DISSERTATION

COOKING UP A BETTER AR EXPERIENCE: NOTIFICATION DESIGN AND THE LIABILITIES OF IMPERFECT CUES IN AUGMENTED REALITY

Submitted by

Aditya R. Raikwar

Department of Computer Science

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2024

Doctoral Committee:

Advisor: Francisco R. Ortega

Indrakshi Ray
Marcia Moraes
Hortensia Soto

ABSTRACT

COOKING UP A BETTER AR EXPERIENCE: NOTIFICATION DESIGN AND THE LIABILITIES OF IMPERFECT CUES IN AUGMENTED REALITY

This dissertation investigates optimizing user experience in Augmented Reality (AR). A virtual cooking environment (ARtisan Bistro) serves as a testbed to explore factors influencing user interaction with AR interfaces. The research starts with notification design, examining strategically placed visual and audio notifications in ARtisan Bistro (Chapter 4). Building on this, Chapter 5 explores optimizing these designs for user awareness and delivering critical information, especially when audio is impractical. This involved exploring visual-only notifications, revealing consistent user performance and attention capture comparable to combined visual-audio notifications (no significant difference found).

The research demonstrates that well-designed notifications can significantly improve user experience, but it also raises a crucial question: can users always trust the information presented in AR environments? The possibility of imperfect information delivery underscores the importance of reliable information delivery. Chapter 6 explores the impact of imperfect cues generated by machine learning (ML) on user performance in AR visual search tasks. This research highlights the potential for automation bias when users rely heavily on unreliable cues.

By investigating both notification design and the limitations of ML systems for reliable information delivery, this dissertation emphasizes the importance of creating a well-rounded user experience in AR environments. The findings underscore the need for further research on optimizing visual notifications, mitigating automation bias, and ensuring reliable information delivery in AR applications.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to everyone who has supported me throughout my doctoral journey. This dissertation would not have been possible without many individuals' guidance, support, and encouragement.

First and foremost, I wish to thank my advisor, Dr. Francisco Raul Ortega, for his insightful guidance, patient mentorship, and unwavering belief in my abilities. You have not only shaped my academic path but have also profoundly influenced my personal growth. Thank you for your wisdom and your time and for pushing me to strive for excellence. I am truly grateful for your mentorship.

I also extend my heartfelt thanks to my dissertation committee members, Dr. Indrakshi Ray, Dr. Marcia Moraes, and Dr. Hortensia Soto, for their diverse perspectives, encouragement, and support throughout this process. Your expertise has greatly enriched my research.

To my colleagues and friends at NUILAB, thank you for fostering an environment of camaraderie and a collaborative spirit, especially Dr. Adam Coler, Faezeh Borhani, Xiaoyan Zhou, and Richard Rodriguez. We've been on this journey together since the beginning, and your camaraderie and support have been invaluable. The intellectual discussions and shared laughter have made this journey enjoyable and fulfilling. Finally, a special thanks to Lucas Plabst for your invaluable assistance with the experiment development. All your friendship and support have meant the world to me.

I would like to acknowledge the financial support from National Science Foundation Awards 2327569 and 2238313, and Office of Naval Research grants N00014-24-1-2214 and N00014-21-1-2580, which made this research possible. Additionally, I am thankful to the participants who volunteered their time to take part in my study.

I am deeply indebted to my family for their unconditional love, patience, and encouragement. To my parents, Ravindra Raikwar and Dr. Vaishali Raikwar, thank you for instilling in me the values of hard work and perseverance – values that have been crucial during challenging moments. To my sister, Kanishka Raikwar, your unwavering support and belief in me have been a constant source of strength. Her drive and overachievements have constantly pushed me to work harder and strive for excellence. Knowing she looks up to me has been a powerful motivator throughout this journey.

Lastly, I would like to thank all my friends, Saurabh Chaudhary, Dr. Dhruva Patil, Abhishek Shetye, and others, who have been there for me, providing support and encouragement when I needed it the most. Your belief in me has been a source of strength and motivation.

This dissertation is dedicated to all of you. Thank you for being an integral part of my academic journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

The field of Human-Computer Interaction (HCI) and adjacent fields (human factors and cognitive psychology) have long researched the concept of interruptions. Human cognition allows for multitasking capabilities [2], leading individuals to frequently shift between tasks either voluntarily or due to external influence. Interruptions are a core part of human-computer interaction as users often engage in multitasking activities, such as checking email while working on a project. Gould et al. highlighted the continuing interest in understanding interruptions in multitasking environments and argued that interruptions can have both positive and negative effects on user attention and task performance [3]. As computers have become ubiquitous in daily life, the research on interruptions caused by notifications on traditional computer interfaces (e.g., desktop computers, laptops, smartphones) has been at the forefront [4–9]. Capturing user attention effectively and ensuring clear information delivery are crucial aspects of successful interaction within Augmented Reality (AR) environments. This study explored the impact of notification design on these key factors, evaluating user experience and usability through a series of experiments.

