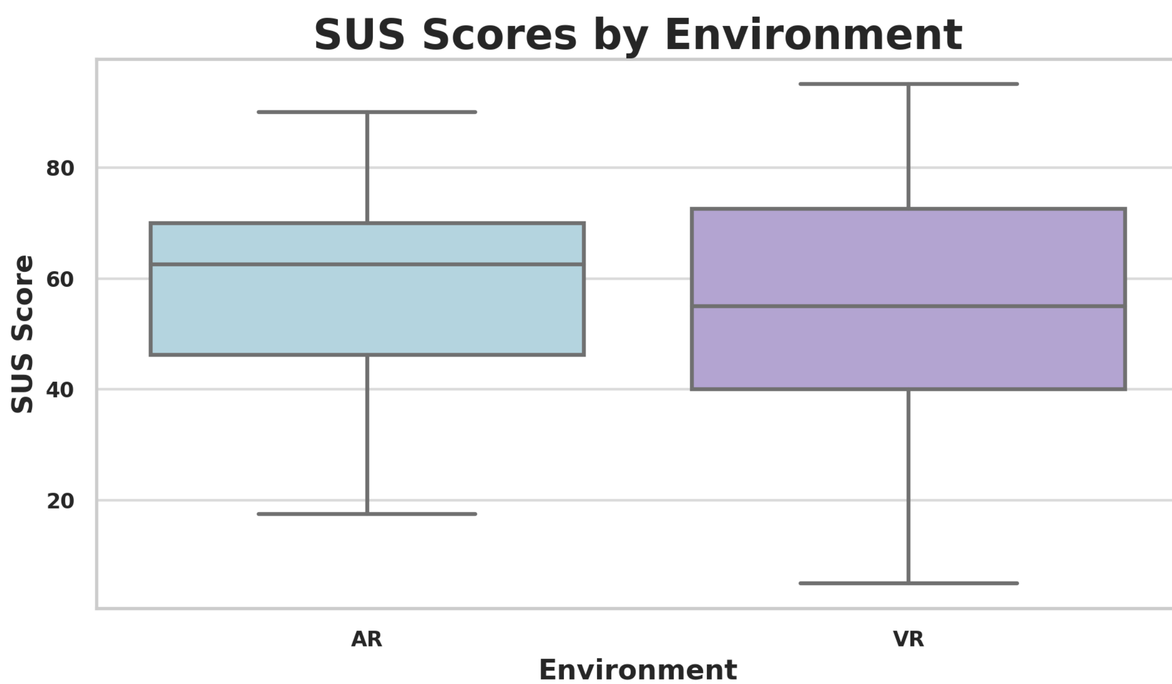


Figure 4.4: SUS Scores of AR and VR

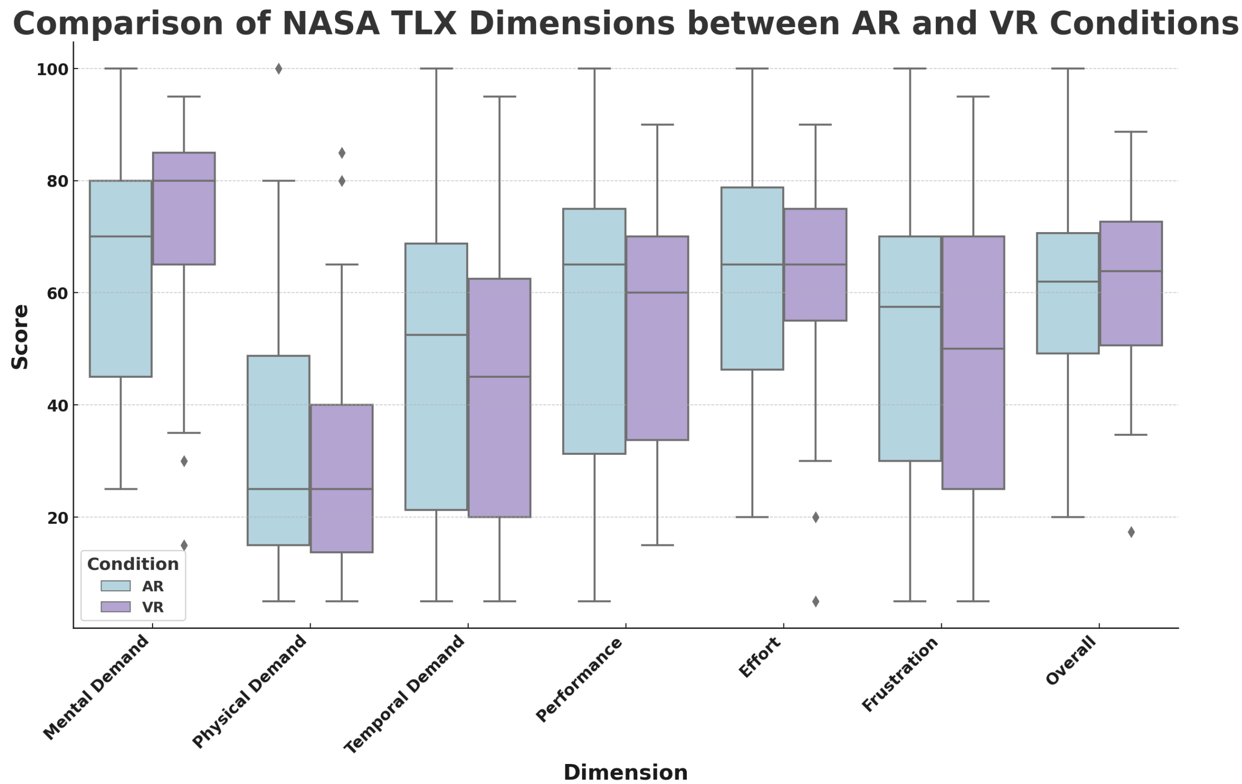


Figure 4.5: Comparison of NASA TLX Overall Scores

(Figure 4.4) to technical difficulties we encountered multiple times with XR-3 during the study, which were promptly resolved upon identification. Participants reported experiencing visual lagging and latency response from the headset once connection issues arose, likely impacting usability ratings. Additionally, a wide spread of SUS scores supports these technical challenges, as some users found the system highly usable while others experienced abnormal usability issues.

4.5.1 NASA TLX and Presence

A paired samples t-tests was conducted on NASA TLX scores from each dimension with both conditions. None of the dimensions show a statistically significant difference between the AR and VR conditions (using a conventional alpha level of 0.05). As shown in Figure 4.5, the mental demand scores are generally higher in VR than in AR, and frustration in both conditions is comparable. In addition, correlation analysis (refer to Figure 4.6) revealed that a smaller radar area for frustration in VR, which indicated frustration is less interconnected with other workload aspects

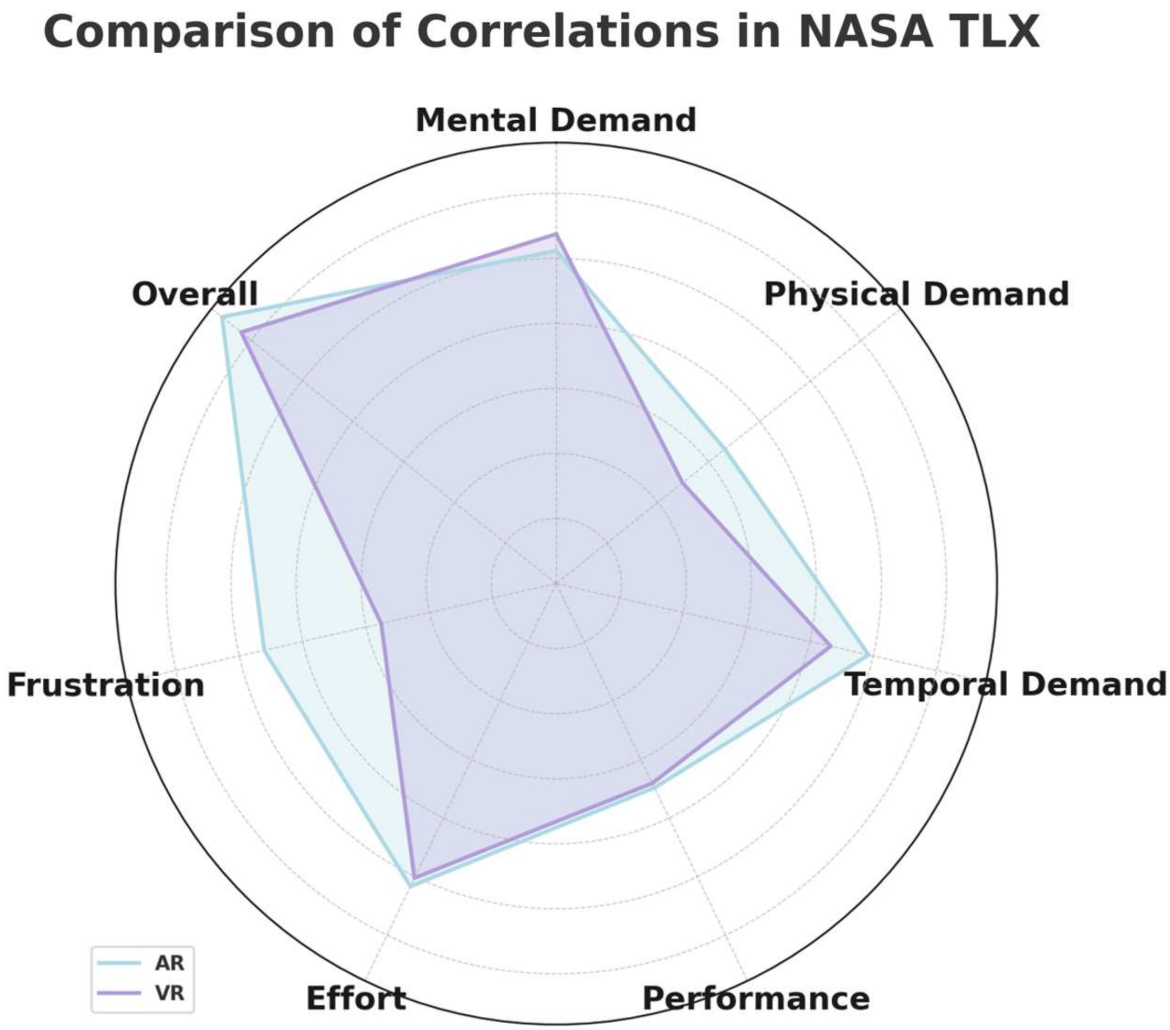


Figure 4.6: Correlation Between Categories in NASA TLX

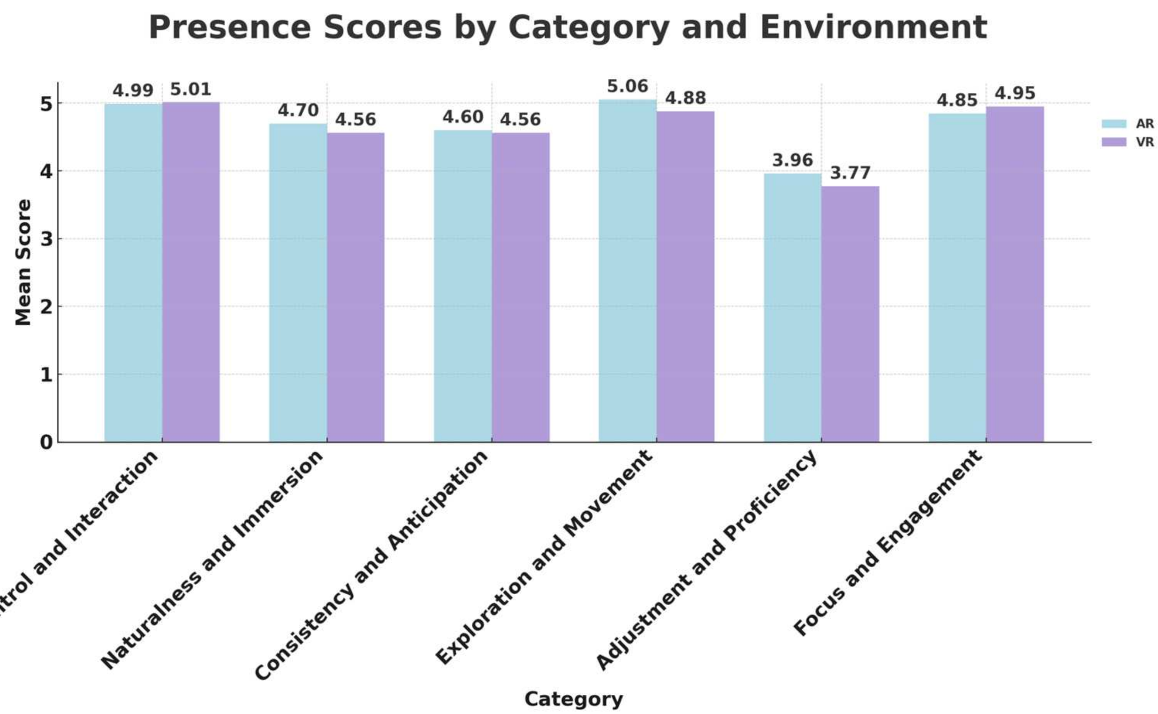


Figure 4.7: Comparison of Presence Scores

in VR than in AR. These findings suggest that users might be more tolerant of errors or learning curves associated with a fully immersive experience.

The overall average presence score across all questions and environments is 4.56 on a scale of 1 to 7, suggesting a moderate to high level of presence experienced by participants. Due to the unequal variances across conditions, we conducted Wilcoxon signed-rank tests on aggregated categories of questionnaire items to compare the presence scores between AR and VR conditions. The results revealed no statistically significant differences across most categories at a alpha level of 0.05, suggesting comparable immersion and interaction for participants in both settings. In Figure 4.7, However, some categories approached significance, suggesting nuanced differences in participant experiences between the two environments. For example, categories of naturalness and immersion approached significance ($p = 0.051$), indicating a possible difference in how natural and immersive the environments were perceived. Also, categories of adjustment and proficiency approached significance ($p = 0.078$), hinting at potential differences in how participants adjusted to and interacted with the environments.

Figure 4.8: Postures in the Experiment

4.5.2 Video Analysis and Interview Feedback

In their study, Lisle et al. found that the tether of the HMD caused concern among participants, potentially affecting their choice to walk in the VR environment [3]. To address this issue, a hook was used in our study to hold the tether above the participants, eliminating distraction and preventing incidents. After collecting postures shown by each participant in each environment, three major postures were identified: sitting, standing, and walking. A chi-square test revealed no significant difference between conditions across three postures that we found among participants throughout the entire study (Figure 4.8). Additionally, participants showed more varied postures in VR compared to AR; particularly an increased amount of walking activity. This result contradicts our hypothesis H3 as it shows more physical navigation occurring in VR than AR. Furthermore, 22 participants demonstrated greater variation within similar postures in VR for sitting and standing compared to AR (Figure 4.9).

Interestingly, six participants replied that they did not really notice that they were in different environments or the differences between AR and VR. *"I did not notice that I am in different environments, all I noticed is that oh I am more familiar with how to use these tools."* Nevertheless, for participants who noticed the differences, they indicated that they have greater situational awareness in AR, enabling them to better understand the space they are navigating and facilitating the sensemaking process. Video analysis further supports this, showing that five participants walked a longer distance in AR compared to VR while exploring visualizations under both conditions.

Similar to Whitlock et al. discovered in their study [94], three participants mentioned that text-reading in AR is difficult due to the presence of physical objects in the background. Additionally,

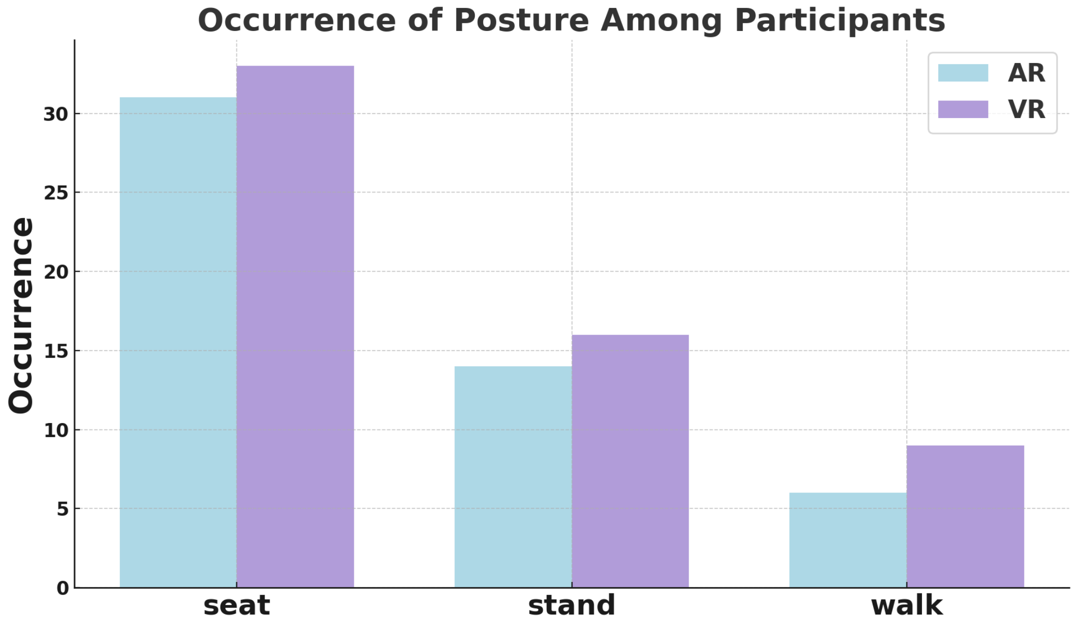


Figure 4.9: Participants Postures Distribution in AR and VR

two participants brought up concerns about clutter and its distracting nature. However, as shown in Figure 5.1, there does not appear to be any significant clutter in the room - it resembles a typical office setting. This observation validates that ordinary objects in an AR environment can disrupt users when they are engaged in demanding cognitive tasks, even without influencing performance [217]. Nonetheless, participants prefer AR due to the safeness and also the natural feeling it provides. *"VR feels fake, I like AR better, cause VR just a little bit silly."*, *"AR is really cool. You safely see everything around you, but at the same time, you are totally doing some virtual reality thing..."*, *"In AR, can see the real world, feels more natural."*, *"I kinda like the AR more, because it feels more like I am here."*

Half of the participants mentioned that practice and training helped them become familiar with the tools and data, allowing them to answer questions more quickly. *"You just need to get used to using it and focus on it."*, *"As long as you're accustomed to the tool, answering the question is easy."* Individual preferences also play a role in how participants choose to navigate through the data: *"Once I figured out how to manipulate the visualization for my needs, it became much easier."*, *"Being able to move around was really helpful..."*.

4.6 Discussion

The NASA TLX responses gave us insights into the importance of immersion and engagement in virtual environments for enhancing the user experience. The high correlation between Mental Demand and Effort, along with the low correlation between Frustration and Effort in VR, suggests that even when tasks in VR require significant effort and cognitive load, the immersive experience may help alleviate feelings of frustration. The immersive nature of VR could potentially induce a *flow* state where participants become so engrossed in the activity that they lose track of external frustrations. Additionally, by offering a more seamless and entirely digital environment, users are spared from having to integrate digital elements into the physical world, which can reduce potential sources of frustration. In designing future iterations of immersive analysis tools, it is important to consider the complexity of tasks and provide options for different realities based on cognitive load demands. For instance, cross-virtuality systems could offer seamless integration and transitions between different reality implementations. This would allow users involved in extensive data analysis or prolonged sensemaking processes to enter a VR environment through a "portal" while being able to switch back to AR when tackling simpler tasks or wanting to maintain the connection with the real world.

Based on the responses to the presence questionnaire, participants felt that both environments provided realistic experiences and allowed for seamless exploration and movement, creating a strong sense of involvement and immersion. However, technical issues served as reminders of the artificial nature of the environment and could result in frustration or disengagement. In addition to preventing technical difficulties, it is imperative for HCI researchers to optimize visual presentation and enhance input device responsiveness. One approach could be implementing a hybrid user interface [218] that incorporates familiar input methods like tablet or smartwatch interfaces and also compensates for the downsides of mixed reality HMDs, which would help ease users into the 3D interaction while enhancing their overall presence in their experience.

The posture analysis of the participants indicates that they exhibited a wider range of postures in VR compared to AR, with notably more walking in the VR environment. This suggests that

Figure 4.10: Various of Posture in VR

participants felt secure and fully engaged with the tasks within the immersive virtual environment. Further detailed findings also affirm that immersion enhances engagement and focus. For instance, participants displayed a greater variety of still postures, as demonstrated in Figure 4.10. This reflects their relaxed state while comfortably adapting their posture, indicative of being "in the zone." These insights highlight to researchers that when aiming to introduce new interaction techniques or demanding tasks requiring high user concentration, increasing immersion during the design phase should be carefully considered.

The diversity of our participants suggests that our results are generalizable for users from various backgrounds. It also resulted in various perspectives from different ages and individual preferences. One participant highlighted the importance of considering *"different ages of users"* and the *"generation gap,"* while praising the system. Additionally, one participant expressed concerns about *"looking ridiculous"* with stretching arms for interaction but became more engaged in the experiment after realizing that *"no one is watching."* This confirmed our assumption that users exhibit heightened self-awareness when they know the experimenter is observing from a distance. As anticipated, the folding screen used in this study effectively ensured privacy, indicating that future study designs should replicate real user scenarios, such as simulating personal offices to encourage natural interaction and increase participants' engagement.

4.7 Limitation and Future Work

Technical difficulties arising from unstable headset cable connectivity have caused some participants to experience issues during the experiments, potentially introducing bias to the system's usability evaluation and overall user experience. While we were able to address these hardware issues as they occurred, future endeavors must take into account the potential for technical challenges during studies and establish contingency plans to mitigate their impact.

Our findings confirm that implementation with AR or VR is influenced by the immersive environment's nature, required efforts, and individual preferences. Our upcoming work will concentrate on integrating various realities into the immersive analytics platform, affording users the option to select their preferred reality for specific tasks or at any given time. Additionally, this involves incorporating input devices and interfaces with a minimal learning curve so that users can reduce cognitive load while becoming familiar with analysis tools and 3D visualization.

Although this study has been conducted with the best mixed reality device available on the market, the study results will vary with different environments and headsets. We are planning to validate the results with other mixed reality HMDs and environment setups.

4.8 Conclusion

This study compares user sensemaking strategies with immersive analytics using quantitative data in both AR and VR environments. By simulating the physical environment in VR and using a mixed-reality headset for both AR and VR conditions, we significantly mitigate confounding variables. Our results show comparable user performance in both conditions, as well as similar subjective feedback regarding presence experience, overall workload, and system usability. However, we discovered that users exhibit a higher tolerance for mental demand and effort in VR than in AR. Regarding navigation, more walking was shown in VR while further travel distance was found in AR. Additionally, users showed more posture variants in VR, indicating greater engagement and focus on tasks at hand. This work provides insights into how different realities facilitate the sensemaking process and offers guidelines for future immersive analytics platform design.

Chapter 5

Hybrid User Interface for Situated Analytics in Mixed-Reality

Integrating different devices into interaction and compensating for each other's drawbacks were explored in prior studies. However, previous work either used the second user interface only for input or showed the same mcontent in a different format and layout. Moreover, no research has been done by using the hybrid user interface in situated analytics, in which the availability of the user interface can change based on the user's location. To utilize the hybrid user interface to improve user interaction and navigation efficiency during decision-making, we need to understand how users use them, what factors influence their choice, and what are the opportunities and challenges for adopting HUI in situated analytics. This chapter reports on the design and implementation of a prototype that simulates a smart home with situated AR visualization with a HUI that users can interact with to assist their decision-making. Our observations and discussion provide insights into designing an efficient and helpful HUI for SA and the factors that need to be considered for implementation.

The study presented in this chapter is still under review by a peer-reviewed conference. Therefore, this chapter has been rewritten and adapted to comply with copyright requirements and to provide additional context and insight.

Figure 5.1: Images of our prototype for situated analytics using hybrid user interfaces, simulating AR in VR. Left: In the foreground, the user holds a tablet at chest height with a dashboard of situated visualizations that show energy usage data of lights in the apartment. In the background, embedded visualizations of energy usage are next to their respective lights. Right: The apartment's living room and kitchen with more embedded visualizations. Leader lines connect each embedded visualization to one or more lights that it relates to.

5.1 Overview

Augmented reality headsets offer new perspectives on viewing and interacting with data, and when combined with traditional computing devices, they can create a hybrid user interface. One potential application for taking advantage of this combination is situated analytics, where AR provides embedded views linked to physical references while a separate device like a tablet offers a situated overview of the integrated information. Our study aimed to explore how individuals perceive and utilize visualizations presented in both embedded (in AR) and situated (on a tablet) formats to accomplish their objectives. Additionally, we investigated how people perceive using a conventional 2D user interface for interaction in a 3D environment. To achieve this goal, we conducted an exploratory study using a common scenario: adjusting light levels in a smart home based on personal preference and energy usage. In our prototype simulating AR within virtual reality, embedded visualizations were positioned next to lights throughout an apartment while situated visualizations were displayed on handheld tablets. Through observation and interviews involving 19 participants interacting with the prototype, we found that although participants completed the task easily, they utilized different visualization methods according to individual goals and preferences. Our results indicated that using the 2D user interface for interaction as well as presenting integrated information proves advantageous for meeting diverse needs during situated analytics tasks. We also identified distinct roles for both situated and embedded visualizations which could lead to improved user satisfaction by minimizing attention-switching overheads within this hybrid user interface setup. We conclude by emphasizing the importance of considering users' needs, goals, and physical environments when designing effective applications for situated analytics.

5.2 Introduction

The tangible world is abundant with information that we interpret and comprehend in our everyday lives. However, this information can sometimes be imperceptible, requiring the use of external tools to enable us to comprehend the objects and phenomena in our surroundings. For instance, a specialist in grape cultivation might utilize aerial imaging data to assess the current condition of their vineyards and make decisions on how best to improve crop productivity for the upcoming season [219]. In more theoretical terms, the 'logical world' where the imaging data exists embodies directly represents the 'physical world' [220] , which in this context refers to the vineyards. Though it is within logical constructs that may guide actions, actual changes are implemented at ground level within physical reality. Performing both of these stages simultaneously in the same location can be advantageous or, at times, essential. For example, if a viticulturist arrives on-site shortly after an aerial imaging and discovers a pest infestation that was not captured by the imaging, new decisions would need to be made using current knowledge derived from both logical and physical environments simultaneously. This method of analyzing data within its specific physical context is referred to as "situated analytics" (SitA) [89].

Spatial Augmented Reality can be enabled by any technology that allows access to data near the objects it pertains to, known as "referents" [220]. Immersive technologies like virtual reality and particularly augmented reality have been utilized to merge visualizations with physical referents. These referents may be located in close proximity, creating a situated visualization (*SitVis*), or directly embedded onto them, resulting in an embedded visualization (*EmbVis*) [220–222]. The use of EmbVis reduces the spatial gap between data and referent, making it easier to establish mental connections between the two. While there may only be a single referent at times such as a street [223], there can also be numerous referents that need collective consideration like products in a supermarket [91] or elements of a museum exhibit [224]. Although having individual EmbVis for each reference is common(*One-to-One configuration*) [221], Sayara et al. point out this approach becomes impractical when supplied all spread far apart. Therefore, they suggested using one unified situated dashboard for information aggregation instead of separate visuals.

Combining a SitVis as a situated dashboard with multiple distinct EmbVis in AR forms what is known conceptually as a *hybrid user interface* (HUI) [53]. In HUIs, different display modes are merged to create a "symbiosis of interfaces" [225], where each device compensates for the limitations of the other. The use of HUIs has now become associated with AR and spatial computing and has been showcased in visualization (e.g., [64, 226]) and general contexts, such as with the Apple Vision Pro [227]. For SitA, the AR headset can offer embedded views of individual referent's data while another separate display not involved in mediating the AR experience(e.g. tablet or monitor) provides a situated dashboard overview encompassing all referent's data.

The combination of SitA and HUIs introduces a new scenario with unique constraints that must be taken into account. Initially, the layout of AR visualizations is determined by the physical referents, which may not be familiar to the user, unlike typical spatial computing applications where users have control over the layout (e.g., [228, 229]). Furthermore, interacting with and considering the physical world and its referents is crucial in SitA but not as significant in conventional HUIs that use physical space primarily for AR views. Additionally, factors such as navigating a large physical environment while carrying a non-AR display and shifting attention between AR and non-AR visualizations are likely to influence the use of non-AR displays. These constraints and requirements could significantly impact design decisions and user strategies. Prior research conducted by Whitlock et al. [230] has explored the integration of mobile and AR devices in real-world settings, with a primary focus on supporting data collection and analysis during fieldwork. However, their study did not examine how individuals actually utilize such a setup. We anticipate that this convergence of SitA and HUIs will become a standard approach for combining AR devices, smartphones, tablets, Internet of Things (IoT) devices, and sensors to enable spatial augmented reality anywhere and everywhere [231].

In order to explore the obstacles and possibilities presented by the combination of SitA and HUIs, we conducted an exploratory study of the design and utilization of a SitA application, wherein users having access to a SitVis dashboard on a tablet along with several EmbVis in close proximity to their referents. We intentionally designed our prototype and study to resemble a simple baseline scenario: optimizing energy usage of lights in a smart home setting [232]. This approach allows us to observe how people interact with the system in the absence of helper functions such as occlusion mitigators [233] and navigational aids [234]. From this, we seek to understand what aspects of the combination of SitA and HUIs users like and dislike, and as a result, identify relevant research challenges and opportunities. With this, we hope to build towards a formal set of guidelines for spatially situated visualization and analytics [235]

Our work contributes to both SitA and HUIs with the following:

- Identification of suitable scenarios and the physical environments that hybrid SitA can be deployed in (Section 5.4)
- A prototype (Section 5.5) and exploratory user study (Section 5.6) using simulated AR in VR wherein 19 participants used situated visualizations on a tablet with embedded visualizations in the environment for a self-directed optimization task
- Analysis of user preferences and behaviors while interacting with such a system (Section 5.7), and a discussion of takeaways and recommendations for both SitA and HUIs (Section 5.8)

5.3 Related Work

This work lies at the intersection of three research fields: the use of virtual and augmented reality (VR/AR) for data visualization, also known as immersive analytics (IA), and the aforementioned situated analytics (SitA) and hybrid user interfaces (HUI).

5.3.1 Immersive and Situated Analytics

The relevance of immersive analytics has increased with the widespread adoption of consumer VR and AR devices such as Apple Vision Pro and Meta Quest 3. According to Marriott et al. [78], IA involves using interactive, physical tools to aid in understanding data and making decisions. VR and AR offer advantages like stereoscopic 3D visualizations including space-time cubes [236], 3D scatterplots [237], and 3D trajectories [238]. While perceiving depth for these visualizations is easier in VR/AR than on desktops [239], caution should be taken when using them, particularly without justification [240]. Research suggests a middle ground by combining 2D and 3D visualizations, either side-by-side or through animated transformations between the two formats [241–243] . A third choice is to opt out of using 3D visualizations altogether. In this approach, the 3D space functions mainly as a workspace for spatially aligned 2D visualizations, which can be organized by the user's preferences [244, 245] or (partially) automated [246, 247], and potentially influenced by the surrounding context [228, 237].

In contrast, spatial augmented reality is distinguished by its focus on the physical environment. Instead of serving as a mere background or a method for organizing visualizations, the environment holds significance and meaning alongside the data [248]. Consequently, the design and placement of SitVis are influenced by and limited to the involved referents [221]. This was evidenced in a recent study conducted by Shin et al. [222], where most surveyed AR systems for SitA situate views in an absolute world manner. In other words, visualizations are typically directly integrated onto the referent to visually overlap with them—such as on grocery products' fronts [91] or buildings' facades [249]. While handling scenarios with few referents poses no significant challenges, Lee et al. [221] highlight difficulties in managing large quantities of widely dispersed referents. For example, comparing EmbVis representations of buildings at different locations would require physically moving between these locations - hence suggesting that only situated visualizations be employed under such circumstances [220, 221]. Although this approach lacks direct spatial connections between data and referents it offers a single representation providing an overview of all referents.

However, the two approaches do not have to be mutually exclusive: simultaneous provision can include both SitVis offering an overview while EmbVis provides more detailed information [250].

Our research has commonalities with three other works in this area. One of these is a study by Jahn et al. [232] where they presented a framework for displaying energy usage information from smart home sensors on both a fixed screen and an AR display on a handheld mobile phone. This involved using the mobile phone to overlay labels showing the name and energy consumption of objects when pointed at them, such as a desk lamp, in AR. However, it's important to note that their study did not include user testing nor did it incorporate the use of an AR headset. Doerr et al. [251] also employ a comparable setup involving a (virtual) tablet and referents, but their emphasis is mainly on visual highlighting in basic brushing and linking tasks, which do not fully capture the analytical aspect of SitA. Thirdly, Whitlock et al. [230] conducted a preliminary study to explore the potential of using SitVis (from mobile devices) in combination with EmbVis (from AR headsets) to facilitate data collection and analysis during field work. While they gathered input from 10 field analysts through a design probe and guided walkthrough, they did not directly observe how the analysts would independently utilize the system or encounter any associated difficulties.

To the best of our knowledge, no existing research has thoroughly investigated the natural user behavior with HUIs for SitA. Therefore, we aim to examine the advantages, obstacles, and possible approaches of utilizing a single SitVis dashboard [252] in conjunction with multiple EmbVis within a SitA setting. Specifically, our study delves into understanding the requirements of users in SitA when references and their visual representations are located at distant points within various fields of view, all arranged according to an environment's predefined layout.

5.3.2 Hybrid User Interfaces with AR and Mobile Devices

2D representations are frequently utilized in augmented reality, often taking the form of floating virtual panels [244–246]. While virtual panels offer a versatile spatial environment, there are advantages to incorporating physical displays as well. Mobile devices have gained significance in visualization research [253] due to their ubiquity [254, 255], mobility, and portability [256]. Integration of these mobile devices with augmented reality has been explored for various purposes in human-computer interaction. One common approach involves using mobile devices as input tools to manipulate AR visualizations by leveraging their touch and spatial tracking capabilities (e.g., [247, 257, 258]). Additionally, mobile devices can present 2D data views while AR provides spatially registered 3D views. For instance, MARVIS by Langner et al. [64] showcases how AR enhances data visualizations on tablets placed on a tabletop setting . On the other hand, STREAM by Hubenschmid et al. focuses primarily on facilitating 3D visualizations through AR while utilizing a handheld tablet that is both an input device and display for alternate 2D visualizations. In this study, we employ a handheld tablet as a familiar and portable interface enabling users to access and interact with 2D visualizations.

When users view digital content that is spread across multiple displays and locations, they are required to shift their attention between them. This transition has been noted to result in performance costs as the user must readjust to new visual elements at varying distances, sizes [259], or even different visual presentations [260]. Rashid et al. [261] identify numerous factors influencing the cost of switching displays, including the degree to which users need to turn in order to see all relevant displays (*angular coverage*), and whether there is semantic coherence among the content on various displays (*content coordination*). Grubert et al. [262] later summarize additional factors found in literature such as users' ability for (*divided* or *split attention* [263]) when spreading their focus across multiple devices. Therefore, systems with human-user interfaces can be developed in a way that reduces the need for switching between different modes. One study by Wang and Lindeman [264] demonstrated that aligning 3D interaction methods across a tablet view and an immersive VR view decreased this cost during a virtual world creation activity. In the context of visualization, this presents a challenge similar to that of coordinating and integrating multiple view [265], where data from various views must be examined and evaluated simultaneously. We anticipate, however, that the introduction of SitA brings about new elements that transform any straightforward guideline into a tradeoff. For example, relocating EmbVis from one room to another automatically in order to improve angular coverage could disconnect the visualizations from their physical surroundings, potentially leading to confusion.

To the best of our knowledge, there has been no research that specifically examines scenarios in which AR views in a HUI are dispersed in space and linked to tangible references. Moreover, the exploration of whether users can effectively manage this setup remains unexplored. This entails investigating the compromises involved in optimizing the placement of visualizations for alignment with references as opposed to ensuring clear visibility for facilitating comparisons across different views.

5.4 Scenarios for Hybrid Situated Analytics

Sensemaking and understanding of data in real-world settings is now a common practice, which is the same as SitA. It involving information from both digital sources [220], typically accessed through mobile devices, and the user's own perception of the physical environment. Augmented reality headsets have the potential to seamlessly integrate logical world data with physical phenomena while maintaining access to familiar tools and input methods available on mobile devices.

For example. a traffic engineer is conducting an on-site examination at a bustling intersection. Their laptop is equipped with a traffic monitoring application that generates map visualizations of traffic patterns in the nearby area using data from road sensors, and also offers access to additional details such as timing of traffic lights. In order to comprehend the reasons behind current traffic conditions, they analyze the flow of vehicles to determine if enough can pass within the allotted green time, if vehicles are moving in groups as planned, and other factors. By utilizing AR headsets for these observations, EmbVis can provide support by displaying cycle timings overlaid on physical traffic lights and presenting reported vehicle counts over individual sensors. Consequently, while the laptop provides an overview of surrounding conditions, AR technology enhances real-world observations to assist the engineer in developing a comprehensive understanding of how effectively the intersection functions and why. Consider, another example, a scenario where a group of friends in an urban setting is looking for a place to dine. Using their mobile devices, they seek out nearby restaurants and peruse customer feedback and menu prices to aid in making a decision. Additionally, they assess the attractiveness, cleanliness, and crowd levels of each potential restaurant as they walk by before choosing whether to enter. With AR headsets, all this information could be amalgamated through EmbVis displayed on the front window of the restaurant or used to guide them towards another potential dining spot via AR navigation aids [234]. Herein lies a configuration wherein the smartphone initially presents an overview of all restaurants in the vicinity using SitVis while AR offers detailed insights into each individual restaurant through EmbVis.

5.4.1 Study Scope

The fundamental research question of this work is the following: *What are the opportunities and challenges of using AR headsets to support situated analytics with mobile devices?* Shin et al. [222] stated that there exist at least four archetypal designs of SitA that aid the user in different ways and for different purposes based on the context. In other words, the design of SitA and the visualizations within it are dependent on the goals of the user and the configuration of the physical environment [221]. We will outline the scenarios in which our research is relevant, beginning with an overview of three primary attributes of the physical settings within our scope. These have been selected to reflect what we anticipate will become standard in future contexts.

- **More physical referents.** There can oftentimes be many objects and referents of interest, such as books in a library [266, 267]. As Sousa Calepso et al. [224] point out, a very small number of physical referents are typically examined in SitA research. To further explore this gap, we investigate a scenario in which many referents are considered together at once.

- **Spatially scattered referents.** Referents can also be spread far apart in sub-optimal viewing arrangements, such as in two separate rooms, thus making tasks such as pairwise comparisons [250] challenging. The layout of these referents might not even be adjustable by the user, particularly if they are stationary in nature.

- **Interaction with the environment.** Another key aspect of SitA, as Sousa Calepso et al. [224] again argue, is the combination of both physical and analytical tasks. That is, the environment should considered and/or interacted with for SitA to have proper value, otherwise it is simply IA.

These three environmental characteristics inform the baseline setup to which a relevant SitA application can be designed. These effectively form the basis of our exploratory prototype in Section 5.5.

- **Situated visualizations on the mobile device.** As a baseline, we assume that SitVis are accessible on a mobile device, be it a tablet, smartphone, etc. The design and layout of these visualizations may either be bespoke (e.g., [230]), similar to a dashboard [252], or simply an existing mobile application.

- **Embedded visualizations in AR.** EmbVis enables a closer integration between data and referent. Information can be seen hands-free directly within their context, thus allowing the referent to be physically interacted with and manipulated.

Under the classification of Shin et al. [222], this form of SitA application would be considered an *Assistant*. Each referent has some associated data that supports the user's decision-making process. The *situating trigger* is, therefore, the referents themselves as they come into view, the *view situatedness* a combination of device-relative for the mobile device and world absolute for AR, and the *visual encoding* and *data abstraction* dependent on the type of data that is being visualized.

5.4.2 Example Working Scenario

We now present a possible scenario in the near future that illustrates mobile devices and AR as parts of HUI might be used in a SitA application in an environment with the aforementioned characteristics. We use the same scenario for our following prototype in Section 5.5 and exploratory study in Section 5.6.

- **The environment.** Emma has recently moved into a new five-room apartment at the beginning of the month. Each room has several light bulbs that are all connected to a pre-installed smart lighting system, which supports manual adjustment of brightness levels. The system provides the estimated energy usage per month (in kWh) of each light bulb to inform the user of their energy consumption. The previous tenants have also consented to provide Emma the historical energy consumption of each light bulb in the previous six months.

- **The task.** Emma, being environmentally conscious, optimizes the brightness levels of each light bulb based on their current and historical energy usage within each room and across the entire apartment. She also takes into account other real-world factors not typically captured as data, such as the presence of natural lighting, the actual perceptible brightness of each different light bulb, and her own subjective preferences of what she considers pleasant to live in.

- **The visualizations.** Emma installs the smart home application on her mobile device and synchronizes it with her AR headset. The mobile application summarizes the energy usage of all the lights in each room and the overall apartment in a convenient dashboard. After putting on the headset, Emma sees several small EmbVis overlaid on the light bulbs throughout her apartment, which indicates the energy usage of each light. With this, Emma can make fine-grained adjustments to each light bulb individually, and see both how bright the light shines and how much energy it is now estimated to use in real time. Emma also walks from room to room to check on and adjust each set of lights. Referencing to the mobile application allows Emma to confirm that the overall energy usage is satisfactorily low enough, or if there are other light bulbs she would like to adjust.

The above scenario follows the same smart home setting presented by Jahn et al. [232] as mentioned earlier. The task was also inspired by Schröder et al. [268], who also employed heterogeneous devices (i.e., desktop, tablet AR, headset VR) used in an asynchronous manner to solve a spatial optimization task with lamps in a virtual park. Unique to our scenario however is the *synchronous* use of multiple devices, and our focus on single-user contexts and the absence of quantified "score" that the user optimizes for.

5.5 A Prototype for Situated Analytics with Mobile Device and AR

The purpose of this work is to build a preliminary understanding of how people use a SitA application in the form of a HUI. To do this, we develop a SitA prototype as shown in Figure 5.1 which we later use in an exploratory study in Section 5.6, employing the scenario described in Section 5.4.2. We intentionally keep the system's functions simple for it to serve as a baseline. This is to understand the extent that known factors such as occlusion and the need to attention switch (Section 5.3) negatively affect people's ability to use the baseline system.

5.5.1 Experiment Setup

The prototype is developed using Unity 2020.3.4f1 and tested and deployed for use with the Meta Quest Pro in VR with two tracked handheld controllers. We made the decision to simulate AR using VR for several key reasons. VR offers a controlled environment that is unconstrained by technological limitations such as poor tracking, calibration, network connectivity between devices (e.g., IoT sensors), ergonomics, and field-of-view (e.g., with the HoloLens 2). Previous studies have cautioned against the use of AR as it may mask the potential benefits of SitA due to these limitations [224, 230], whereby the benefits (and drawbacks) of SitA are precisely what we hope to understand from our later exploratory study. Moreover, research has suggested that insights from AR studies can be replicated when simulating AR in VR [269–271]. Recent AR headsets like the Apple Vision Pro may have overcome some of these technical challenges, though the headset was not available when this prototype was developed and the study conducted. The use of VR also avoids the logistical issues of conducting a study in a real apartment, be it owned by the authors or temporarily rented, which would be personally invasive or expensive. While a less personal scenario would be an alternative (e.g., lights in an office), we wanted a scenario that would be relatable to almost any participant.

The virtual environment is an apartment, as seen in Figure 5.1 right, that is approximately 10 m by 12 m and consists of five rooms: a living room, two bedrooms, and two bathrooms. As per our scenario, the referents in question are the light bulbs that are spread throughout the apartment. There are 9 ceiling lights, 5 drop lights, and 9 counter lights and lamps. Note that we did not artificially distribute the lights to keep their locations as organic as possible (e.g., no drop lights in the middle of a bathroom). To adjust the brightness of each light, the user points at the desired light with their right-hand controller and presses the trigger button, cycling between 4 brightness levels. The lowest brightness level turns the light off. A raycast beam indicates what the controller is pointed at, and changes from red to green when the pointed object is a valid target. This can be seen in Figure 5.3 right. Some lights are treated together as a group where logical, such as the adjacent ceiling lights in a room, and thus cycle their brightness levels together. All light sources update in real time, illuminating their surrounding areas accordingly.

Using VR for large virtual environments necessitates using locomotion techniques, such as teleportation, which can cause cybersickness [272] and distract from the true aims of our study. We therefore ensured our prototype worked in a large real-world space, allowing users to physically and naturally walk from one end of the apartment to the other, thus more accurately simulating AR.