Previous research (e.g., [9–12]) suggests that task-independent notifications will decrease the performance. While notifications can be disruptive, hindering attention and task performance, well-designed ones can provide timely updates on ongoing processes and ultimately aid in task management and reduce user anxiety [13]. This duality of notifications becomes even more critical in AR environments. With the rise of Augmented and Virtual Reality technology, another dimension is added to the notifications, requiring careful design considerations to balance effectiveness and user experience.

Unlike traditional desktop or mobile interfaces, AR displays information directly within the user's field of view (FoV). This limited FoV, especially compared to the human eye, presents a challenge for our study in displaying notifications without obstructing the user's vision, particularly during movement. While AR devices have come a long way, one of the main shortcomings of these devices is the FoV. It is smaller (Hololens 2 has 52°FoV) when compared to average Virtual Reality VR (90°-100°) and even smaller when compared to the human eye ($\sim$135°). Traditional notification frameworks used on desktop and mobile devices may not be directly applicable to AR HMDs due to these unique limitations and user interaction paradigms. There are a few things to keep in mind while showing information or interrupting users: Where to display information?; How much occlusion is tolerable in different contexts?; When to display information?; What should be the interaction method for the user? Given this, the framework utilized by other smart devices cannot be applied to information display in Mixed Reality (MR).

The evolution of display technology is leading to a future where screens are becoming increasingly integrated into our daily lives. From desktops to laptops, handheld devices, and now AR HMDs, screens are moving closer to our faces, offering high-resolution and immersive experiences. AR allows the virtual and real worlds to be co-located [14–16]. It's important to comprehend how these omnipresent AR devices will be used in daily life. Understanding the user's context is important in order to present accurate and pertinent information and control how pervasive gadgets behave [17]. On smartphones, information can be retrieved and interacted with in a variety of ways using touch interfaces and speech/audio responses. But there are times when these kinds of interactions are inadequate [18]. When a user's hands are occupied, or they are unable to focus on anything other than their primary task, using a smartphone to gather information becomes impossible.

AR is less immersive compared to VR and necessitates more thought when deciding where and how to display virtual content. New information suddenly appearing in front of someone and obscuring their FoV can cause inconvenience and be highly distracting in an attention-sensitive situation [19,20]. Understanding the user's context is crucial for delivering relevant information and interacting with the pervasive devices [17]. Effective information delivery within AR environments presents significant challenges, particularly regarding notification design (Chapters 4 and 5) and visual search support (Chapter 6). Notifications must balance being informative and attention-grabbing without causing undue distraction or cognitive overload. Visual search tasks in complex environments require efficient methods to guide users to targets, and AR HMDs with visual cues present a promising solution. However, both notification design and visual search effectiveness are highly dependent on user experience and can be negatively impacted by factors like poorly designed notifications or unreliable visual cues.

This dissertation investigates these challenges and explores how to optimize user experience with information delivery and visual search tasks within AR environments. Chapters 4 and 5 focus on notification design, exploring the effectiveness of visual and audio modalities for notifications within the context of AR. These chapters examine how strategically placed visual and audio notifications can enhance user experience in AR by considering the limitations of FoV and the potential for user distraction.

While understanding how to design effective notifications and visual cues is crucial for a seamless AR experience, it's equally important to ensure the reliability of those cues themselves. Even with the best design, imperfect cues can significantly harm user experience and performance. Imagine searching for a critical medical tool in an AR-assisted surgery only to be directed to the wrong location due to an inaccurate visual cue. This wastes valuable time and can lead to frustration and potentially compromised safety. Therefore, exploring the impact of imperfect cues becomes paramount after investigating notification design. Chapter 6 delves into visual search tasks within AR. In complex visual environments, searching for specific targets can be effortful and time-consuming [21–23]. AR HMDs with visual cues can support visual search by guiding users to objects of interest. However, the effectiveness of these cues depends on their accuracy. This chapter explores how errors in visual cues generated by machine learning (ML) systems can impact user performance and potentially lead to automation bias, where users over-rely on cues even when they are incorrect [24].

Building on the established knowledge about interruption management in HCI (explored in more detail in Chapter 2), this dissertation investigates how notification design (Chapters 4 and 5) and visual search support (Chapter 6) can be optimized for AR environments. A controlled AR environment called ARtisan Bistro (described in Chapter 3) serves as the foundation for user studies for the studies described in Chapters 4 and 5. This environment allowed me to isolate variables and assess the impact of different design choices on user experience in a simulated but realistic setting. The findings from these studies provide valuable insights for designing AR interfaces that are both informative and user-friendly, ultimately fostering more efficient and effective interaction with information within AR applications.