5.5.2 Situated Visualization on Tablet

On the left-hand controller of the user is attached a virtual tablet, as seen in Figure 5.2 left, which represents the SitVis used for the task. This mimics a person carrying a tablet with them with one hand. The tablet is 24 cm by 16 cm—large enough for text to be easier to read, but small enough to be similar in size to a real tablet. It can be held in any manner the user wishes, such as in Figure 5.1 left at chest height. The tablet's interface takes on a dashboard-style layout and consists of four separate views in the four separate corners.

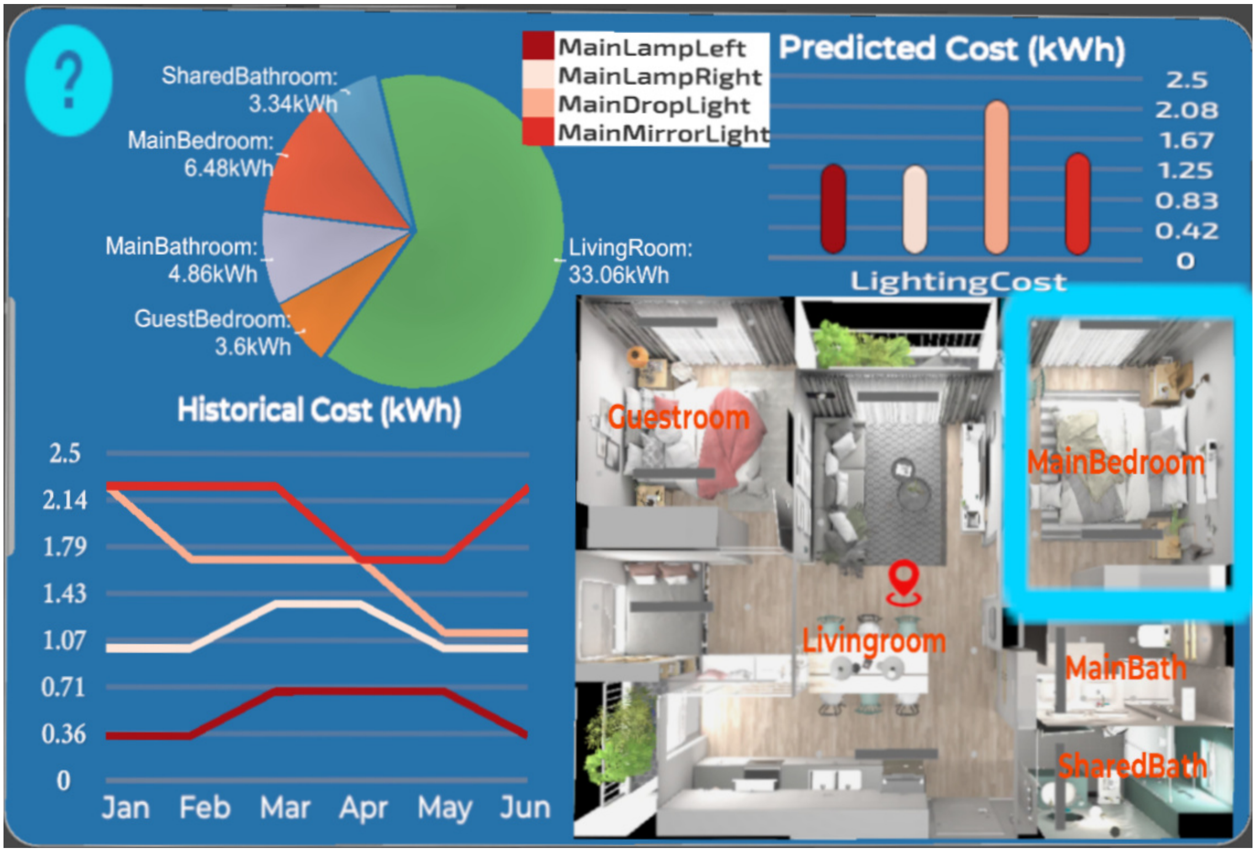


Figure 5.2: The dashboard of situated visualizations (referred to as SitVis) that appears on a virtual tablet attached to the user's left hand with four linked views: a pie chart, bar chart, time series, and map.

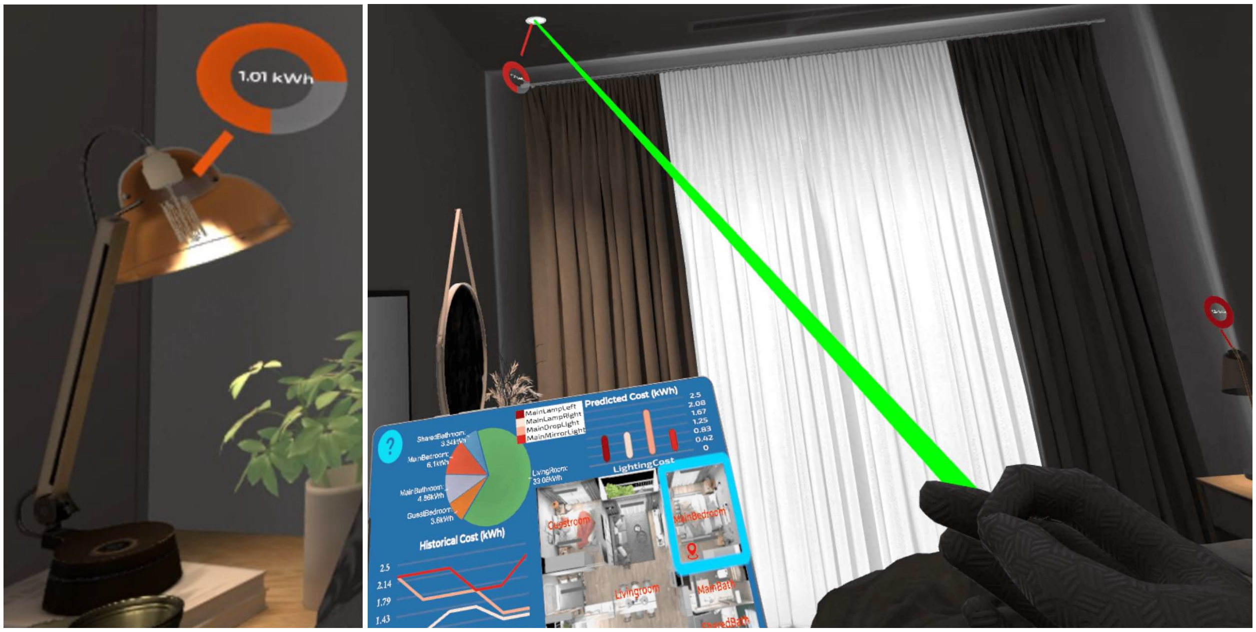


Figure 5.3: Left: An example of an embedded visualization (referred to as EmbVis) that is attached to a bedside lamp. Right: The user points at a light with their right controller and presses the trigger to adjust the brightness level at incremental stages.

The **pie chart** on the top left provides an overview of the energy consumption of lights across the five rooms of the apartment. It allows the user to understand the part-to-whole relationship between each individual room and the rest of the apartment. Using the raycast controller, the user can also select a slice (i.e., room) to filter the other three linked views accordingly. The **bar chart** on the top right therefore shows the breakdown of each set of lights and their individual predicted energy usage per month. This is color-coded with the **time series** on the bottom left, showing the historical energy usage of each set of lights over the previous six months (as per the scenario in Section 5.4.2). Lastly, the **map** on the bottom right provides a spatial reference for which room the user had selected on the pie chart as indicated by the blue outline, and where they are currently located in the apartment as indicated by the red "location" icon which moves in real-time.

5.5.3 Embedded Visualizations in AR

Next to each light is a simple **donut chart**, an example seen in Figure 5.3 left, which are the EmbVis that would be seen in AR. It shows the predicted energy usage per month of the light.

The donut, in particular, encodes the part-to-whole relationship of its current usage at the active brightness level versus its maximum possible usage. In cases where lights are logically grouped together (e.g., all illuminating the same area), the donut chart shows the sum of all lights, with leader lines indicating the respective lights that it is linked to (Figure 5.1 left). As this work is preliminary in nature, we intentionally chose not to include more complex visualizations, particularly those that are multi-dimensional, to focus on the usability of the overall SitA setup and not on the understandability of the chosen visualizations.

To mitigate the context switch cost, we designed the HUI considering aspects mentioned in Section 2.4.2 regarding transitional interfaces:

- **Maintain context.** Our prototype does not support users to leave annotations or highlight data points to assist users in maintaining context and spatial memory while switching between interfaces.

- **Liking content.** We used the same color encoding for both interfaces, which helps users establish and maintain a mental connection between visualization on the tablet and the 3D space,

- **Switching Cost between devices.** With the mixed reality headset used in our design, users do not need to remove the headset to see the content displayed on the tablet. Although we did not verify this design due to simulated AR in our study, it is believable that users can read it on a tablet with an MR headset, especially since the MR headset display quality is improving rapidly.

- **Consistent user experience.** Our tablet prototype was simulated with VR so that users can use the same controller to interact with the tablet and the virtual object, which represents the referent in an AR environment. As a result, the interaction between different interfaces in our study is consistent. With the real tablet and AR environment, we imagine the interactions are easy to keep consistent with the hand gestures in AR and finger-touching with the tablet.

5.6 Exploratory User Study

We now describe the exploratory study that we conducted using the SitA prototype described in Section 5.5. The goal was to observe how people use the prototype, elicit the perceived benefits that such a SitA application can have, and understand the drawbacks and challenges that need to be resolved to support future SitA systems.

5.6.1 Study Design and Data Collected

Being exploratory in nature, we give participants only the single scenario and task in Section 5.4.2 that we believe to be common in the future. We discuss other possible task types later in Section 5.8 that would further explore the breadth of SitA. The data provided to participants during the study was artificial, but chosen within ranges that would be plausible. This was particularly relevant for the time series data that needed to be created per light. Else, the energy usage of each light was dependent on the brightness level which the participant was tasked with adjusting. These levels were set to their maximums by default.

We collected subjective quantitative data in the form of three questionnaires. The Simulator Sickness Questionnaire [273] (SSQ) was used to determine if our use of VR had inadvertently induced cybersickness in our participants, which may bias their feedback. The NASA Task Load Index [1] (NASA-TLX) and System Usability Scale [274] (SUS) was used to compare our prototype with known benchmarks [275, 276] and help validate our prototype as a usable SitA system. The NASA-TLX questionnaire included the pairwise weighting procedure. We also conducted semi-structured interviews after the task to obtain subjective qualitative data about their general thoughts on the prototype, suggestions for improvement, and whether they could see themselves using a similar system in AR in the future. Several questions were also guided by the considerations raised in Sections Introduction and 5.3. These questions relate to the strategies participants used to: solve the task as a whole; to look at and use both SitVis and EmbVis together; and to navigate the apartment. Lastly, we collected logs and first-person video recordings of participant movements and interactions with the system. The study was approved by our university's ethics review board.

5.6.2 Apparatus

The same setup described in Section 5.5.1 was used to conduct the study. The study was conducted in a university classroom that had a usable space of approximately 12 m by 14 m—larger than the virtual apartment. We used the Oculus Air Link feature to stream from the experimenter's PC running Unity to the Meta Quest Pro, allowing participants to freely walk around the virtual apartment without the risk of tripping over a tethered cable. The PC was equipped with an Intel Core i9-13980HX 2.20 GHz with 32 GB of RAM, with an NVIDIA RTX 4070 Laptop GPU with 8 GB of VRAM. When answering the questionnaires, participants entered their responses on the experimenter's tablet. We recorded the semi-structured interviews with a GoPro Hero7 Black, set up in such a manner to ensure a high-quality audio recording to be transcribed afterward.

5.6.3 Study Procedure

The study comprised five sections, totaling approximately 50 minutes.

Pre-study (5 minutes). Participants first signed a consent form and filled in a demographics questionnaire. The experimenter then helped the participant put on and adjust to the Meta Quest Pro.

Training (10 minutes). The experimenter then introduced the scenario of the study as per Section 5.4.2. They then guided the participant through the virtual environment and its visualizations and features. Participants were then given five minutes to freely explore, use the system, and ask questions. While the same apartment and lights were used during training, the energy usage data shown to participants was heavily simplified compared to the actual task.

Short break (5 minutes). Participants were then allowed to take a short break before the study began if requested.

Study (15 minutes). Participants were then instructed to complete the task as per Section 5.4.2. They were instructed to think aloud during the experiment and were also told that there was no correct or incorrect answer to the task. A text version of the task could be accessed at any time by selecting the "?" button on the top left of the tablet (Figure 5.2). The participants had to indicate to the experimenter when they felt like they had adequately completed the task, which ended the study.

Post-study (15 minutes). Participants answered the SSQ, NASA-TLX, and SUS questionnaires on a physical tablet. They were then interviewed by the experimenter. At the end, participants received a US$20 gift card as compensation for their time.

5.6.4 Participants

We recruited 19 participants (six female) in total. To mitigate against the effects of novelty bias, especially as the majority of our data collection is subjective in nature, we sought to recruit participants who already had experience with VR and/or AR. 12 experienced participants (four female) were thus recruited from both a VR/AR course taught at the local university and from our VR/AR research group via internal mailing lists and word of mouth. These participants are labelled P1 to P12. To still capture a broad range of opinions however, we also recruited seven novice participants (two female) from the same university who had next to no prior experience with VR or AR via the same advertising methods. These participants are labelled P13 to P19. All participants had normal or corrected-to-normal vision. One participant was red-green color blind and had difficulties seeing the red and green raycast from the right-hand controller, though they did not report any major difficulties in reading the visualizations. One other participant was left-handed, though they reported no difficulties in using the prototype.

5.6.5 Data Analysis

For the semi-structured interviews, the audio recordings were first transcribed using AI and then cleaned by the author who conducted the interviews (the experimenter). The same author and a separate author who was not present for the interviews (the non-experimenter) then used thematic analysis [277] on the transcripts together using an inductive approach. The transcripts were read and discussed by the two authors together and, whenever a statement was deemed relevant, a code was assigned that the two authors agreed on. Codes were progressively merged as similarities were found throughout the analysis and were finally assigned to overall themes. The non-experimenter then reviewed, analyzed, and structured a narrative report of the themes, which was later reviewed and validated by the experimenter. For the video coding, a technical issue caused the video footage of P14 to not be saved. Regardless, the same experimenter first compiled an initial set of codes based on handwritten notes and a review of six of the 18 videos. The codes were then refined after discussion with the non-experimenter. The experimenter then applied the refined codebook to all 18 videos.

5.7 Results

We present our results from participant demographics (Section 5.7.1), questionnaire responses (Section 5.7.2), interview feedback (Section 5.7.3), and observations from video coding (Section 5.7.4).

5.7.1 Demographics

The mean age of our participants was 25.05 years old (*sd* = 4.21). Regarding their experience with AR and VR, six reported using either daily, five weekly, and two monthly, and the remaining six reported N/A. For using 2D data visualizations, four participants reported using them daily, seven weekly, four monthly, and the remaining four reported N/A. For using 3D data visualizations, only seven participants reported any prior experience.

5.7.2 Questionnaire Responses

The results from the three administered questionnaires are as follows.

SSQ. The calculated total simulator sickness score [273] has a mean value of 20.47 (*sd* = 17.11). According to Stanney et al. [278], this puts the prototype slightly above the threshold of 20 for “a bad simulator”.

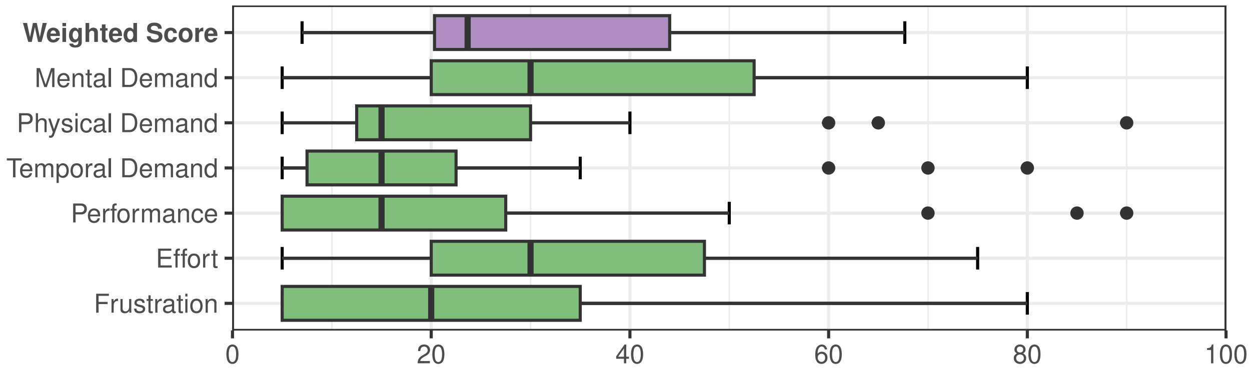


Figure 5.4: The distributions of the overall weighted scores and the six individual unweighted subscales from the NASA-TLX questionnaire [1].

NASA-TLX. Figure 5.4 shows the distribution of scores for each of the six unweighted subscales and the overall weighted workload score. The weighted workload had a mean score of 31.65 (*sd* = 18.07), which according to Grier [275] places the prototype at slightly below the 20th percentile of global NASA-TLX scores which is 33.00.

SUS. The mean overall score for the SUS was 80.66 (*sd* = 10.27). When breaking this down by participants' experience level to account for novelty bias, experienced participants gave a lower average score of 77.50 (*sd* = 10.50) than novice participants with 86.07 (*sd* = 7.75). According to Bangor et al. [276], the overall and experienced scores correspond to "Good" and the novice scores correspond to "Excellent".

5.7.3 Interview Feedback

We now report on the relevant topics raised by our participants during the interview, categorized by the coded themes from our analysis. Note that in this section, quotes have been modified to remove filler words for readability, and that "tablet" is used synonymously for SitVis, and likewise for "donut chart" and EmbVis.

General impressions of the prototype. Participants provided positive feedback on the system. They thought it was *"straightforward and easy to use"* [P15], *"seems like something that would be useful"* [P3] and that *"it was very neat, about the tablet [and] light kind of combination"* [P5]. They also reported feeling immersed in VR, admitting that *"I know I could walk through the bed* *or walk through the table, but I just kept walking around"* [P1] and that they were *"definitely immersed"* [P17]. However, the use of VR did make it *"hard to read [the tablet]"* [P14], with at least eight participants making similar statements. There were also some technical issues as when *"you turn too fast or walk too fast it kind of lags"* [P16]. From a task perspective, several participants reported that the SitA prototype was well suited as *"the system allowed me to have the information I need when I needed it"* [P7], with one participant acknowledging that *"we have a tablet that can also display data, it makes sense to potentially capitalize on that"* [P3]. Some comments also signified the benefits of situatedness, such as one making the observation that *"some rooms will need more energy just because there's more lights in them and they're bigger"* [P12] and another saying that *"it's good to see that the bedside lamp took a lot of energy, but it wasn't that bright"* [P17].

Impressions of the visualizations. Regarding the SitVis on the tablet, feedback was mixed and sometimes contradictory. For instance, P3 said that the pie chart was the most useful, whilst P12 did not see the purpose of it. Likewise, four participants found the bar chart helpful, whilst P15 *"never even looked at [it]"*. Only the line chart received clearly negative feedback, summed up by the following: *"I was more focused on how I wanted to set up my space, not what someone else did before me"* [P3]. In contrast, eight participants gave clearly positive remarks about the EmbVis. They *"really liked that although I had the tablet, there were [donut] charts all by the lights as well"* [P15] and thought that *"having them floating there is a good indication"* [P8]. It is possible however that they interpreted the EmbVis not as a data visualization, but as a "gauge" that signifies the current lighting level of each light. Four participants made statements that suggest this, such as *"it was nice to remember that I liked the lighting at a quarter, and I could go back to it easily"* [P10], and *"I used less the number on [the donut chart], and more just the percent of what was left"* [P17]. This is reinforced by a fifth participant who thought the donut charts *"were useful to know what lights I could control, but other than that, I didn't really use the power numbers too much"* [P5].

Navigating the space and its referents. This theme focuses on how participants decided **where** in the apartment they should go to first. While not all participants gave clear responses, we coded seven participants who suggested that they did zero prioritization and wandered into any room, turning out to be the most common approach. For instance, *"I just walked around and walked into the rooms exploring"* [P17] and *"I was able to click on the room [on the tablet] as I walked up to it and then look at the lights"* [P7]. Two participants said the opposite, as *"I clicked on the pie chart for each room and I went back to that room and saw if I could make any adjustments"* [P2] and *"I would look at [the map] and say, okay, I'm gonna go there next"* [P19]. Many participants described how the map helped with the task. From seven participants, this was to *"identify which room I'm in, either the main bedroom or the guest bedroom"* [P5] as the two looked alike, which then let them *"figure out which [room] to click on the pie chart"* [P1]. Five participants used the map as a reminder tool, such as *"during the training, I almost missed the bathroom next to the door, but then the map helped me realize there's a bathroom right there"* [P4]. On the other hand, four participants reported not using the map at all. Several comments hint as to why: P15 applied her intuition of how apartments are typically laid out, and P4 and P16 became familiarized with the apartment's layout after the training phase.

Workflows for using visualizations. This theme focuses on **what** visualizations participants used and **how** they used them for the task, after having entered a room with lights in it. We coded four distinct workflows for how the SitVis and EmbVis were used, though we could not confidently assign one to each participant. The most common workflow from six participants was to first focus on the EmbVis, adjusting the lights as necessary, then validating their changes on the SitVis on the tablet. For example, *"when I was when I thought I was done with the room, then I would check the chart on the tablet and see if I felt like that was good enough"* [P12] and *"I used the donut graph [...] and then I used the [tablet] a lot to see the actual statistical differences of changing lighting levels in the room [...] this was mostly just seeing if I had half lighting that I could only get maybe a quarter more lighting down"* [P17]. Two participants reported the opposite workflow, with one saying *"as I walked into a room, I would click on the pie chart, and I would look at what* *areas were taking more power and especially at anything that was at the top of those charts, and then I would look around the room and find the light that was presenting the highest amount of energy usage"* [P7]. In contrast to these two workflows, several participants claimed not to have even bothered with one set of visualizations or the other. Three participants suggested they only used the tablet, as they *"haven't concentrated much on the graph close to the light"* [P6]. Two others instead felt that they only needed the EmbVis, as *"for me to accomplish what you told me to do, I don't think the tablet is needed"* [P18].

Interaction with Hybrid User Interfaces. This theme focuses on **how** participants felt about the interaction with the tablet in an immersive environment. Due to the tablet being simulated instead of using a physical tablet in our experiment, participants reported it was hard to read with the true size of a tablet but less clarity of the content presentation in VR. Also, some participants thought the casting ray from the controller could be thinner, so it is easier to choose the section on the pie chart. Some participants mentioned that they wished the tablet could be bigger so that the graphs and visualization were big enough to read and interact with easily. Moreover, a couple of participants said it would be much easier to interact with the tablet if it is in the real world and they could use touch input.

Priorities in decision-making. This theme now focuses on **why** participants adjusted the lights the way that they did. As with the preceding two themes, we identified three distinct mindsets based on participant responses. Five participants would first adjust the lights based on their own personal preferences, and then make further adjustments and refinements by referring to the energy data. For instance, *"I was definitely prioritizing my preference over the cost. If I prefer [the] room to look like this, I'd make the cost work"* [P1]. Four participants would instead try to find a balance between their personal preferences and energy usage. For instance, *"[it was] personal preference, which is directly also tied to keeping energy levels down. I want to use the least amount of energy possible, while also keeping comfort in"* [P15], and *"I just tried to make it comfortable, but still use as little lighting as I could"* [P17]. Perhaps the most extreme option, four participants reported having ignored the data entirely, only following their personal preferences. For instance, *"I just* *went with my personal preference, whether I want the lights or not. And mostly, I keep my lights off"* [P13] and *"I feel comfortable without light, or I usually prefer natural light. So I just turn off all the lights and that's it"* [P18].

Further improvements to the prototype. The most common suggestions relate to the design of the SitVis on the tablet and the interactions that it can support. This includes the ability to click on the room in the map to link views together [P2, P3, P7, P10], showing the total energy consumption of the room on the bar chart [P8], showing dollar prices instead of kWh [P17], the ability to make pairwise comparisons of rooms [P5], or even using 3D visualizations [P6, P8]. The only suggestion specific to the EmbVis was to have them always rotate to face the user [P8]. One participant suggested using the position of the user as a means to control which information is shown, *"so when I walk into the living room, then I see the pop ups, because there's not really a reason for me to see what light in the other room is doing, it's more about where I currently am"* [P3]. Lastly, one participant suggested showing the location of the lights on the map itself, *"if it was a bigger map, in each room you can see on the tablet, and it shows me where the light is on the map, that would be useful"* [P8].

Future use of SitA. While participants hinted at being willing to use an AR version of the prototype in the future, this came with caveats. An obvious caveat was due to the high price [P17] and relatively poor ergonomics of current AR headsets [P8]. This also depended on the actual situation, such as *"if I had a house though, I would be willing to spend a little bit of money on an app that allowed me to visualize my cost like that"* [P7]. Two participants did state that they would likely only use a similar prototype in more complex scenarios. For example, *"if it was like an industry where there are multiple machines and there can be emergency hazards and stuff, where you have to actively go and interact with the equipment, having that information [...] can be really helpful"* [P11]. This may also be the case in scenarios where *"there was a temporal demand where I'll have to look around or look for that information [...] I would like to have information right next to the object itself, but if it's not, then I can just refer to it on my tablet or my phone and not have some stuff floating"* [P11].

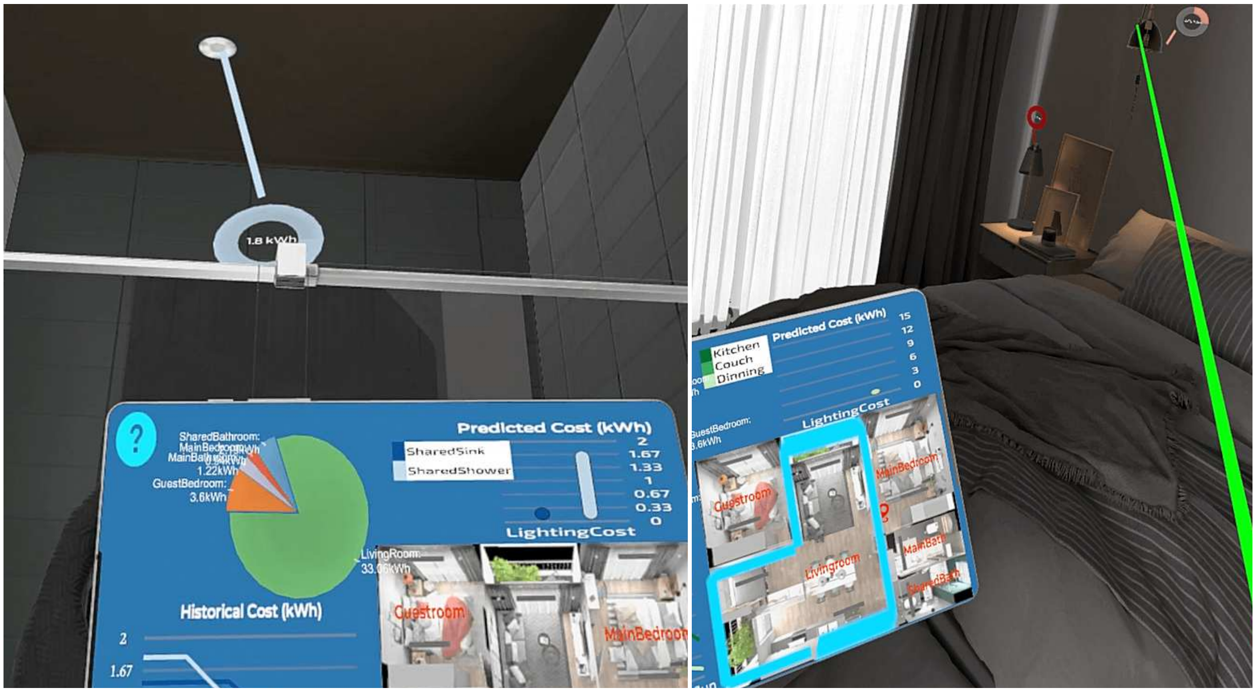


Figure 5.5: Left: P9 standing close to the light and EmbVis and needing to hold the SitVis high. Right: P6 standing far away from the light and EmbVis while holding the SitVis at a comfortable height.

5.7.4 Observations and Video Coding

In order to enhance our analysis, we present findings and outcomes derived from our video coding of participant behaviors and interactions with the system. We focus solely on the most significant results obtained through our analysis. For details on our coding process, please refer to the supplementary material provided.

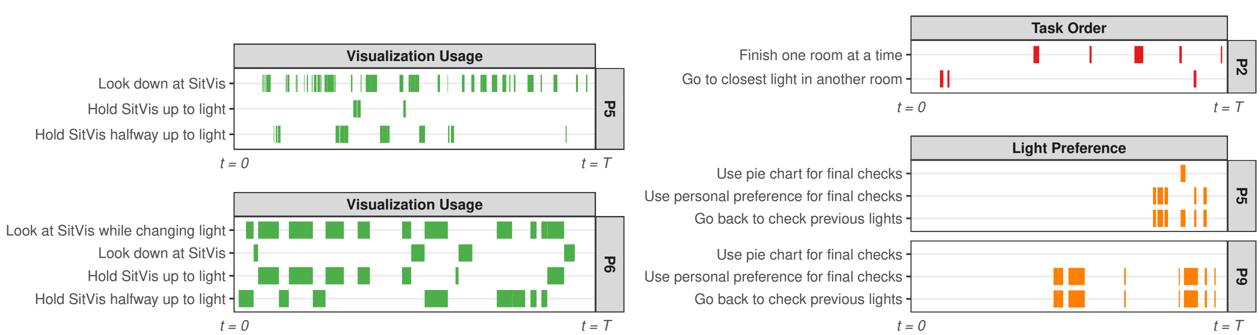


Figure 5.6: Timelines of coded behaviors. See supplemental material for complete timelines of all participants.

When and how to use visualizations. Aligned with the interview responses, there was significant variation among participants in their utilization of the visualizations. Regarding SitVis usage, two participants [P13, P18] used it for less than 6% of the experiment, while seven others [P1, P5, P6, P8, P10, P12, P17] used it for at least 40%. The timing and manner of SitVis use also varied among participants. For example, P3, P4, and P7 only looked at the SitVis after adjusting a light; whereas P6 primarily used it while in the process of adjusting a light. In contrast, P1, P5, and P10 consistently referred back to the SitVis throughout the entire experiment regardless of their activity or location (e.g., as seen in Figure 5.6 top left).

Relative positioning of the visualizations. Participants displayed varying ways of holding the SitVis device. Typically, when not engaging with a light, the SitVis was held at a relaxed hip level. However, when participants interacted with a light, they would hold the SitVis at different heights, impacting the relative angular distance between the SitVis and EmbVis. In summary, some participants such as P1, P5, and P17 maintained slight proximity to but not perfect alignment with the EmbVis when holding the SitVis. For example, they might hold it at chest height while looking at an EmbVis close to ceiling level. Conversely, some individuals, such as P6, P8, and P10, opted to keep the SitVis closely positioned to the EmbVis, even when the latter was near the ceiling (see Figure 5.5 left), for approximately 13% of the trial. However, their approach to holding the SitVis evolved over time (see Figure 5.6 bottom left). One effective strategy they employed to reduce physical exertion was standing farther away from the EmbVis, thus reducing required angular coverage while maintaining a more comfortable lower height for SitVis (see Figure 5.5 right).

Navigation strategy. Aligned with our interview results, we observed two general approaches for prioritizing which lights to navigate to. Five individuals [P2, P12, P13, P15, P19] seemed to act opportunistically by moving from one visible light to the next in their peripheral vision, regardless of whether it was in the same room or not (e.g., in Figure 5.6 top right). This behavior was particularly noticeable when the lights were scattered throughout a room, potentially causing participants to overlook them. On the contrary, four individuals (P2, P3, P4, and P19) had evidently customized the SitVis to display data from a particular room before proceeding to navigate to it. However, this behavior was not consistent across all rooms, and there was variability among participants in terms of when during the task they demonstrated this conduct. For example, P4 did so twice in the first half of the experiment, while P3 also did it twice but in the second half.

Refining light settings. In the later phases of the study, numerous participants revisited lights that they had previously modified to make additional improvements. The extent to which they consciously consulted the data during this process seemed to differ among them. Four participants [P5, P6, P17, P19] had deliberately checked the SitVis and its data whilst making these refinements. In comparison, nine individuals [P2, P3, P4, P5, P9, P10, P11, P13, P18] participated but did not specifically observe the SitVis. Instead, they focused solely on the light itself and its associated EmbVis. P5 was the sole individual to perform both actions, as depicted in Figure **??**, where a comparison is made with P9 who was just mentioned.

Beyond the lights and visualizations. During our coding process, we also observed various occasions when participants engaged with the prototype beyond the light and visual stimuli. For instance, five participants [P10, P13, P17, P18, P19] paid attention to the furniture and décor in the background during the experiment. Additionally, some others tried to interact with the physical environment by attempting to open drawers and closets. Significantly, 17 participants were seen turning their heads around after adjusting a light at some point in order to likely assess how the room looked with the new lighting level.

5.8 Discussion & Future Work

Based on our study, we now discuss key takeaways, provide several guidelines, and highlight avenues for future work.

The prototype was usable and valid. As a whole, the prototype we developed to represent the baseline of using HUIs for SitA was functional and suitable for its intended purpose. While the SSQ results show some shortcomings [278], it's worth noting that VR studies often report average scores above 20 [279], suggesting that the use of VR did not significantly hinder participants. In contrast, the NASA-TLX and SUS scores were positive, indicating that the prototype was functional and appropriate for its purpose. The low taskload may have been influenced by participants' varying levels of involvement due to self-determined task aims. The high SUS scores could also be attributed to novelty, although an average score of 77.50 from only experienced participants is still encouraging. Therefore, we believe that combining HUIs with SitA holds promise and should be further explored; however, potential differences when deploying on a physical mobile device or AR must be acknowledged. Future work should likely focus on understanding usability while carrying a physical device during interaction with physical objects—and analyzing and interpreting data at the same time.

The needs and goals of users SitA are varied, and so too should their design and evaluation. The participants' interpretations of the task and scenario varied widely, leading to different utilization patterns of the prototype. Some relied on visualizations for decision-making while others disregarded them altogether. This lack of strict guidance, such as through gamification methods [224, 268], allowed participants to make autonomous choices similar to real-world scenarios. Interestingly, no participant felt incapable of completing the task satisfactorily. In retrospect, this outcome prompts considerations about user requirements and testing strategies for validating SitA applications. It is evident that not all individuals possess an inherent "analytical need" that can be addressed by a complex SitA application. While one individual may simply want to view the overall health rating of a grocery product, another person may focus on and deliberate over the nutritional value breakdown. Given that our participants varied across this spectrum, it is reasonable to assume that this variability exists in real life as well. Hence, SitA applications should have the capability to display varying amounts of information based on users' specific goals and requirements. When assessing SitA applications in a usability context, the traditional performance metrics in visualization such as time, accuracy, or "insights gained" may be considered less significant. This is especially true when users establish their own objectives based on their expertise and preferences. While we do not propose an alternative metric for this purpose—see a related article by Wang et al. [280] for further discussion—it's important to consider the diverse goals of users when designing and evaluating SitA applications. It should be noted that our perspective on SitA focuses on its role as an *Assistant* [222], which may not necessarily apply to other archetypes like *Simulators* or *Planners* with more distinct use cases and goals.

Hybrid user interface helped maintain a low effort for interaction in an immersive environment. There is no participant reported arm fatigue during the study, which is common feedback for interaction in AR or VR environments due to the mid-air poses. Even though we couldn't utilize the tangible features of a real tablet, and users still need to hold up controllers to click on the tablet, the benefit of using HUI is already shown in our study. One participant who had limited movement in the neck and arm mentioned that the HUI setup did not cause any pain during the study since using the tablet provides options to users to choose the suitable height for individual needs. Future studies will be conducted with a real tablet in AR which offers the benefit of a tangible interface, and we can further examine the opportunities and challenges brought by adopting HUI with MR HMDs.

SitVis for cognitive tasks, EmbVis for physical tasks. The EmbVis utilized in our research were intentionally kept simple and one-dimensional. It was surprising that none of the participants proposed increasing the amount of information displayed on the EmbVis; instead, they expressed positive feedback about its current design. This suggests their appreciation for the simplicity of the EmbVis, as it directly supported their task of adjusting light levels. Any suggestions to enhance visualizations focused on adding more functionality to SitVis, such as enabling pairwise comparisons of rooms [P5]. Given our prototype's hybrid setup, we hypothesize that SitVis served as a "data analysis" tool supporting cognitive tasks, while EmbVis functioned as "helper" tools supporting physical tasks. Accordingly, a SitA designer might consider offloading complex data and visualizations onto SitVis while maintaining simplicity in EmbVis. The role(s) assigned to EmbVis may also include providing visual guidance by using glyphs and trajectories [221]. Rashid et al. [260] have recommended an asymmetric distribution of visual elements and information for minimizing attention-switching costs in HUIs; however further research is needed to confirm this hypothesis particularly regarding varying levels of information shown on SitVis versus EmbVis for SitA applications.

Attention-switching between SitVis and EmbVis was not a problem, though fatigue should still be considered. Based on our video coding analysis, it was observed that the majority of participants utilized both SitVis and EmbVis in combination for the task, with a few exceptions. One unexpected finding was that none of the participants mentioned facing challenges when switching their attention between the two visualizations, which is noteworthy considering this aspect aligns with one of our primary research interests. It's important to note that we refrained from directly asking about this during interviews to prevent any potential bias. This observation may be linked to an earlier point indicating that while EmbVis was straightforward and unique, it differed from SitVis. Nevertheless, we observed that participants held the SitVis at different heights, ranging from a comfortable hip or chest level to head height when the EmbVis was near the ceiling. Despite this variation, no participants mentioned experiencing fatigue during the interviews. It is worth mentioning that there were some outliers in the NASA-TLX physical demand score. Additionally, it is significant to highlight that compared to an Apple iPad Pro 11-inch (164 g versus 466 g), the Meta Quest Pro controller is lighter and has a more favorable center of gravity. Nonetheless, it is probable that exploring potential strategies to alleviate fatigue for similar HUIs remains essential. When there is a need for visual comparisons between SitVis and EmbVis, one clear approach would be to reduce the distance between them, although this may vary depending on the task at hand. One option could be relocating the EmbVis alongside the SitVis; however, this might not serve any purpose if users need to observe lighting changes in their environment and therefore must keep their focus around the light source. Alternatively, positioning the SitVis above the EmbVismay involve creating a mirrored image of it to preserve touch input on mobile devices. Future studies could explore effective techniques for integrating HUIs and spatial augmented reality, particularly when there is a necessity to support traditional coordinated multiple view techniques [265] such as brushing and linking [251]. As a side note, in Figure 5.1 on the right side, the donut chart section of the EmbVis appears noticeably distant from the four lights.

This situation may prompt consideration as to whether attention-switching was required between the EmbVis and its reference point. Despite not being highlighted during our interviews, we could not definitively confirm this based on video coding data without eye gaze information. Naturally, if the donut chart were positioned further away (e.g., at a distance as great as that mentioned in *viewer space* [249]), it would no longer be regarded as embedded. Specific visualizations such as glyphs and decals depend on sharing the same spatial coordinate system as their reference points [221], thus requiring innate shared attention in these instances too.

Support revisiting prior referents through bookmarks or annotations. Video analysis revealed that many participants, in the later stages of the task, revisited and made final adjustments to the lights. While their reasons for these refinements varied, some were based on data while others were purely personal preference. The recurring nature of this behavior highlights its significance. In our context, we assume that participants could easily do this due to the task's pre-attentive nature. They could effortlessly move from room to room and quickly assess whether it was slightly too bright or dark for their liking. However, this may not be applicable in a more complex task that requires thorough examination of similar-looking components of a machine rather than just a cursory look at past references. In such situations, having a way to bookmark or annotate important references may be beneficial in supporting *process and provenance* of SitA [281]. Further research could explore effective methods for tracking provenance using simple 3D drawings in AR, traditional 2D annotations on mobile devices, or novel approaches like world-in-miniature representations for embodied provenance [282].

Personalizing SitA based on the user's familiarity with the environment. Due to the physical exploration required by participants in the apartment, we naturally included a map to assist them. However, our interviews revealed that many participants did not ultimately use it for navigation. Despite their initial unfamiliarity with the apartment, they quickly became accustomed to it through training and their understanding of modern apartment layouts. As a result, they were able to complete tasks without much need for navigational or attentional support. We argue that this factor of familiarity should be considered in personalization strategies. While homeowners might not require navigational or attentional assistance, it may be necessary for tradespeople who are new to the space. Providing such information "just in case" could lead to visual and information overload when using AR cues [251]. Therefore, future research should explore how personal digital assistants can adapt based on individuals' changing needs, considering factors like familiarity among others.

There is a need to understand the relationship between physical environment and SitA. It is evident that our selection of scenario and setting directly impacted the behavior of our participants, thereby influencing our observations and conclusions. In Section 5.4.1, we outlined the scope and environment to demonstrate the real-world contexts to which our work applies. Nevertheless, a slight modification in these parameters could lead to changes in user behavior. For instance, if participants were not required to interact with the environment, they might express dissatisfaction about having to move around, potentially making a *ProxSituated* approach more attractive [283]. Furthermore, altering the environment itself could also affect user behavior and preferences; for example visually distinct rooms may eliminate issues encountered by participants when differentiating between similar-looking rooms on a map. This indicates the necessity to empirically assess how the various attributes of physical surroundings impact the usability of SitA, similar to Doerr et al.'s examination of highlighting methods based on reference layouts [251]. From a design perspective, this implies the necessity for applications to smartly adjust to evolving environments and goals, as no two settings will be identical. This could involve integration with artificial intelligence (as suggested by Shin et al. [222]), visualizations whose layout [284] or even appearance [285] adapts to real-world scenarios, or other innovative approaches.

Limitations. The apparent limitation of our research lies in utilizing VR as a means to simulate AR. While we believe that our findings can be applied to AR, as suggested by previous studies [269–271], it is clear that VR has influenced our results to some extent. For instance, some participants were focused on interacting with virtual furniture which could indicate their deep engagement. Nonetheless, this did not noticeably hinder us from obtaining valuable feedback in such cases. Our prototype also encountered performance challenges, especially when both users turned their heads quickly or moved to the remote corner of the physical space with weaker reception for the Oculus AirLink. The SSQ findings indicate that these performance effects were probably not significant in comparison to other VR research studies. One noticeable consequence of VR, however, was our inability to incorporate a task that involved physical manipulation of objects, as exemplified in the study by Sousa Calepso et al. [224]. It is possible that a task with a more practical focus [286] (i.e., working towards a physical goal) could impact the usability of SitVis and EmbVis in this context, for example if the user had to physically install light bulbs of varying color temperatures. Equally, a task that is more focused on acquiring information [286] (i.e., discovering knowledge) may also receive different types of feedback, for instance if the user had to identify an underlying issue. Our task was inherently simpler in terms of SitA, and did not necessitate the same level of analytical thinking as a field worker might need to utilize [230]. Subsequent research could aim to examine and differentiate the requirements of users when undertaking tasks with either pragmatic or epistemic objectives.

5.9 Conclusion

Motivated by the potential of HUIs to facilitate specific situations, we conducted a study on the natural utilization of both SitVis and EmbVis in an optimization task. Our goal was to gain insight into the underlying difficulties and prospects. We designed a prototype using VR technology to simulate AR within an apartment environment with dispersed smart lights. The objective was to optimize energy consumption by adjusting lighting levels. The SitVis included multiple views displayed on a handheld virtual tablet, while the EmbVis consisted of donut charts associated with individual lights. We used this prototype in an exploratory study of 19 participants, who had the freedom to move around the apartment and customize the lighting according to available data. Our results indicate that users' approaches to determining the "correct" solution varied widely, with some carefully considering the data while others disregarded it. Despite these differences, both methods led to satisfactory outcomes, demonstrating their equal validity. Additionally, our study confirmed the usability of the hybrid user interface among participants and did not identify attention-switching cost as a significant issue. These findings suggest potential for integrating heterogeneous devices in SitA applications in the future. Moreover, our results highlight the importance of broadening perspectives when designing and evaluating SitA applications to account for diverse user aims, goals, and physical environments. Subsequent studies should explore how changes in these variables may impact both human factors and system adaptability within an SitA environment.