This dissertation is a compilation of several works completed during my studies on notifications and visual cues. Chapter 4, currently under review at a top-tier peer-reviewed venue, has been slightly modified from its original form to add more details. Chapter 6 is presented as it was published in a top-tier peer-reviewed venue. These works are introduced with explanations of their relevance and necessity in the context of this research. Including these publications provides important insights into the key decisions made throughout this research journey.

1.1 Contributions

This work advances the understanding of HCI and AR through

1. **Improved Notification Design (Chapters 4 and 5):** By analyzing user performance and preferences, I identified optimal design strategies for the placement and modality of notifications in AR HMDs. Our findings provide clear guidelines for developers.

2. **Mitigating Disruption and Enhancing User Experience (Chapters 4 and 5):** The research demonstrates how AR notifications can be designed to minimize disruption and even assist users in completing tasks. This knowledge will inform the development of user-centered AR applications.

3. **Understanding User Behavior in AR (Chapters 4 and 5):** The study provides valuable insights into users' preferences with different notifications within an AR environment. This knowledge can be used to develop future AR notification design principles.

4. **Impact of Imperfect Cues in AR Visual Search (Chapter 6):** This research investigates a critical factor in AR visual search tasks: the impact of errors in visual cues generated by ML systems. The study explores how these errors can affect user performance and potentially lead to automation bias. These findings can inform the design of reliable AR guidance systems and user training strategies to mitigate overreliance on cues, ultimately fostering safer and more effective interaction with AR visual search tools.

5. **Quantifying Automation Bias in AR Visual Search with Machine Learning (Chapter 6):** The study demonstrates a high level of automation bias when users rely on visual cues generated by an ML system during a visual search task. Furthermore, the research reveals that the magnitude of automation bias was even greater using the ML system compared to a WoZ study, likely due to the higher perceived reliability of the imperfect visual cue generated by the ML system due to better performance.

This research has paved the way for a future where AR notifications and visual cues are seamlessly integrated into user workflows, enhancing productivity and user satisfaction within the rapidly evolving world of Augmented Reality.

Chapter 2

Related Work

This chapter delves into the existing body of research relevant to the experiments specifically dealing with "Notifications" detailed in Chapters 4 and 5. Well-designed interfaces are crucial for a user to interact effectively within an AR environment. These interfaces need to deliver information clearly, minimize disruption to the user's primary task, and cater to user preferences. By exploring related research, I gained valuable insights into existing design strategies and identified potential areas for improvement.

In the future, AR-Head Mounted Displays (AR-HMDs) are expected to be used often, just like smartphones, due to the increase in the use of AR technology in a variety of fields. One of the crucial features of a typical digital device in daily use is the notification mechanism [25]. However, AR-HMDs are a novel computing environment that is distinct from conventional smart gadgets. As a result, it's critical to comprehend how to use the new technique to register AR notifications on HMDs in practical situations [26].

2.1 Notifications

A typical illustration of a multimodal notification is the visual pop-up window that shows some basic information when an email is received on a smartphone. In the event that the device's owner is not nearby or is not focusing on it, there is an additional audible cue. Furthermore, a vibration sequence is included to notify the owner when the gadget is in their pocket or in noisy environments. But there are other sorts of notifications as well; this is just the most popular and well-known. Notifications vary according to the user's needs and the context. Research is continuing to determine the best time to interrupt users without involving humans. Most of the time it comes down to the user to select their preference of being interrupted. Kern et al. suggested 5 factors for user interruptibility:

location, the importance of the event, the activity of the user, social situation, and social activity [27].

There are 2 main types of notifications [28]:

- **Action-required notifications:** User is required to take some immediate action based on the information provided by the notification. e.g., Windows asking for administrative authentication.

- **Passive notifications:** User is presented with some information that does not require the user to take any action. e.g., calendar notification.

User-generated notifications focus on content created by other users, such as Gmail messages. In contrast, *context-generated notifications* are triggered by programs based on user preferences, like Google Alert. *System-generated notifications* originate from the system itself and can range from essential updates to spam or re-engagement notifications, depending on the app or device. Finally, *push notifications* are clickable pop-up alerts designed to inform users about new features, performance improvements, or suggestions.

In the following subsections, we'll go through two main mediums of notifications.