Chapter 6

Conclusion

This thesis included three user-centered studies that aim to improve user experience and performance in interaction and navigation with AR and VR environments. The elicitation study discovered user-designed hand gestures and speech for interaction with AR HMD. The comparison study in AR and VR explored the differences between user interaction and navigation strategies during sensemaking with 3D data visualization. Lastly, the exploratory study disclosed various user experiences with using hybrid user interfaces for interaction and navigation in decision-making scenarios with a simulated AR environment. Through three studies, we found that user interaction and navigation were influenced by different realities, users' goals, and users' preferences. Future research should improve interaction and navigation experiences based on the level of reality, individual goals, and preferences. Furthermore, hybrid user interfaces showed big potential in AR interaction and navigation, which should be taken into consideration with more investigation.

6.1 Summary: Multimodal Interaction in AR

In Chapter 3 we explored user-designed gestures and speech for interaction with AR HMD. A multi-object AR prototype was developed and used for the elicitation study. With a Wizard of Oz approach, we showed the participants text-based prompts (referents) with the AR display, to elicit their proposal to accomplish the task, then showed them the animation as the result of the interaction. In this way, participants feel that their proposal manipulated the object. Based on previous elicitation research in AR, we generated 22 referents including interaction of abstract manipulation, scale, translation, and rotation. The study has three conditions, which include both multimodal and unimodal interaction. The proposed gestures and speech were collected from 24 participants, and after binning and analysis, we presented the top three proposed gestures for each referent. These results offer future researchers references for designing the gestures when considering user preferences. We also compared our results to previous elicitation work in a single-object

AR environment and discussed how participants tended to perform more real-life-like interactions in a multi-object AR scenario than in a single-object AR scenario. Based on our video coding results, we also present the syntax format of the speech proposal in both unimodal and multimodal interaction. This can be used as guidance for future speech interaction technique design. Our findings also included the association between gestures and speech in multimodal interaction, which can be beneficial for increasing the accuracy and detection efficiency of the gesture or speech recognition model.

Later, we presented a guideline for the gesture binning procedure, which is a critical step during proposal analysis. Using our study results as an example, we generated the formula and walked through each decision-making step during the gesture binning process. We also presented our approach for binning variations of the gesture, and hopefully, this will lead to future work to continue perfecting the binning guideline for the elicitation study.

Nevertheless, designing prompts for elicitation studies is always a complicated decision due to the bias caused by using either animation or text. Using text-based prompts has influenced the proposed speech in our study, as the user will read the prompts out loud as a response to the referent. Utilizing other modalities, such as audio for prompts, can be a potential solution for future elicitation research, or showing the start and end position of the object as a prompt to the participants.

6.2 Summary: Cross-Reality for Sensemaking

In Chapter 4 we explored users' interaction and navigation strategy during sensemaking with an immersive analytics platform. An immersive analytics prototype was developed in both AR and VR and used for a controlled study with 40 participants. We used a within-subject design for the experiment, and each participant performed the task in both AR and VR environments, with the same mixed-reality headset. Participants answered 6 questions in each environment and analyzed the dataset with 3D data visualization and analysis tools that they could manipulate. The results did not show significant differences in user performance. However, differences were found between workloads perceived by participants. The findings showed that even with a similar amount of perceived mental demand and effort, participants have a higher tolerance for frustration in VR compared to AR. This means that a complicated sensemaking system should be considered to implement in VR or provide options for users to choose to work in VR. Furthermore, we discovered that participants showed more walking behavior in VR; however, they reached further locations when they walked in AR. This verified the feedback from participants that more safeness and naturalness they felt in AR. Nevertheless, the consistency and immersion that VR provides are well-received by participants. All findings suggested that both realities can be very helpful and have their benefits and drawbacks, and exploring cross-reality is a promising direction for future work to facilitate users' sensemaking process. Lastly, the observation and users' feedback on interaction provides insight into improving user experience by adopting hybrid user interfaces, including both conventional 2D user interfaces and innovative 3D user interfaces. With familiarity, users can adapt to mixed-reality technology easily without an entry barrier caused by uneasy-to-use hand gestures or controllers.

6.3 Summary: Hybrid User Interface for Decision-making

In Chapter 5, we explored using a hybrid user interface to assist users' decision-making in situated analytics scenarios. A situated analytics prototype was designed and developed in VR, which aims to simulate a physical environment with AR visualization. A tablet was used as a 2D user interface that users can interact with and display the result of all user interactions. Users can interact with virtual objects which represent the referents in a situated analytics scenario. A visualization was displayed close to each referent which also changes along with users' interaction to the according referent. An exploratory study was conducted with 19 participants, including 12 experienced users and 9 inexperienced users. Users freely explore the space with physical walking and having a mini-map on the tablet to assist navigation. The task is to optimize the lighting based on the visualization of hybrid user interfaces and personal preferences. The results showed various approaches and strategies for using HUI during decision-making. While users have positive feedback about the system's usability, they have different goals and expectations for the HUI in this situated analytics scenario. This finding demonstrated the necessity of adopting HUI with mixed-reality HMD and the flexibility it provides to meet individual needs.

6.4 Future Research Directions

In this section, we conclude the thesis by discussing possible future research directions based on the findings and the insights we have gained from the three studies in this thesis.

6.4.1 Hybrid User Interface

The results of Chapter 5 showed a promising use case with HUI in mixed reality. It can be used to facilitate interaction efficiency by deploying heavy inaction to the conventional 2D user interface, which mitigates the learning curve for users to use MR HMD. It also offloads the visualization presentation in 3D space and eliminates the cluttered AR scenario. However, the distribution and layout of visualizations can vary depending on individual preferences and needs, as well as environment and context. Future work can explore how to use artificial intelligence models to help make decisions about the distribution of the visualization in real time based on the changes in factors. In terms of facilitating interaction with HUI, we can explore how to reduce the workload caused by context-switching between user interfaces, such as arm fatigue and content loss.

6.4.2 Cross-Reality system

The results of Chapter 4 demonstrated that different realities provide users with assistance and setbacks. With the popularity of mixed-reality headsets, it is plausible for users to switch between different realities in a soon to come future. To this end, we should develop cross-reality systems with smooth transition techniques, so that users can choose between realities based on their needs. We can also include artificial intelligence into the picture and train the model to recognize situations requiring a transition between realities. Inevitably, the transition between realities could induce context-switching costs, which means a high mental workload for users to connect the environments and relocate themselves, not to mention if the task in hand strongly relies on the content. Using HUI to maintain the mental model between realities can be a potential solution to facilitate the transition procedure.

Bibliography

[1] Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*, Advances in Psychology, 52, pages 139–183. North-Holland, Oxford, England, 1988.

[2] Ruth B. Ekstrom. *Manual for kit of factor referenced cognitive tests*. Educational Testing Service, 1996.

[3] Lee Lisle, Kylie Davidson, Edward J. K. Gitre, Chris North, and Doug A. Bowman. Different realities: a comparison of augmented and virtual reality for the sensemaking process. *Frontiers in Virtual Reality*, 4, 2023.

[4] Adam Sinclair Williams Xiaoyan Zhou and Other Francisco Raul Ortega. Eliciting multimodal gesture+speech interactions in a multi-object augmented reality environment. *28th ACM Symposium on Virtual Reality Software and Technology (VRST '22), November 29-December 1, 2022, Tsukuba, Japan*, November 2022.

[5] Xiaoyan Zhou, Adam S. Williams, and Francisco R. Ortega. Towards establishing consistent proposal binning methods for unimodal and multimodal interaction elicitation studies. In Masaaki Kurosu, editor, *Human-Computer Interaction. Theoretical Approaches and Design Methods*, pages 356–368, Cham, 2022. Springer International Publishing.

[6] Xiaoyan Zhou, Anil Ufuk Batmaz, Adam S. Williams, Dylan Schreiber, and Francisco Ortega. I did not notice: A comparison of immersive analytics with augmented and virtual reality. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI EA '24, New York, NY, USA, 2024. Association for Computing Machinery.

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

[8] Carolina Cruz-Neira, Daniel J Sandin, Thomas A DeFanti, Robert V Kenyon, and John C Hart. The cave: Audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–73, 1992.

[9] Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti. *Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE*. Association for Computing Machinery, New York, NY, USA, 1 edition, 2023.

[10] OculusRift. https://www.kickstarter.com/projects/1523379957/oculus-rift-step-into-the-game.

[11] Hololens 2. https://www.microsoft.com/en-us/hololens/hardwaredocument-experiences.

[12] Edwin L Hutchins, James D Hollan, and Donald A Norman. Direct manipulation interfaces. *Human–computer interaction*, 1(4):311–338, 1985.

[13] Yvonne Rogers. 2012.

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

[15] Kim Marriott, Falk Schreiber, Tim Dwyer, Karsten Klein, Nathalie Henry Riche, Takayuki Itoh, Wolfgang Stuerzlinger, and Bruce H Thomas. *Immersive Analytics*, volume 11190. Springer, 2018.

[16] Gunnar Borg. *Borg's perceived exertion and pain scales*. Human kinetics, 1998.

[17] Ken Hinckley, Robert JK Jacob, Colin Ware, Jacob O Wobbrock, and Daniel Wigdor. Input/output devices and interaction techniques., 2014.

[18] Eugenie Brasier, Olivier Chapuis, Nicolas Ferey, Jeanne Vezien, and Caroline Appert. Arpads: Mid-air indirect input for augmented reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 332–343, 2020.

[19] Florian Müller, Joshua McManus, Sebastian Günther, Martin Schmitz, Max Mühlhäuser, and Markus Funk. Mind the tap: Assessing foot-taps for interacting with head-mounted displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–13, New York, NY, USA, 2019. Association for Computing Machinery.

[20] Yukang Yan, Chun Yu, Xin Yi, and Yuanchun Shi. Headgesture: Hands-free input approach leveraging head movements for hmd devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(4), dec 2018.

[21] Wilder Penfield and Theodore Rasmussen. The cerebral cortex of man; a clinical study of localization of function. 1950.

[22] Barrett Ens, Aaron Quigley, Hui-Shyong Yeo, Pourang Irani, Thammathip Piumsomboon, and Mark Billinghurst. Counterpoint: Exploring mixed-scale gesture interaction for ar applications. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, page 1–6, New York, NY, USA, 2018. Association for Computing Machinery.

[23] Klen Čopič Pucihar, Christian Sandor, Matjaž Kljun, Wolfgang Huerst, Alexander Plopski, Takafumi Taketomi, Hirokazu Kato, and Luis A. Leiva. The missing interface: Micro-gestures on augmented objects. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, page 1–6, New York, NY, USA, 2019. Association for Computing Machinery.

[24] Barrett Ens, Benjamin Bach, Maxime Cordeil, Ulrich Engelke, Marcos Serrano, Wesley Willett, Arnaud Prouzeau, Christoph Anthes, Wolfgang Büschel, Cody Dunne, Tim Dwyer, Jens Grubert, Jason H. Haga, Nurit Kirshenbaum, Dylan Kobayashi, Tica Lin, Monsurat Olaosebikan, Fabian Pointecker, David Saffo, Nazmus Saquib, Dieter Schmalstieg, Danielle Albers Szafir, Matt Whitlock, and Yalong Yang. Grand challenges in immersive

analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[25] Joseph J LaViola Jr, Sarah Buchanan, and Corey Pittman. Multimodal input for perceptual user interfaces. *Interactive Displays: Natural Human-Interface Technologies*, pages 285–312, 2014.

[26] Jean claude Martin. Tycoon: Theoretical framework and software tools for multimodal interfaces. 1997.

[27] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, and E. Wilson. Immersive analytics. In *2015 Big Data Visual Analytics (BDVA)*, pages 1–8, 2015.

[28] Adrien Fonnet and Yannick Prie. Survey of immersive analytics. *IEEE transactions on visualization and computer graphics*, 27(3):2101–2122, 2019.

[29] Sophie Stellmach, Markus Jüttner, Christian Nywelt, Jens Schneider, and Raimund Dachselt. Investigating freehand pan and zoom. In *Mensch & computer*, pages 303–312. Oldenbourg Verlag, 2012.

[30] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. Pinpointing: Precise head- and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–14, New York, NY, USA, 2018. Association for Computing Machinery.

[31] P. R. Cohen, M. Dalrymple, D. B. Moran, F. C. Pereira, and J. W. Sullivan. Synergistic use of direct manipulation and natural language. *SIGCHI Bull.*, 20(SI):227–233, mar 1989.

[32] Ed Kaiser, Alex Olwal, David McGee, Hrvoje Benko, Andrea Corradini, Xiaoguang Li, Phil Cohen, and Steven Feiner. Mutual disambiguation of 3d multimodal interaction in

augmented and virtual reality. In *Proceedings of the 5th International Conference on Multimodal Interfaces*, ICMI ’03, page 12–19, New York, NY, USA, 2003. Association for Computing Machinery.

[33] S. Oviatt, B. Schuller, P. Cohen, D. Sonntag, and G. Potamianos. *The Handbook of Multimodal-Multisensor Interfaces, Volume 1: Foundations, User Modeling, and Common Modality Combinations*. ACM Books. Association for Computing Machinery and Morgan & Claypool Publishers, 2017.

[34] Sharon Oviatt. Mutual disambiguation of recognition errors in a multimodel architecture. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’99, page 576–583, New York, NY, USA, 1999. Association for Computing Machinery.

[35] Sharon Oviatt and Philip Cohen. Perceptual user interfaces: Multimodal interfaces that process what comes naturally. *Commun. ACM*, 43(3):45–53, mar 2000.

[36] T. Piumsomboon, D. Altimira, H. Kim, A. Clark, G. Lee, and M. Billinghurst. 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)*, pages 73–82, Munich, Germany, 2014. IEEE.

[37] M. Kolsch, R. Bane, T. Hollerer, and M. Turk. Multimodal interaction with a wearable augmented reality system. *IEEE Computer Graphics and Applications*, 26(3):62–71, 2006.

[38] Minkyung Lee, Mark Billinghurst, Woonhyuk Baek, Richard Green, and Woontack Woo. A usability study of multimodal input in an augmented reality environment. *Virtual Real.*, 17(4):293–305, November 2013.

[39] Taehee Lee and Tobias Hollerer. Handy ar: Markerless inspection of augmented reality objects using fingertip tracking. In *2007 11th IEEE International Symposium on Wearable Computers*, pages 83–90, 2007.

[40] Otmar Hilliges, David Kim, Shahram Izadi, Malte Weiss, and Andrew Wilson. Holodesk: Direct 3d interactions with a situated see-through display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 2421–2430, New York, NY, USA, 2012. Association for Computing Machinery.

[41] Hrvoje Benko, Ricardo Jota, and Andrew Wilson. Miragetable: Freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 199–208, New York, NY, USA, 2012. Association for Computing Machinery.

[42] Lode Vanacken, Tovi Grossman, and Karin Coninx. Exploring the effects of environment density and target visibility on object selection in 3d virtual environments. In *2007 IEEE Symposium on 3D User Interfaces*, 2007.

[43] Yiqin Lu, Chun Yu, and Yuanchun Shi. Investigating bubble mechanism for ray-casting to improve 3d target acquisition in virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 35–43, 2020.

[44] Douglas A Bowman. *Interaction techniques for common tasks in immersive virtual environments: design, evaluation, and application*. Georgia Institute of Technology, 1999.

[45] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. The go-go interaction technique: Non-linear mapping for direct manipulation in vr. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*, UIST '96, page 79–80, New York, NY, USA, 1996. Association for Computing Machinery.

[46] Regis Kopper, Felipe Bacim, and Doug A. Bowman. Rapid and accurate 3d selection by progressive refinement. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 67–74, 2011.

[47] Jeffrey Cashion, Chadwick Wingrave, and Joseph J. LaViola Jr. Dense and dynamic 3d selection for game-based virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):634–642, 2012.

[48] Felipe Bacim de Araujo e Silva. Increasing selection accuracy and speed through progressive refinement. 2015.

[49] Tovi Grossman and Ravin Balakrishnan. The bubble cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, page 281–290, New York, NY, USA, 2005. Association for Computing Machinery.

[50] Roman Rädle, Hans-Christian Jetter, Simon Butscher, and Harald Reiterer. The effect of egocentric body movements on users' navigation performance and spatial memory in zoomable user interfaces. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '13, page 23–32, New York, NY, USA, 2013. Association for Computing Machinery.

[51] Sarah S. Chance, Florence Gaunet, Andrew C. Beall, and Jack M. Loomis. Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. *Presence*, 7(2):168–178, 1998.

[52] Tovi Grossman and Ravin Balakrishnan. The design and evaluation of selection techniques for 3d volumetric displays. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, page 3–12, New York, NY, USA, 2006. Association for Computing Machinery.

[53] Steven Feiner and Ari Shamash. Hybrid user interfaces: Breeding virtually bigger interfaces for physically smaller computers. In *Proceedings of the 4th Annual ACM Symposium on User Interface Software and Technology - UIST '91*, pages 9–17, Hilton Head, South Carolina, United States, 1991. ACM Press.

[54] Pascal Knierim, Dimitri Hein, Albrecht Schmidt, and Thomas Kosch. The smartphone controller. *i-com*, 20(1):49–61, 2021.

[55] Bongshin Lee, Arjun Srinivasan, John Stasko, Melanie Tory, and Vidya Setlur. Multimodal interaction for data visualization. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, AVI '18, New York, NY, USA, 2018. Association for Computing Machinery.

[56] Erwan Normand and Michael J. McGuffin. Enlarging a smartphone with ar to create a handheld vesad (virtually extended screen-aligned display). In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 123–133, 2018.

[57] Katja Vock, Sebastian Hubenschmid, Johannes Zagermann, Simon Butscher, and Harald Reiterer. Idiar: Augmented reality dashboards to supervise mobile intervention studies. In *Proceedings of Mensch Und Computer 2021*, MuC '21, page 248–259, New York, NY, USA, 2021. Association for Computing Machinery.

[58] Fengyuan Zhu and Tovi Grossman. Bishare: Exploring bidirectional interactions between smartphones and head-mounted augmented reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–14, New York, NY, USA, 2020. Association for Computing Machinery.

[59] Jens Grubert, Matthias Heinisch, Aaron Quigley, and Dieter Schmalstieg. Multifi: Multi fidelity interaction with displays on and around the body. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 3933–3942, New York, NY, USA, 2015. Association for Computing Machinery.

[60] Rahul Arora, Rubaiat Habib Kazi, Tovi Grossman, George Fitzmaurice, and Karan Singh. Symbiosissketch: Combining 2d & 3d sketching for designing detailed 3d objects in situ. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–15, New York, NY, USA, 2018. Association for Computing Machinery.

[61] Francesco Riccardo Di Gioia, Eugenie Brasier, Emmanuel Pietriga, and Caroline Appert. Investigating the use of ar glasses for content annotation on mobile devices. *Proc. ACM Hum.-Comput. Interact.*, 6(ISS), nov 2022.

[62] Tobias Drey, Jan Gugenheimer, Julian Karlbauer, Maximilian Milo, and Enrico Rukzio. Vrsketchin: Exploring the design space of pen and tablet interaction for 3d sketching in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–14, New York, NY, USA, 2020. Association for Computing Machinery.

[63] Sebastian Hubenschmid, Johannes Zagermann, Simon Butscher, and Harald Reiterer. Stream: Exploring the combination of spatially-aware tablets with augmented reality head-mounted displays for immersive analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2021.

[64] Ricardo Langner, Marc Satkowski, Wolfgang Büschel, and Raimund Dachselt. MARVIS: Combining Mobile Devices and Augmented Reality for Visual Data Analysis. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–17, Yokohama Japan, May 2021. ACM.

[65] Weizhou Luo, Eva Goebel, Patrick Reipschläger, Mats Ole Ellenberg, and Raimund Dachselt. Exploring and slicing volumetric medical data in augmented reality using a spatially-aware mobile device. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 334–339, 2021.

[66] Mohammad Rajabi Seraji and Wolfgang Stuerzlinger. Hybridaxes: An immersive analytics tool with interoperability between 2d and immersive reality modes. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 155–160, 2022.

[67] Hemant Bhaskar Surale, Aakar Gupta, Mark Hancock, and Daniel Vogel. Tabletinvr: Exploring the design space for using a multi-touch tablet in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–13, New York, NY, USA, 2019. Association for Computing Machinery.

[68] Sebastian Hubenschmid, Jonathan Wieland, Daniel Immanuel Fink, Andrea Batch, Johannes Zagermann, Niklas Elmqvist, and Harald Reiterer. ReLive: Bridging In-Situ and Ex-Situ Visual Analytics for Analyzing Mixed Reality User Studies. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, pages 1–20, New York, NY, USA, April 2022. Association for Computing Machinery.

[69] Sebastian Hubenschmid, Johannes Zagermann, Daniel Fink, Jonathan Wieland, Tiare Feuchtner, and Harald Reiterer. Towards asynchronous hybrid user interfaces for cross-reality interaction. 2021.

[70] Patrick Reipschlager, Tamara Flemisch, and Raimund Dachselt. Personal augmented reality for information visualization on large interactive displays. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1182–1192, 2021.

[71] Tianchen Sun, Yucong Ye, Issei Fujishiro, and Kwan-Liu Ma. Collaborative visual analysis with multi-level information sharing using a wall-size display and see-through hmds. In *2019 IEEE Pacific Visualization Symposium (PacificVis)*, pages 11–20, 2019.

[72] Verena Biener, Daniel Schneider, Travis Gesslein, Alexander Otte, Bastian Kuth, Per Ola Kristensson, Eyal Ofek, Michel Pahud, and Jens Grubert. Breaking the screen: Interaction across touchscreen boundaries in virtual reality for mobile knowledge workers. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3490–3502, 2020.

[73] Simon Butscher, Sebastian Hubenschmid, Jens Müller, Johannes Fuchs, and Harald Reiterer. Clusters, trends, and outliers: How immersive technologies can facilitate the collaborative analysis of multidimensional data. In *Proceedings of the 2018 CHI Conference on*

Human Factors in Computing Systems, CHI '18, page 1–12, New York, NY, USA, 2018. Association for Computing Machinery.

[74] Patrick Reipschläger and Raimund Dachselt. Designar: Immersive 3d-modeling combining augmented reality with interactive displays. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, ISS '19, page 29–41, New York, NY, USA, 2019. Association for Computing Machinery.

[75] Maurício Sousa, Daniel Mendes, Soraia Paulo, Nuno Matela, Joaquim Jorge, and Daniel Simões Lopes. Vrrrroom: Virtual reality for radiologists in the reading room. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 4057–4062, New York, NY, USA, 2017. Association for Computing Machinery.

[76] Sebastian Hubenschmid, Johannes Zagermann, Daniel Leicht, Harald Reiterer, and Tiare Feuchtner. Around the smartphone: Investigating the effects of virtually-extended display size on spatial memory. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[77] Abdelrahman Zaky, Johannes Zagermann, Harald Reiterer, and Tiare Feuchtner. Opportunities and challenges of hybrid user interfaces for optimization of mixed reality interfaces. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 215–219, 2023.

[78] K. Marriott, F. Schreiber, T. Dwyer, K. Klein, N. Henry Riche, T. Itoh, W. Stuerzlinger, and B. H. Thomas. *Immersive Analytics*. Lecture Notes in Computer Science. Springer International Publishing, 2018.

[79] R. Sicat, J. Li, J. Choi, M. Cordeil, W. Jeong, B. Bach, and H. Pfister. Dxr: A toolkit for building immersive data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):715–725, 2019.

[80] M. Cordeil, A. Cunningham, B. Bach, C. Hurter, B. H. Thomas, K. Marriott, and T. Dwyer. Iatk: An immersive analytics toolkit. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 200–209, 2019.

[81] Jorge A. Wagner Filho, Wolfgang Stuerzlinger, and Luciana Nedel. Evaluating an immersive space-time cube geovisualization for intuitive trajectory data exploration. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):514–524, 2020.

[82] Natalie Hube and Mathias Müller. The data in your hands: Exploring novel interaction techniques and data visualization approaches for immersive data analytics. In *VisBIA@AVI*, 2018.

[83] Christophe Hurter, Nathalie Henry Riche, Steven M. Drucker, Maxime Cordeil, Richard Alligier, and Romain Vuillemot. Fiberclay: Sculpting three dimensional trajectories to reveal structural insights. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):704–714, 2019.

[84] M. Kraus, N. Weiler, D. Oelke, J. Kehrer, D. A. Keim, and J. Fuchs. The impact of immersion on cluster identification tasks. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):525–535, 2020.

[85] Mark Simpson, Jiayan Zhao, and Alexander Klippel. Take a walk: Evaluating movement types for data visualization in immersive virtual reality. In *Proceedings of Workshop on Immersive Analytics (IA)*, pages 1–6, 2017.

[86] Martin Luboschik, Philip Berger, and Oliver Staadt. On spatial perception issues in augmented reality based immersive analytics. In *Proceedings of the 2016 ACM Companion on Interactive Surfaces and Spaces*, ISS '16 Companion, page 47–53, New York, NY, USA, 2016. Association for Computing Machinery.

[87] Wolfgang Büschel, Patrick Reipschläger, Ricardo Langner, and Raimund Dachselt. Investigating the use of spatial interaction for 3d data visualization on mobile devices. In *Proceed-*

ings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, ISS '17, page 62–71, New York, NY, USA, 2017. Association for Computing Machinery.

[88] Benjamin Bach, Ronell Sicat, Johanna Beyer, Maxime Cordeil, and Hanspeter Pfister. The hologram in my hand: How effective is interactive exploration of 3d visualizations in immersive tangible augmented reality? *IEEE Transactions on Visualization and Computer Graphics*, 24(1):457–467, 2018.

[89] Neven ElSayed, Bruce Thomas, Kim Marriott, Julia Piantadosi, and Ross Smith. Situated Analytics. In *2015 Big Data Visual Analytics (BDVA)*, pages 1–8, September 2015.

[90] Bruce H. Thomas, Gregory F. Welch, Pierre Dragicevic, Niklas Elmqvist, Pourang Irani, Yvonne Jansen, Dieter Schmalstieg, Aurélien Tabard, Neven A. M. ElSayed, Ross T. Smith, and Wesley Willett. *Situated Analytics*, pages 185–220. Springer International Publishing, Cham, 2018.

[91] Neven A. M. ElSayed, Bruce H. Thomas, Kim Marriott, Julia Piantadosi, and Ross T. Smith. Situated Analytics: Demonstrating immersive analytical tools with Augmented Reality. *Journal of Visual Languages & Computing*, 36:13–23, October 2016.

[92] Sungbok Shin, Andrea Batch, Peter W. S. Butcher, Panagiotis D. Ritsos, and Niklas Elmqvist. The reality of the situation: A survey of situated analytics. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–19, 2023.

[93] Yalong Yang, Maxime Cordeil, Johanna Beyer, Tim Dwyer, Kim Marriott, and Hanspeter Pfister. Embodied navigation in immersive abstract data visualization: Is overview+ detail or zooming better for 3d scatterplots? *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1214–1224, 2020.

[94] Matt Whitlock, Stephen Smart, and Danielle Albers Szafir. Graphical perception for immersive analytics. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 616–625. IEEE, 2020.

[95] Fengfeng Zhang, Long Chen, Wenhua Miao, and Lining Sun. Research on accuracy of augmented reality surgical navigation system based on multi-view virtual and real registration technology. *IEEE Access*, 8:122511–122528, 2020.

[96] Adam Drogemuller, Andrew Cunningham, James Walsh, Maxime Cordeil, William Ross, and Bruce Thomas. Evaluating navigation techniques for 3d graph visualizations in virtual reality. In *2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA)*, pages 1–10. IEEE, 2018.

[97] Jorge A Wagner Filho, Carla MDS Freitas, and Luciana Nedel. Comfortable immersive analytics with the virtualdesk metaphor. *IEEE computer graphics and applications*, 39(3):41–53, 2019.

[98] Johannes Sorger, Manuela Waldner, Wolfgang Knecht, and Alessio Arleo. Immersive analytics of large dynamic networks via overview and detail navigation. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 144–1447. IEEE, 2019.

[99] Jorge Wagner, Wolfgang Stuerzlinger, and Luciana Nedel. The effect of exploration mode and frame of reference in immersive analytics. *IEEE Transactions on Visualization and Computer Graphics*, 28(9):3252–3264, 2021.

[100] Daniel Zielasko, Sven Horn, Sebastian Freitag, Benjamin Weyers, and Torsten W Kuhlen. Evaluation of hands-free hmd-based navigation techniques for immersive data analysis. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 113–119. IEEE, 2016.

[101] Wolfgang Büschel, Annett Mitschick, Thomas Meyer, and Raimund Dachselt. Investigating smartphone-based pan and zoom in 3d data spaces in augmented reality. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–13, 2019.

[102] Raphaël James, Anastasia Bezerianos, Olivier Chapuis, Maxime Cordeil, Tim Dwyer, and Arnaud Prouzeau. Personal+ context navigation: combining ar and shared displays in network path-following. *arXiv preprint arXiv:2005.10612*, 2020.

[103] Ze Dong, Adrian Clark, Barrett Ens, and Thammathip Piumsomboon. Localization in immersive analytics with world-in-miniature, landmarks and constellations. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction*, pages 1–2, 2022.

[104] M. Kraus, N. Weiler, D. Oelke, J. Kehrer, D. A. Keim, and J. Fuchs. The impact of immersion on cluster identification tasks. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):525–535, 2020.

[105] Dmitry Nekrasovski, Adam Bodnar, Joanna McGrenere, François Guimbretière, and Tamara Munzner. An evaluation of pan amp; zoom and rubber sheet navigation with and without an overview. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, page 11–20, New York, NY, USA, 2006. Association for Computing Machinery.

[106] Yu Zhao, Jeanine Stefanucci, Sarah Creem-Regehr, and Bobby Bodenheimer. Evaluating augmented reality landmark cues and frame of reference displays with virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2710–2720, 2023.

[107] Kadek Ananta Satriadi, Barrett Ens, Maxime Cordeil, Bernhard Jenny, Tobias Czauderna, and Wesley Willett. Augmented reality map navigation with freehand gestures. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*, pages 593–603. IEEE, 2019.

[108] Shun Sekiguchi and Takashi Komuro. Volgrab: Realizing 3d view navigation by aerial hand gestures. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2064–2069, 2017.

[109] Sriram Karthik Badam, Arjun Srinivasan, Niklas Elmqvist, and John Stasko. Affordances of input modalities for visual data exploration in immersive environments. In *2nd Workshop on Immersive Analytics*, volume 2, 2017.

[110] Wolfgang Büschel, Jian Chen, Raimund Dachselt, Steven Drucker, Tim Dwyer, Carsten Görg, Tobias Isenberg, Andreas Kerren, Chris North, and Wolfgang Stuerzlinger. Interaction for immersive analytics. *Immersive analytics*, pages 95–138, 2018.

[111] Mark Mine. Virtual environment interaction techniques. 06 1995.

[112] Jorge Wagner, Wolfgang Stuerzlinger, and Luciana Nedel. Comparing and combining virtual hand and virtual ray pointer interactions for data manipulation in immersive analytics. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2513–2523, 2021.

[113] Joseph J. LaViola, Ernst Kruijff, Ryan P. McMahan, Doug A. Bowman, and Ivan Poupyrev. *3D user interfaces theory and practice*. Addison-Wesley, 2017.

[114] Wallace S Lages and Doug A Bowman. Move the object or move myself? walking vs. manipulation for the examination of 3d scientific data. *Frontiers in ICT*, 5:15, 2018.

[115] Christoph Anthes. Design considerations for cross-virtuality applications, 2022.

[116] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, volume 2351, pages 282–292. Spie, 1995.

[117] Adalberto L Simeone, Mohamed Khamis, Augusto Esteves, Florian Daiber, Matjaž Kljun, Klen Čopič Pucihar, Poika Isokoski, and Jan Gugenheimer. International workshop on cross-reality (xr) interaction. In *Companion Proceedings of the 2020 Conference on Interactive Surfaces and Spaces*, pages 111–114, 2020.

[118] Andreas Riegler, Christoph Anthes, Hans-Christian Jetter, Christoph Heinzl, Clemens Holzmann, Herbert Jodlbauer, Manuel Brunner, Stefan Auer, Judith Friedl, Bernhard Fröhler, et al. Cross-virtuality visualization, interaction and collaboration. *XR@ ISS*, 11, 2020.

[119] Arnaud Prouzeau, Antoine Lhuillier, Barrett Ens, Daniel Weiskopf, and Tim Dwyer. Visual link routing in immersive visualisations. In *Proceedings of the 2019 ACM international conference on interactive surfaces and spaces*, pages 241–253, 2019.

[120] Max Krichenbauer, Goshiro Yamamoto, Takafumi Taketom, Christian Sandor, and Hirokazu Kato. Augmented reality versus virtual reality for 3d object manipulation. *IEEE transactions on visualization and computer graphics*, 24(2):1038–1048, 2017.

[121] Victoria Interrante, Brian Ries, and Lee Anderson. Distance perception in immersive virtual environments, revisited. In *IEEE virtual reality conference (VR 2006)*, pages 3–10. IEEE, 2006.

[122] Joan Sol Roo, Jean Basset, Pierre-Antoine Cinquin, and Martin Hachet. Understanding users' capability to transfer information between mixed and virtual reality: Position estimation across modalities and perspectives. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[123] Richard Skarbez, Nicholas F Polys, J Todd Ogle, Chris North, and Doug A Bowman. Immersive analytics: Theory and research agenda. *Frontiers in Robotics and AI*, 6:82, 2019.

[124] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. The magicbook: a transitional ar interface. *Computers & Graphics*, 25(5):745–753, 2001.

[125] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. Magicbook: transitioning between reality and virtuality. In *CHI'01 extended abstracts on Human factors in computing systems*, pages 25–26, 2001.

[126] Kiyoshi Kiyokawa, Haruo Takemura, and Naokazu Yokoya. A collaboration support technique by integrating a shared virtual reality and a shared augmented reality. In *IEEE SMC'99*

Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028), volume 6, pages 48–53. IEEE, 1999.

[127] Hrvoje Benko, Edward W Ishak, and Steven Feiner. Collaborative mixed reality visualization of an archaeological excavation. In *Third IEEE and ACM international symposium on mixed and augmented reality*, pages 132–140. IEEE, 2004.

[128] Joan Sol Roo and Martin Hachet. One reality: Augmenting how the physical world is experienced by combining multiple mixed reality modalities. In *Proceedings of the 30th annual ACM symposium on user interface software and technology*, pages 787–795, 2017.

[129] Raphael Grasset, Julian Looser, and Mark Billinghurst. Transitional interface: concept, issues and framework. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 231–232. IEEE, 2006.

[130] Raphaël Grasset, Andreas Dünser, and Mark Billinghurst. Moving between contexts-a user evaluation of a transitional interface. 2008.

[131] Felipe G Carvalho, Daniela G Trevisan, and Alberto Raposo. Toward the design of transitional interfaces: an exploratory study on a semi-immersive hybrid user interface. *Virtual Reality*, 16:271–288, 2012.

[132] Fabian Pointecker, Hans-Christian Jetter, and Christoph Anthes. Exploration of visual transitions between virtual and augmented reality. In *Workshop on Immersive Analytics: Envisioning Future Productivity for Immersive Analytics//@ CHI*, 2020.

[133] Hans-Christian Jetter, Jan-Henrik Schröder, Jan Gugenheimer, Mark Billinghurst, Christoph Anthes, Mohamed Khamis, and Tiare Feuchtner. Transitional interfaces in mixed and cross-reality: A new frontier? In *Companion Proceedings of the 2021 Conference on Interactive Surfaces and Spaces*, pages 46–49, 2021.

[134] Bernhard Fröhler, Christoph Anthes, Fabian Pointecker, Judith Friedl, Daniel Schwajda, Andreas Riegler, Shailesh Tripathi, Clemens Holzmann, Manuel Brunner, Herbert Jodlbauer,

et al. A survey on cross-virtuality analytics. In *Computer Graphics Forum*, volume 41, pages 465–494. Wiley Online Library, 2022.

[135] Judith Friedl, Janine Mayer, and Christoph Anthes. Germane elements for the evaluation of transitional interfaces. In *ISS'21: Interactive Surfaces and Spaces*, 2021.

[136] Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum. A procedure for developing intuitive and ergonomic gesture interfaces for hci. In Antonio Camurri and Gualtiero Volpe, editors, *Gesture-Based Communication in Human-Computer Interaction*, pages 409–420, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[137] Niloofar Dezfuli, Mohammadreza Khalilbeigi, Max Mühlhäuser, and David Geerts. A study on interpersonal relationships for social interactive television. In *Proceedings of the 9th European Conference on Interactive TV and Video*, EuroITV '11, page 21–24, New York, NY, USA, 2011. Association for Computing Machinery.

[138] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1083–1092, New York, NY, USA, 2009. ACM.

[139] Muhammad Zeeshan Baig and Manolya Kavakli. Qualitative analysis of a multimodal interface system using speech/gesture. In *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 2811–2816, Wuhan, China, 2018. IEEE, IEEE.

[140] A A Karpov and R M Yusupov. Multimodal interfaces of Human–Computer interaction. *Her. Russ. Acad. Sci.*, 88(1):67–74, January 2018.

[141] Meredith Ringel Morris. Web on the wall: Insights from a multimodal interaction elicitation study. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '12, pages 95–104, New York, NY, USA, 2012. ACM.

[142] Jaime Ruiz, Yang Li, and Edward Lank. User-defined motion gestures for mobile interaction, 2011.

[143] Aurélie Cohé and Martin Hachet. Understanding user gestures for manipulating 3d objects from touchscreen inputs. In *Proceedings of Graphics Interface 2012*, GI '12, pages 157–164, Toronto, Ont., Canada, Canada, 2012. Canadian Information Processing Society.

[144] Panagiotis Vogiatzidakis and Panayiotis Koutsabasis. 'address and command': Two-handed mid-air interactions with multiple home devices. *International Journal of Human-Computer Studies*, 159:102755, 2022.

[145] Adam S. Williams, Jason Garcia, and Francisco Ortega. Understanding multimodal user gesture and speech behavior for object manipulation in augmented reality using elicitation. *IEEE Transactions on Visualization & Computer Graphics*, 2020 In press.

[146] Minkyung Lee and Mark Billinghurst. A wizard of oz study for an ar multimodal interface. In *Proceedings of the 10th International Conference on Multimodal Interfaces*, ICMI '08, page 249–256, New York, NY, USA, 2008. Association for Computing Machinery.

[147] Adam S. Williams, Jason Garcia, and Francisco Ortega. Understanding gesture and speech multimodal interactions for manipulation tasks in augmented reality using unconstrained elicitation. *Pre-Print*, 2020.

[148] Radu-Daniel Vatavu and Jacob O. Wobbrock. Clarifying agreement calculations and analysis for end-user elicitation studies. *ACM Trans. Comput.-Hum. Interact.*, 29(1), jan 2022.

[149] Tran Pham, Jo Vermeulen, Anthony Tang, and Lindsay MacDonald Vermeulen. Scale impacts elicited gestures for manipulating holograms: Implications for AR gesture design. In *Proceedings of the 2018 Designing Interactive Systems Conference*, pages 227–240. ACM, June 2018.

[150] Jacob O Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A Myers. Maximizing the guessability of symbolic input, 2005.

[151] Miguel A. Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. Memorability of pre-designed and user-defined gesture sets. In *Proceedings of the SIGCHI Conference*

on Human Factors in Computing Systems, CHI '13, page 1099–1108, New York, NY, USA, 2013. Association for Computing Machinery.

[152] Meredith Ringel Morris, Jacob O Wobbrock, and Andrew D Wilson. Understanding users' preferences for surface gestures. In *Proceedings of graphics interface 2010*, pages 261–268. Canadian Information Processing Society, 2010.

[153] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob A Wobbrock. A systematic review of gesture elicitation studies: What can we learn from 216 studies. In *Proceedings of ACM Int. Conf. on Designing Interactive Systems (DIS'20)*, page NA, Eindhoven, 2020. ACM Press.

[154] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. User-defined gestures for augmented reality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, page 955–960, New York, NY, USA, 2013. Association for Computing Machinery.

[155] Markus L Wittorf and Mikkel R Jakobsen. Eliciting Mid-Air gestures for Wall-Display interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, NordiCHI '16, pages 3:1–3:4, New York, NY, USA, 2016. ACM.

[156] Andreea Danielescu and David Piorkowski. Iterative design of gestures during elicitation: Understanding the role of increased production, Jan 2022.

[157] F. R. Ortega, A. Galvan, K. Tarre, A. Barreto, N. Rishe, J. Bernal, R. Balcazar, and J. Thomas. Gesture elicitation for 3d travel via multi-touch and mid-air systems for procedurally generated pseudo-universe. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 144–153, Los Angeles, CA, USA, 2017. IEEE.

[158] Theophanis Tsandilas. Fallacies of agreement: A critical review of consensus assessment methods for gesture elicitation. *ACM Trans. Comput. Hum. Interact.*, 25(3):18, June 2018.

[159] Adam S Williams and Francisco Ortega. Insights on visual aid and study design for gesture interaction in limited sensor range augmented reality devices. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 19–22, 2020.

[160] Radu-Daniel Vatavu and Jacob O. Wobbrock. Formalizing agreement analysis for elicitation studies: New measures, significance test, and toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 1325–1334, New York, NY, USA, 2015. Association for Computing Machinery.

[161] Adam S. Williams and Francisco R. Ortega. A concise guide to elicitation methodology, May 2021.

[162] David Mcneill. *Gesture and Thought*. the University of Chicago Press, USA, 01 2005.

[163] Marie-Luce Bourguet and Akio Ando. Synchronization of speech and hand gestures during multimodal human-computer interaction. In *CHI 98 Conference Summary on Human Factors in Computing Systems*, CHI '98, page 241–242, New York, NY, USA, 1998. Association for Computing Machinery.

[164] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, M c Schraefel, and Jacob O Wobbrock. Reducing legacy bias in gesture elicitation studies. *Interactions*, 21(3):40–45, May 2014.

[165] SHENG KAI TANG and David Coulter. Start gesture - mixed reality, 2022.

[166] Giulia Wally Scurati, Michele Gattullo, Michele Fiorentino, Francesco Ferrise, Monica Bordegoni, and Antonio Emmanuele Uva. Converting maintenance actions into standard symbols for augmented reality applications in industry 4.0. *Computers in Industry*, 98:68–79, 2018.

[167] Xiaoyan Zhou, Adam S. Williams, and Francisco R. Ortega. Towards establishing consistent proposal binning methods for unimodal and multimodal interaction elicitation studies. In

Masaaki Kurosu, editor, *Human-Computer Interaction. Theoretical Approaches and Design Methods*, pages 356–368, Cham, 2022. Springer International Publishing.

[168] R Balakrishnan and K Hinckley. Symmetric bimanual interaction. *GROUP ACM SIGCHI Int. Conf. Support. Group Work*, 2000.

[169] Robert J.K. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. Reality-based interaction: A framework for post-wimp interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, page 201–210, New York, NY, USA, 2008. Association for Computing Machinery.

[170] Adam S. Williams and Francisco R. Ortega. Evolutionary gestures: When a gesture is not quite legacy biased. *Interactions*, 27(5):50–53, sep 2020.

[171] Sumbul Khan and Bige Tunçer. Gesture and speech elicitation for 3d cad modeling in conceptual design. *Automation in Construction*, 106:102847, 2019.

[172] Teddy Seyed, Chris Burns, Mario Costa Sousa, Frank Maurer, and Anthony Tang. Eliciting usable gestures for multi-display environments. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '12, page 41–50, New York, NY, USA, 2012. Association for Computing Machinery.

[173] Adam Williams and Francisco Ortega. A concise guide to elicitation methodology, 05 2021.

[174] Abdullah X. Ali, Meredith Ringel Morris, and Jacob O. Wobbrock. Crowdsourcing similarity judgments for agreement analysis in end-user elicitation studies. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, page 177–188, New York, NY, USA, 2018. Association for Computing Machinery.

[175] Michael Nebeling, Alexander Huber, David Ott, and Moira C. Norrie. Web on the wall reloaded: Implementation, replication and refinement of user-defined interaction sets. In

Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces, ITS '14, page 15–24, New York, NY, USA, 2014. Association for Computing Machinery.

[176] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O. Wobbrock. A systematic review of gesture elicitation studies: What can we learn from 216 studies? In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, page 855–872, New York, NY, USA, 2020. Association for Computing Machinery.

[177] Radu-Daniel Vatavu. User-defined gestures for free-hand tv control. In *Proceedings of the 10th European Conference on Interactive TV and Video*, EuroITV '12, page 45–48, New York, NY, USA, 2012. Association for Computing Machinery.

[178] Hoo Yong Leng, Noris Mohd Norowi, and Azrul Hazri Jantan. A user-defined gesture set for music interaction in immersive virtual environment. In *Proceedings of the 3rd International Conference on Human-Computer Interaction and User Experience in Indonesia*, CHIuXiD '17, page 44–51, New York, NY, USA, 2017. Association for Computing Machinery.

[179] Ze Dong, Thammathip Piumsomboon, Jingjing Zhang, Adrian Clark, Huidong Bai, and Rob Lindeman. 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*, CHI EA '20, page 1–8, New York, NY, USA, 2020. Association for Computing Machinery.

[180] Zhuochao Peng and Jiaxin Xu. User-defined gestures for taking self-portraits with smartphone based on consistency. In Tareq Ahram and Christianne Falcão, editors, *Advances in Usability, User Experience, Wearable and Assistive Technology*, pages 316–326, Cham, 2020. Springer International Publishing.

[181] Radu-Daniel Vatavu. The dissimilarity-consensus approach to agreement analysis in gesture elicitation studies. In *Proceedings of the 2019 CHI Conference on Human Factors*

in Computing Systems, CHI '19, page 1–13, New York, NY, USA, 2019. Association for Computing Machinery.

[182] Lisa Anthony, Radu-Daniel Vatavu, and Jacob O Wobbrock. Understanding the consistency of users' pen and finger stroke gesture articulation. In *Proceedings of Graphics Interface 2013*, pages 87–94. Citeseer, 2013.

[183] Yu-Chun Chen, Chia-Ying Liao, Shuo-wen Hsu, Da-Yuan Huang, and Bing-Yu Chen. Exploring user defined gestures for ear-based interactions. *Proc. ACM Hum.-Comput. Interact.*, 4(ISS), nov 2020.

[184] Tilman Dingler, Rufat Rzayev, Alireza Sahami Shirazi, and Niels Henze. *Designing Consistent Gestures Across Device Types: Eliciting RSVP Controls for Phone, Watch, and Glasses*, page 1–12. Association for Computing Machinery, New York, NY, USA, 2018.

[185] Guiying Du, Auriol Degbelo, and Christian Kray. User-generated gestures for voting and commenting on immersive displays in urban planning. *Multimodal Technologies and Interaction*, 3(2), 2019.

[186] Daniela Grijincu, Miguel A. Nacenta, and Per Ola Kristensson. User-defined interface gestures: Dataset and analysis. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, ITS '14, page 25–34, New York, NY, USA, 2014. Association for Computing Machinery.

[187] Maria Karam and M c Schraefel. *A Taxonomy of Gestures in Human Computer Interactions*. s.n., 2005.

[188] David McNeill. *Hand and mind: What gestures reveal about thought*. University of Chicago press, 1992.

[189] Jacob H Steffen, James E Gaskin, Thomas O Meservy, Jeffrey L Jenkins, and Iopa Wolman. Framework of affordances for virtual reality and augmented reality. *Journal of Management Information Systems*, 36(3):683–729, 2019.

[190] Adam S. Williams, Xiaoyan Zhou, Anil Ufuk Batmaz, Michel Pahud, and Francisco Ortega. A pilot study comparing user interactions between augmented and virtual reality. In George Bebis, Golnaz Ghiasi, Yi Fang, Andrei Sharf, Yue Dong, Chris Weaver, Zhicheng Leo, Joseph J. LaViola Jr., and Luv Kohli, editors, *Advances in Visual Computing*, pages 3–14, Cham, 2023. Springer Nature Switzerland.

[191] Kim Marriott, Falk Schreiber, Tim Dwyer, Karsten Klein, Nathalie Henry Riche, Takayuki Itoh, Wolfgang Stuerzlinger, and Bruce H Thomas. *Immersive analytics*, volume 11190. Springer, 2018.

[192] Andrea Batch, Andrew Cunningham, Maxime Cordeil, Niklas Elmqvist, Tim Dwyer, Bruce H. Thomas, and Kim Marriott. There is no spoon: Evaluating performance, space use, and presence with expert domain users in immersive analytics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):536–546, 2020.

[193] Maxime Cordeil, Andrew Cunningham, Tim Dwyer, Bruce H. Thomas, and Kim Marriott. Imaxes: Immersive axes as embodied affordances for interactive multivariate data visualisation. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, page 71–83, New York, NY, USA, 2017. Association for Computing Machinery.

[194] Kadek Ananta Satriadi, Barrett Ens, Maxime Cordeil, Tobias Czauderna, and Bernhard Jenny. Maps around me: 3d multiview layouts in immersive spaces. *Proc. ACM Hum.-Comput. Interact.*, 4(ISS), nov 2020.

[195] Yidan Zhang, Barrett Ens, Kadek Ananta Satriadi, Arnaud Prouzeau, and Sarah Goodwin. Timetables: Embodied exploration of immersive spatio-temporal data. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 599–605, 2022.

[196] Ibrahim A. Tahmid, Lee Lisle, Kylie Davidson, Chris North, and Doug A. Bowman. Evaluating the benefits of explicit and semi-automated clusters for immersive sensemaking. In

2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pages 479–488, 2022.

[197] Jiazhou Liu, Arnaud Prouzeau, Barrett Ens, and Tim Dwyer. Design and evaluation of interactive small multiples data visualisation in immersive spaces. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 588–597, 2020.

[198] Lee Lisle, Kylie Davidson, Edward J.K. Gitre, Chris North, and Doug A. Bowman. Sensemaking strategies with immersive space to think. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 529–537, 2021.

[199] Christopher Andrews, Alex Endert, and Chris North. Space to think: large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, page 55–64, New York, NY, USA, 2010. Association for Computing Machinery.

[200] Ying Yang, Tim Dwyer, Michael Wybrow, Benjamin Lee, Maxime Cordeil, Mark Billinghurst, and Bruce H. Thomas. Towards immersive collaborative sensemaking. *Proc. ACM Hum.-Comput. Interact.*, 6(ISS), nov 2022.

[201] Max Krichenbauer, Goshiro Yamamoto, Takafumi Taketom, Christian Sandor, and Hirokazu Kato. Augmented reality versus virtual reality for 3d object manipulation. *IEEE Transactions on Visualization and Computer Graphics*, 24(2):1038–1048, 2018.

[202] Anil Ufuk Batmaz, Mayra Donaji Barrera Machuca, Junwei Sun, and Wolfgang Stuerzlinger. The effect of the vergence-accommodation conflict on virtual hand pointing in immersive displays. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2022.

[203] Jiamin Ping, Yue Liu, and Dongdong Weng. Comparison in depth perception between virtual reality and augmented reality systems. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1124–1125, 2019.

[204] A. U. Batmaz, A. K. Mutasim, and W. Stuerzlinger. Precision vs. power grip: A comparison of pen grip styles for selection in virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 23–28, 2020.

[205] Alexandra Voit, Sven Mayer, Valentin Schwind, and Niels Henze. Online, vr, ar, lab, and in-situ: Comparison of research methods to evaluate smart artifacts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–12, New York, NY, USA, 2019. Association for Computing Machinery.

[206] Kylie Davidson, Lee Lisle, Kirsten Whitley, Doug A. Bowman, and Chris North. Exploring the evolution of sensemaking strategies in immersive space to think. *IEEE Transactions on Visualization and Computer Graphics*, 29(12):5294–5307, 2023.

[207] Ruth B Ekstrom and Harry Horace Harman. *Manual for kit of factor-referenced cognitive tests, 1976*. Educational testing service, 1976.

[208] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, 1988.

[209] Bob G. Witmer and Michael J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240, 06 1998.

[210] J Brooke. Sus: A 'quick and dirty'usability scale. usability evaluation in industry. pw jordan, b thomas, ba weerdmeester and al mcclelland, 1996.

[211] Nov 2023. https://varjo.com/products/varjo-xr-3/.

[212] 2018. https://www.vive.com/us/accessory/controller2018/.

[213] 2020. https://www.vive.com/us/support/camtrack/category_howto/base-station2.html.

[214] Adam S. Williams, Xiaoyan Zhou, Anil Ufuk Batmaz, Michel Pahud, and Francisco Ortega. A pilot study comparing user interactions between augmented and virtual reality. In George Bebis, Golnaz Ghiasi, Yi Fang, Andrei Sharf, Yue Dong, Chris Weaver, Zhicheng Leo, Joseph J. LaViola Jr., and Luv Kohli, editors, *Advances in Visual Computing*, pages 3–14, Cham, 2023. Springer Nature Switzerland.

[215] Adam Greenfeld, Artur Lugmayr, and Wesley Lamont. Comparative reality: Measuring user experience and emotion in immersive virtual environments. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 204–209. IEEE, 2018.

[216] Hung-Jui Guo, Jonathan Z Bakdash, Laura R Marusich, and Balakrishnan Prabhakaran. Augmented reality and mixed reality measurement under different environments: A survey on head-mounted devices. *IEEE Transactions on Instrumentation and Measurement*, 71:1–15, 2022.

[217] Marc Satkowski and Raimund Dachselt. Investigating the impact of real-world environments on the perception of 2d visualizations in augmented reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[218] Steven Feiner and Ari Shamash. Hybrid user interfaces: breeding virtually bigger interfaces for physically smaller computers. In *Proceedings of the 4th Annual ACM Symposium on User Interface Software and Technology*, UIST '91, page 9–17, New York, NY, USA, 1991. Association for Computing Machinery.

[219] G.R. King, W. Piekarski, and B.H. Thomas. ARVino - outdoor augmented reality visualisation of viticulture GIS data. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*, pages 52–55, October 2005.

[220] Wesley Willett, Yvonne Jansen, and Pierre Dragicevic. Embedded Data Representations. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):461–470, January 2017.

[221] Benjamin Lee, Michael Sedlmair, and Dieter Schmalstieg. Design Patterns for Situated Visualization in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–12, 2023.

[222] Sungbok Shin, Andrea Batch, Peter W. S. Butcher, Panagiotis D. Ritsos, and Niklas Elmqvist. The Reality of the Situation: A Survey of Situated Analytics. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–19, 2023.

[223] Sean White and Steven Feiner. SiteLens: Situated visualization techniques for urban site visits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1117–1120, Boston MA USA, April 2009. ACM.

[224] Aimee Sousa Calepso, Philipp Fleck, Dieter Schmalstieg, and Michael Sedlmair. Exploring Augmented Reality for Situated Analytics with Many Movable Physical Referents. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*, VRST '23, pages 1–12, New York, NY, USA, October 2023. Association for Computing Machinery.

[225] Johannes Zagermann, Sebastian Hubenschmid, Priscilla Balestrucci, Tiare Feuchtner, Sven Mayer, Marc O. Ernst, Albrecht Schmidt, and Harald Reiterer. Complementary interfaces for visual computing. *it - Information Technology*, 64(4-5):145–154, August 2022.

[226] Xiyao Wang, Lonni Besançon, David Rousseau, Mickael Sereno, Mehdi Ammi, and Tobias Isenberg. Towards an Understanding of Augmented Reality Extensions for Existing 3D Data Analysis Tools. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Honolulu HI USA, April 2020. ACM.

[227] Apple. Use your Mac with Apple Vision Pro. https://support.apple.com/en-us/HT213971, February 2024.

[228] Weizhou Luo, Anke Lehmann, Hjalmar Widengren, and Raimund Dachselt. Where Should We Put It? Layout and Placement Strategies of Documents in Augmented Reality for Collaborative Sensemaking. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, pages 1–16, New York, NY, USA, April 2022. Association for Computing Machinery.

[229] Lee Lisle, Kylie Davidson, Edward J.K. Gitre, Chris North, and Doug A. Bowman. Sensemaking Strategies with Immersive Space to Think. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 529–537, March 2021.

[230] Matt Whitlock, Keke Wu, and Danielle Albers Szafir. Designing for Mobile and Immersive Visual Analytics in the Field. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):503–513, January 2020.

[231] Niklas Elmqvist. Data Analytics Anywhere and Everywhere. *Communications of the ACM*, 66(12):52–63, November 2023.

[232] Marco Jahn, Marc Jentsch, Christian R. Prause, Ferry Pramudianto, Amro Al-Akkad, and Rene Reiners. The Energy Aware Smart Home. In *2010 5th International Conference on Future Information Technology*, pages 1–8, May 2010.

[233] Ambre Assor, Arnaud Prouzeau, Martin Hachet, and Pierre Dragicevic. Handling Non-Visible Referents in Situated Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):1336–1346, January 2024.

[234] Wolfgang Narzt, Gustav Pomberger, Alois Ferscha, Dieter Kolb, Reiner Müller, Jan Wieghardt, Horst Hörtner, and Christopher Lindinger. Augmented reality navigation systems. *Universal Access in the Information Society*, 4(3):177–187, March 2006.

[235] Barrett Ens, Benjamin Bach, Maxime Cordeil, Ulrich Engelke, Marcos Serrano, Wesley Willett, Arnaud Prouzeau, Christoph Anthes, Wolfgang Büschel, Cody Dunne, Tim

Dwyer, Jens Grubert, Jason H. Haga, Nurit Kirshenbaum, Dylan Kobayashi, Tica Lin, Monsurat Olaosebikan, Fabian Pointecker, David Saffo, Nazmus Saquib, Dieter Schmalstieg, Danielle Albers Szafir, Matt Whitlock, and Yalong Yang. Grand Challenges in Immersive Analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–17, Yokohama Japan, May 2021. ACM.

[236] Jorge A. Wagner Filho, Wolfgang Stuerzlinger, and Luciana Nedel. Evaluating an Immersive Space-Time Cube Geovisualization for Intuitive Trajectory Data Exploration. *IEEE Transactions on Visualization and Computer Graphics*, 26(01):514–524, January 2020.

[237] Benjamin Lee, Xiaoyun Hu, Maxime Cordeil, Arnaud Prouzeau, Bernhard Jenny, and Tim Dwyer. Shared Surfaces and Spaces: Collaborative Data Visualisation in a Co-located Immersive Environment. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1171–1181, February 2021.

[238] Wolfgang Büschel, Anke Lehmann, and Raimund Dachselt. MIRIA: A Mixed Reality Toolkit for the In-Situ Visualization and Analysis of Spatio-Temporal Interaction Data. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15, Yokohama Japan, May 2021. ACM.

[239] Matt Whitlock, Stephen Smart, and Danielle Albers Szafir. Graphical Perception for Immersive Analytics. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 616–625, March 2020.

[240] Tamara Munzner. *Visualization Analysis and Design*. A.K. Peters Visualization Series. CRC Press, Taylor & Francis Group, CRC Press is an imprint of the Taylor & Francis Group, an informa business, Boca Raton, 2014.

[241] Jiayi Hong, Rostyslav Hnatyshyn, Ebrar A. D. Santos, Ross Maciejewski, and Tobias Isenberg. A Survey of Designs for Combined 2D+3D Visual Representations. *IEEE Transactions on Visualization and Computer Graphics*, 2024.

[242] Christophe Hurter, Nathalie Henry Riche, Steven M. Drucker, Maxime Cordeil, Richard Alligier, and Romain Vuillemot. FiberClay: Sculpting Three Dimensional Trajectories to Reveal Structural Insights. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):704–714, January 2019.

[243] Benjamin Lee, Maxime Cordeil, Arnaud Prouzeau, Bernhard Jenny, and Tim Dwyer. A Design Space For Data Visualisation Transformations Between 2D And 3D In Mixed-Reality Environments. In *CHI Conference on Human Factors in Computing Systems*, pages 1–14, New Orleans LA USA, April 2022. ACM.

[244] Kadek Ananta Satriadi, Barrett Ens, Maxime Cordeil, Tobias Czauderna, and Bernhard Jenny. Maps Around Me: 3D Multiview Layouts in Immersive Spaces. *Proceedings of the ACM on Human-Computer Interaction*, 4(ISS):201:1–201:20, November 2020.

[245] Sungwon In, Tica Lin, Chris North, Hanspeter Pfister, and Yalong Yang. This is the Table I Want! Interactive Data Transformation on Desktop and in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–16, 2023.

[246] Jiazhou Liu, Barrett Ens, Arnaud Prouzeau, Jim Smiley, Isobel Kara Nixon, Sarah Goodwin, and Tim Dwyer. DataDancing: An Exploration of the Design Space For Visualisation View Management for 3D Surfaces and Spaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–17, Hamburg Germany, April 2023. ACM.

[247] Jinbin Huang, Shuang Liang, Qi Xiong, Yu Gao, Chao Mei, Yi Xu, and Chris Bryan. SPARVIS: Combining Smartphone and Augmented Reality for Visual Data Analytics. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 111–117, October 2022.

[248] Sean Michael White. *Interaction and Presentation Techniques for Situated Visualization*. PhD thesis, Columbia University, USA, 2009.

[249] Zeinab Ghaemi, Kadek Ananta Satriadi, Ulrich Engelke, Barrett Ens, and Bernhard Jenny. Visualization Placement for Outdoor Augmented Data Tours. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction*, SUI '23, pages 1–14, New York, NY, USA, October 2023. Association for Computing Machinery.

[250] Qian Zhu, Zhuo Wang, Wei Zeng, Tong Wai, Weiyue Lin, and Xiaojuan Ma. Make Interaction Situated: Designing User Acceptable Interaction for Situated Visualization in Public Environments, February 2024.

[251] Nina Doerr, Benjamin Lee, Katarina Baricova, Dieter Schmalstieg, and Michael Sedlmair. Visual Highlighting for Situated Brushing and Linking, March 2024.

[252] Anika Sayara, Benjamin Lee, Carlos Quijano-Chavez, and Michael Sedlmair. Designing Situated Dashboards: Challenges and Opportunities. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 97–102, October 2023.

[253] Bongshin Lee, Matthew Brehmer, Petra Isenberg, Eun Kyoung Choe, Ricardo Langner, and Raimund Dachselt. Data Visualization on Mobile Devices. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages 1–8, New York, NY, USA, April 2018. Association for Computing Machinery.

[254] Niklas Elmqvist and Pourang Irani. Ubiquitous Analytics: Interacting with Big Data Anywhere, Anytime. *Computer*, 46(4):86–89, April 2013.

[255] Irene Ros. MobileVis. https://mobilev.is/.

[256] Bongshin Lee, Raimund Dachselt, Petra Isenberg, and Eun Kyoung Choe, editors. *Mobile Data Visualization*. AK Peters Visualization Series. CRC Press, first edition edition, 2022.

[257] Wolfgang Büschel, Annett Mitschick, Thomas Meyer, and Raimund Dachselt. Investigating Smartphone-based Pan and Zoom in 3D Data Spaces in Augmented Reality. In *Proceedings*

of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '19, pages 1–13, New York, NY, USA, October 2019. Association for Computing Machinery.

[258] Mickael Sereno, Stéphane Gosset, Lonni Besançon, and Tobias Isenberg. Hybrid Touch/Tangible Spatial Selection in Augmented Reality. *Computer Graphics Forum*, 41(3):403–415, 2022.

[259] Desney S. Tan and Mary Czerwinski. Effects of visual separation and physical discontinuities when distributing information across multiple displays. In Matthias Rauterberg, Marino Menozzi, and Janet Wesson, editors, *Human-Computer Interaction INTERACT '03: IFIP TC13 International Conference on Human-Computer Interaction, 1st-5th September 2003, Zurich, Switzerland*. IOS Press, 2003.

[260] Umar Rashid, Miguel A. Nacenta, and Aaron Quigley. The cost of display switching: A comparison of mobile, large display and hybrid UI configurations. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pages 99–106, New York, NY, USA, May 2012. Association for Computing Machinery.

[261] Umar Rashid, Miguel A. Nacenta, and Aaron Quigley. Factors influencing visual attention switch in multi-display user interfaces: A survey. In *Proceedings of the 2012 International Symposium on Pervasive Displays*, pages 1–6, Porto Portugal, June 2012. ACM.

[262] Jens Grubert, Matthias Kranz, and Aaron Quigley. Challenges in mobile multi-device ecosystems. *mUX: The Journal of Mobile User Experience*, 5(1):5, August 2016.

[263] Andy Brown, Michael Evans, Caroline Jay, Maxine Glancy, Rhianne Jones, and Simon Harper. HCI over multiple screens. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 665–674, New York, NY, USA, April 2014. Association for Computing Machinery.

[264] Jia Wang and Robert Lindeman. Coordinated 3D interaction in tablet- and HMD-based hybrid virtual environments. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction*, pages 70–79, Honolulu Hawaii USA, October 2014. ACM.

[265] Jonathan C. Roberts. State of the Art: Coordinated & Multiple Views in Exploratory Visualization. In *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)*, pages 61–71, July 2007.

[266] Markus Tatzgern, Valeria Orso, Denis Kalkofen, Giulio Jacucci, Luciano Gamberini, and Dieter Schmalstieg. Adaptive information density for augmented reality displays. In *2016 IEEE Virtual Reality (VR)*, pages 83–92, March 2016.

[267] Benjamin Bach, Ronell Sicat, Hanspeter Pfister, and Aaron Quigley. Drawing into the AR-CANVAS: Designing embedded visualizations for augmented reality. In *Workshop on Immersive Analytics, IEEE Vis*, 2017.

[268] Jan-Henrik Schröder, Daniel Schacht, Niklas Peper, Anita Marie Hamurculu, and Hans-Christian Jetter. Collaborating Across Realities: Analytical Lenses for Understanding Dyadic Collaboration in Transitional Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pages 1–16, New York, NY, USA, April 2023. Association for Computing Machinery.

[269] Cha Lee, Scott Bonebrake, Tobias Hollerer, and Doug A. Bowman. A replication study testing the validity of AR simulation in VR for controlled experiments. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 203–204, Orlando, FL, USA, October 2009. IEEE.

[270] Cha Lee, Gustavo A. Rincon, Greg Meyer, Tobias Hollerer, and Doug A. Bowman. The Effects of Visual Realism on Search Tasks in Mixed Reality Simulation. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):547–556, April 2013.

[271] Jerônimo G Grandi, Zekun Cao, Mark Ogren, and Regis Kopper. Design and Simulation of Next-Generation Augmented Reality User Interfaces in Virtual Reality. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 23–29, March 2021.

[272] Joseph J. LaViola, Ernst Kruijff, Ryan P. McMahan, Doug A. Bowman, and Ivan Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley Usability and HCI Series. Addison-Wesley, Boston, second edition edition, 2017.

[273] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, July 1993.

[274] John Brooke. SUS: A 'Quick and Dirty' Usability Scale. In *Usability Evaluation In Industry*. CRC Press, 1996.

[275] Rebecca A. Grier. How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59(1):1727–1731, September 2015.

[276] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3):114–123, May 2009.

[277] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, January 2006.

[278] Kay M. Stanney, Robert S. Kennedy, and Julie M. Drexler. Cybersickness is Not Simulator Sickness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 41(2):1138–1142, October 1997.

[279] Pauline Bimberg, Tim Weissker, and Alexander Kulik. On the Usage of the Simulator Sickness Questionnaire for Virtual Reality Research. In *2020 IEEE Conference on Virtual*

Reality and 3D User Interfaces Abstracts and Workshops (VRW), pages 464–467, March 2020.

[280] Yun Wang, Adrien Segal, Roberta Klatzky, Daniel F. Keefe, Petra Isenberg, Jorn Hurtienne, Eva Hornecker, Tim Dwyer, and Stephen Barrass. An Emotional Response to the Value of Visualization. *IEEE Computer Graphics and Applications*, 39(5):8–17, September 2019.

[281] Jeffrey Heer and Ben Shneiderman. Interactive dynamics for visual analysis. *Communications of the ACM*, 55(4):45–54, April 2012.

[282] Yidan Zhang, Barrett Ens, Kadek Ananta Satriadi, Ying Yang, and Sarah Goodwin. Embodied Provenance for Immersive Sensemaking. *Proceedings of the ACM on Human-Computer Interaction*, 7(ISS):435:198–435:216, November 2023.

[283] Kadek Ananta Satriadi, Andrew Cunningham, Ross T. Smith, Tim Dwyer, Adam Drogemuller, and Bruce H. Thomas. ProxSituated Visualization: An Extended Model of Situated Visualization using Proxies for Physical Referents. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pages 1–20, New York, NY, USA, April 2023. Association for Computing Machinery.

[284] João Marcelo Evangelista Belo, Mathias N. Lystbæk, Anna Maria Feit, Ken Pfeuffer, Peter Kán, Antti Oulasvirta, and Kaj Grønbæk. AUIT – the Adaptive User Interfaces Toolkit for Designing XR Applications. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22, pages 1–16, New York, NY, USA, October 2022. Association for Computing Machinery.

[285] Ari Kouts, Lonni Besançon, Michael Sedlmair, and Benjamin Lee. LSDvis: Hallucinatory Data Visualisations in Real World Environments. In *Alt.VIS 2023, an IEEE VIS Workshop*. arXiv, December 2023.

[286] David Kirsh and Paul Maglio. On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18(4):513–549, October 1994.

Appendix A

Appendix

A.1 Questions Asked During Experiments

The questions below were the first question set used in the AR-VR comparison study (Chapter 4).

- Which brand makes the cereal with the highest Fat?
- Which brand makes the cereal with the lowest Dietary Fiber and Sugar?
- What is the Sugar content of the cereal that has the lowest Fat and Protein as well as the highest Carbs?
- Which value is the closest to the mean Dietary Fiber content?
- What is the relationship between Carbs and Sugar?
- What are the Fat content values for the three largest clusters (groups)?

The questions below were the second question set used in the AR-VR comparison study (Chapter 4).

- What are the Protein content values for the three largest clusters (groups)?
- What is the relationship between Sugar and Protein?
- Which value is the closest to the mean Fat content?
- What is the Protein content of the cereal that has the lowest Fat and Carbs as well as the lowest Dietary Fiber?
- Which brand makes the cereal with the lowest Sugar and high Protein?

- Which brand makes the cereal with the highest Dietary Fiber?

The questions below were used in the semi-structured interview of the hybrid user interface with situated analytics study (Chapter 4).

- Do you notice any difference between AR and VR?
- Do you have any insights or strategies you took to get more comfortable with the tasks?
- Do you have any other thoughts you want to share with the research team?
- Do you have any additional comments about your experience today that you'd like to add?

The questions below were used in the semi-structured interview of the hybrid user interface with situated analytics study (Chapter 5).

- What is your impression of the system? Can you please elaborate?
- What strategies did you use to solve the tasks? Did it work? If it didn't, why not?
- What strategies did you use to look at and analyze both the graphs and visualizations on the tablet and in the apartment? Did it work? If it didn't, why not?
- What strategies did you use to navigate through the physical space? How natural did walking around feel to you?
- Do you have any suggestions for improving the system to help you perform similar tasks in the future? In terms of its features, user interfaces, visualizations, and interactions. Tasks would include looking at data on both the tablet and in the environment to help make some decision.
- Do you see yourself using this system in the real world with augmented reality? What about if the augmented reality system included your proposed suggestions?

A.2 Surveys

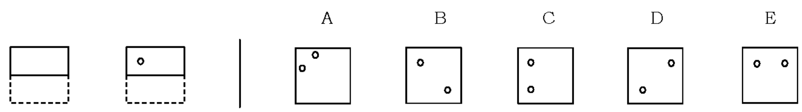
A.2.1 Paper Folding Test

The paper folding test originally was originally released in the "Kit of Factor-Referenced Cognitive Tests" in 1976 [2]. The version used in this study is shown in figures A.1A.2A.3.

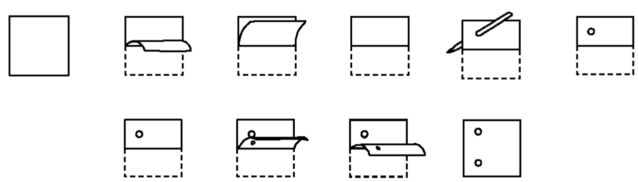
Paper Folding Test—Vz-2-BRACE

In this test you are to imagine the folding and unfolding of pieces of paper. In each problem in the test there are some figures drawn at the left of a vertical line and there are others drawn at the right of the line. The figures at the left represent a square piece of paper being folded, and the last of these figures has one or two small circles drawn on it to show where the paper has been punched. Each hole is punched through all the thicknesses of paper at that point. One of the five figures on the right of the vertical line shows where the holes will be when the paper is completely unfolded. You are to decide which one of these figures is correct and draw an X through that figure.

Now try the sample problem below. (In this problem only one hole was punched in the folded paper).



The correct answer to the sample problem above is C and so it should have been marked with an X. The figures below show how the paper was folded and why C is the correct answer.



In these problems all of the folds that are made are shown in the figures at the left of the line, and the paper is not turned or moved in any way except to make the folds shown in the figures. Remember, the answer is the figure that shows the positions of the holes when the paper is completely unfolded.

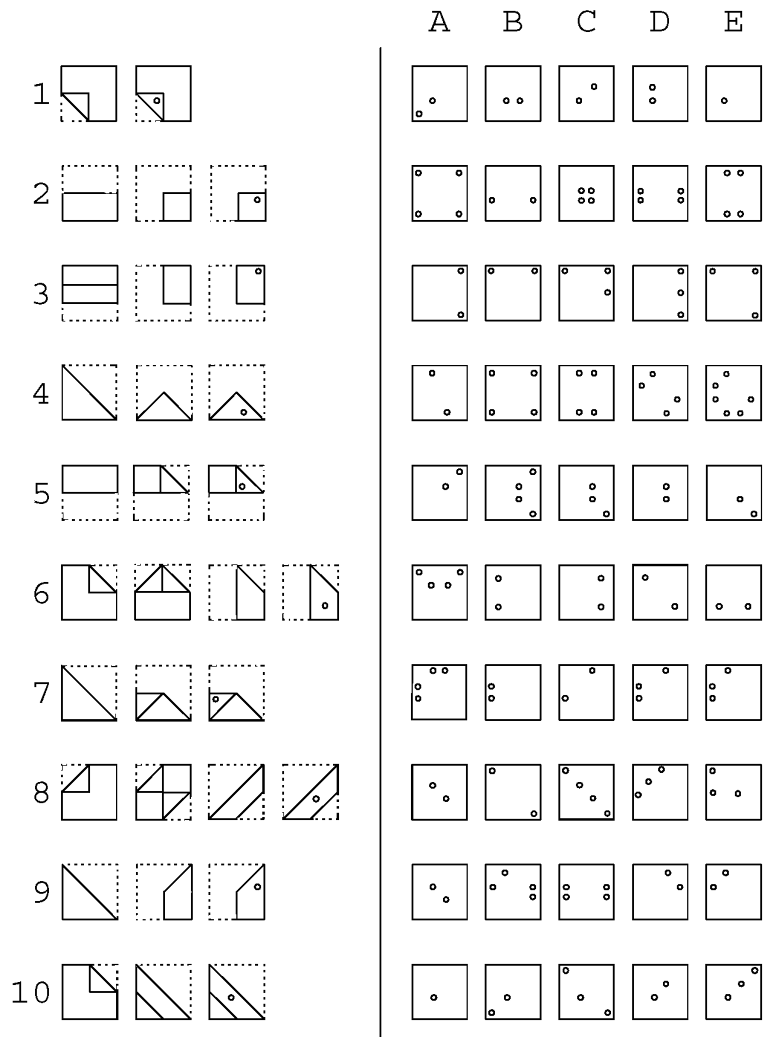
Some of the problems on this sheet are more difficult than others. If you are unable to do one of the problems, simply skip over it and go on to the next one.

You will have three minutes for each of the two parts of this test. Each part has one page. When you have finished Part One, STOP. Please do not go on to Part Two until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

Figure A.1: Paper Folding Test VZ-2, Page one [2]

PART ONE (3 MINUTES)

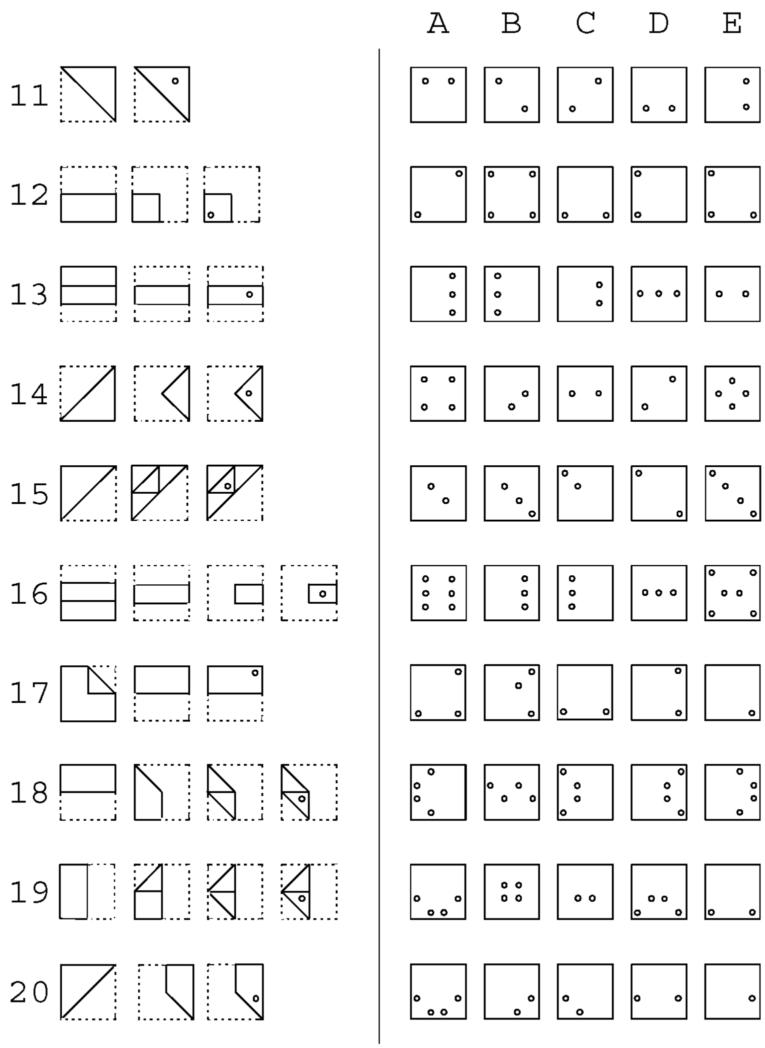


STOP

DO NOT PROCEED TO THE NEXT PAGE UNTIL ASKED TO DO SO

Figure A.2: Paper Folding Test VZ-2, Page two [2]

PART TWO (3 MINUTES)



STOP AND WAIT FOR FURTHER INSTRUCTIONS
DO NOT GO BACK TO PART ONE

Figure A.3: Paper Folding Test VZ-2, Page three [2]

A.3 Supplementary Material

A.3.1 Codebook

This document contains additional visualizations from the video coding process as described in the main paper. The codebook consists of 23 behaviors, with explanations shown in Table A.1. The total duration of codes are shown in Figure A.4 as well as their individual counts in Figure A.5. Complete timelines for each individual participant are shown at the end of the document. As mentioned in the main manuscript, P14 could not be coded due to a technical error in recording their POV.

Table A.1: The codebook of 23 behaviors that were applied to the videos of our participants' POVs during our exploratory study.

Category	Code	Description
Task Order	Finish one room at a time	Adjust all of the lights in the same room before continuing to another room
Task Order	Go to closest light in another room	After adjusting a light, continue to the next closest light even though it is in another room
Movement Pattern	Wander to each light	Move around the apartment from light to light
Movement Pattern	Wander in a circle	Move around the apartment in a circular pattern
Movement Pattern	Wander randomly	Move around the apartment with no pattern
Movement Pattern	Stay close to a light	Stand still directly next to a light
Movement Pattern	Stay in middle of room	Stand still around the middle of the room
Movement Pattern	Stay in corner of room	Stand still in a corner of the room
Movement Pattern	Go to room first then check data	Go into a room first, then look at the SitVis
Movement Pattern	Check data first then go to room	Look at the SitVis, then go into a room
Movement Pattern	Look at other furniture	Look around at furniture not relevant to the task
Visualization Usage	Look up at EmbVis while changing light	Looking up at an EmbVis whilst changing the light level
Visualization Usage	Look up at EmbVis without changing light	Looking up at an EmbVis but not changing the light level
Visualization Usage	Look at SitVis while changing light	Looking at the SitVis whilst changing the light level
Visualization Usage	Look down at SitVis	Looking down at the SitVis
Visualization Usage	Hold SitVis up to light	Holding the SitVis up to a light and its EmbVis such that the two are visually adjacent
Visualization Usage	Hold SitVis halfway up to light	Holding the SitVis slightly upwards to a light and its EmbVis such that the two are still visually separated, requiring head turning
Visualization Usage	Look around to check lighting effects	Looking around the light and its EmbVis to see the effects of the new light level
Interaction with Lights	Walk close to lights	Standing directly next to a light while adjusting it
Interaction with Lights	Stay far from lights	Standing far away from a light while adjusting it
Light Preference	Use pie chart for final checks	Making light adjustments based on the pie chart
Light Preference	Use personal preference for final checks	Making light adjustments without looking at data on the SitVis
Light Preference	Go back to check previous lights	Making light adjustments by going back to previous lights that were already adjusted

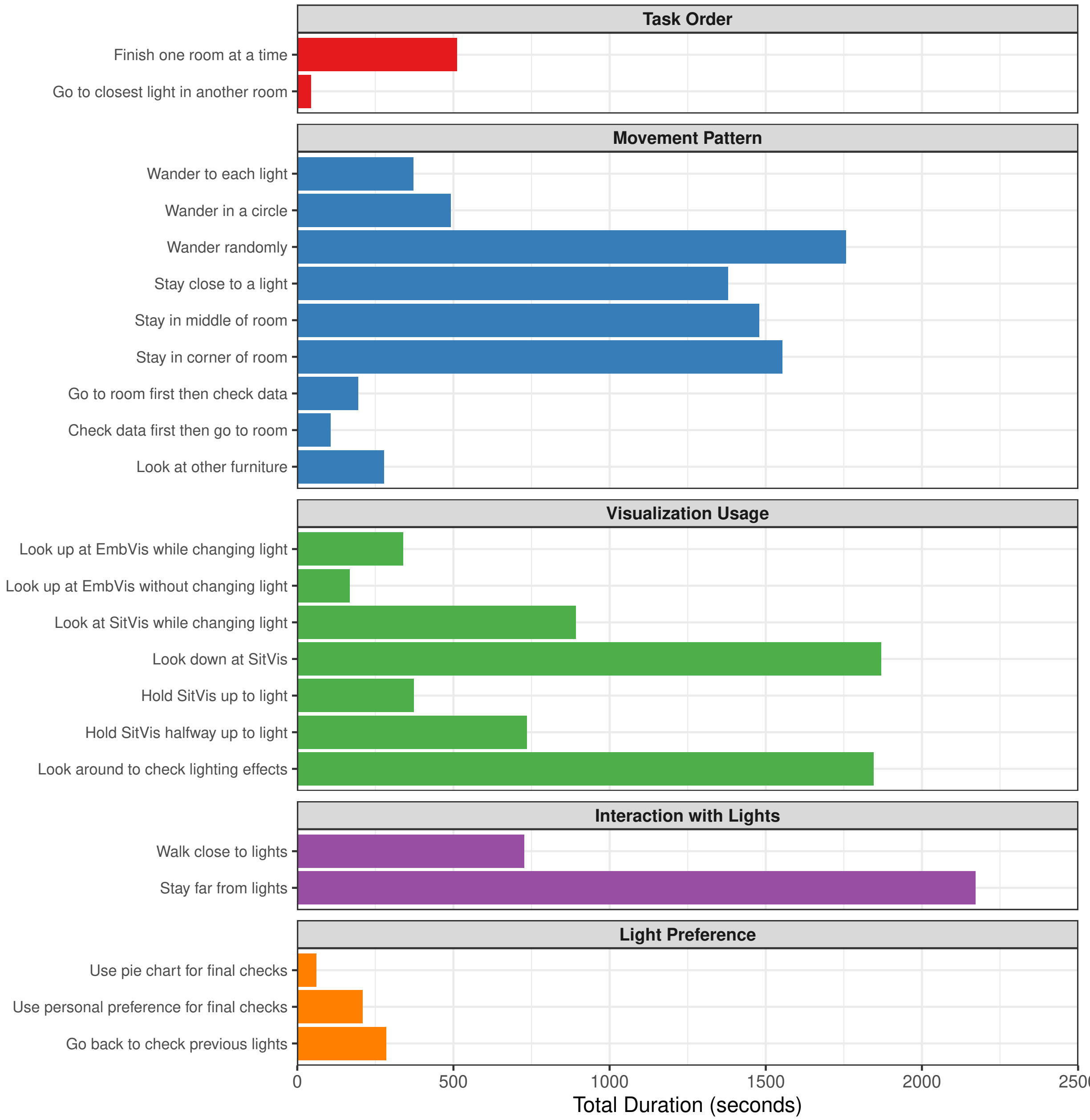


Figure A.4: Total durations of coded behaviors aggregated across all participants.

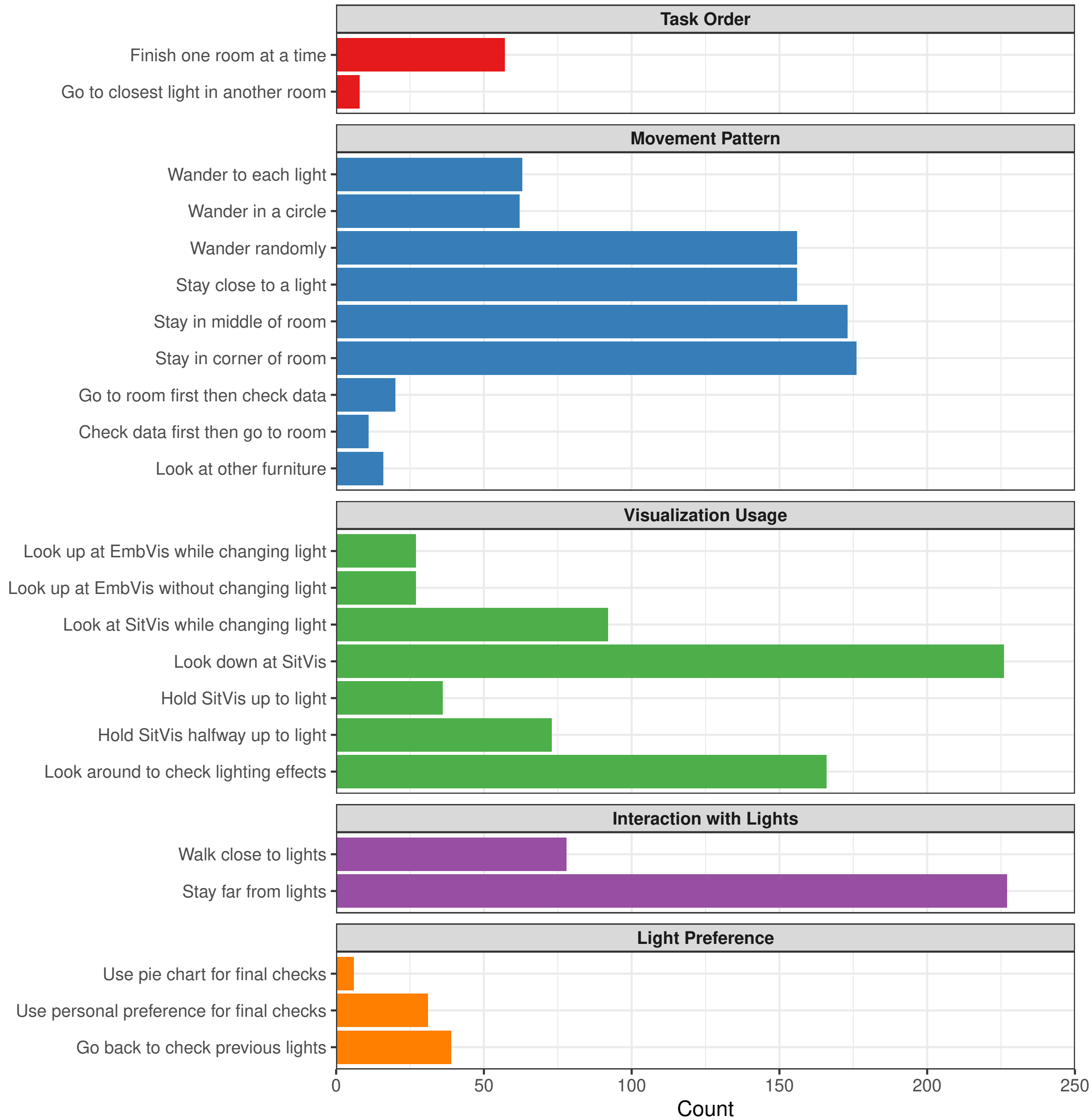


Figure A.5: Total counts of the frequency of coded behaviors aggregated across all participants.