2.1.1 Visual Cues

Visual cues, as the name suggests, are the information provided to the user via the means of vision. These can include something as simple as an error LED light to as complex as AR notifications with interaction capabilities. Wallmyr et al. conducted a study that tested transparent interfaces based on mixed reality to facilitate the display of key information to the operators at a construction site [29]. After comparing mixed reality (head-up display and projected display) and head-down display, they found that users were more reactive to head-up display and experienced lower workloads. The position of the visual cues is also important. In [30], Harrison et al. presented the observations from their study by placing small lights on different body parts serving as visual cues. They observed that performance order, from high to low, is wrist, arm, brooch, shoulder, thigh, waist, and shoe. The reaction time performance highly depends on physical distance and visual accessibility, but also outside factors, like occlusion by furniture.

Visual notifications convey information through sight. Traditionally, they appear on flat screens like phones or monitors. These notifications can be simple pings with minimal animations, similar to clicking a bell icon that reveals more information. More elaborate examples include phone notifications that provide detailed messages. Phone notifications often display the app name and relevant details about the event that triggered the notification. These notifications can originate from apps informing users about status changes (e.g., low battery, game update) or messages from other users [31,32].

Visual notifications are not limited to personalized devices like smart TVs. Since the display can be shared by multiple people at the same time the notification mechanisms of other smart devices cannot be transferred to smart TVs. Three focus groups provided their impressions about notifications on TVs in a study conducted by Weber et al. [33]. They got responses about the duration of notifications, the amount of information displayed, the position of the notifications, and the number of notifications. Based on this, they created an application and conducted a study to create the following guidelines:

- Notifications that are truly important should be displayed.
- Privacy is important when multiple people are using the TV.
- Notifications should be displayed when there is a break.

2.1.2 Audio Cues

Audio cues offer an alternative to visual cues in situations where visual displays might be distracting or even harmful. This includes scenarios like driving, working on heavy machinery, etc. In [34], Lee et al. studied a collision detection system with audio and haptic cues to mitigate driver distraction. They observed that graded warnings (gradually increasing levels) were preferred by users and were the better-performing warning strategy than the single-stage warning strategy. They also observed that graded warnings weren't irritating and were more trustworthy for the users. In the case of haptic vs. audio, although users' performance was the same, their preference was haptic in trust, overall benefit to driving, and annoyance.

Another way audio notifications can be used is as a substitute for visual cues when the user is visually disabled. Crommentuijn et al. conducted a study to test different ways of providing audio or haptic cues when there is an obstacle in the way [35]. These findings suggest that carefully designed audio cues can significantly aid visually impaired users in navigating their surroundings.

2.1.3 Notifications in Mixed Reality

Many technologies combine in mixed reality devices to blur the line between the virtual and real worlds. This covers everything from fully immersive headsets to smartphones with augmented reality capabilities. In this section, we will talk about HMDs when referring to MR devices. The MR devices' primary input is through the eyes. As we move from AR to VR, the amount of immersion increases. While displaying notifications or interrupting users, there are certain considerations to keep in mind:

- Where to display information?
- How much occlusion is tolerable in different contexts?
- When to display information?
- What should be the behavior of the currently running application?
- What should be the interaction method for the user?

VR environments completely immerse users in a virtual world. Considering this, notifications in VR cannot adopt the framework used by other smart devices as this can break the immersion. There have been studies that propose different ways of showing notifications in VR. Zenner et al. proposed a framework where instead of stopping the application, the message is incorporated in the scenario [36]. If it's a medieval setting, a villager comes to the user and hands them a letter containing the information, or if it's a futuristic setting, a drone flies to the user to display a notification. They proposed the method of presenting notifications depends not only on the context of the virtual environment but also on the urgency of the notification. Although this avoids breaking the immersion, it also presents additional work for the developers. The developers must integrate ways to provide notifications, and further research is needed to explore user preferences, optimal notification design principles, and effective standardization methods. Unless VR starts being pervasive technology this will be extremely difficult to implement.

Ghosh et al.'s NotifiVR extends the notifications to cues (visual, auditory, or haptic) to bring the attention of the immersed user to events outside of the virtual environment [37]. In the case of audio cues, Ghosh et al.'s suggestions were to consider *Spatial Sound* (sound played should be at the position of notification in 3D space), *Skeumorphism* (metaphorical design like someone entering space as door opening sound), and *Gradual Intensification* (Gradual increase of sounds to signify increasing urgency). They proposed three haptic cues: *Gradual Intensification* (a gradual increase in vibration to indicate increasing urgency), *Directionality* (a vibration in the direction of notification), and *Skeumorphism* (metaphorical vibrations, such as feet vibrating to signify footsteps). Lastly, recommendations for visual cues included *3D Representation* (notices should be shown in 3D space rather than always being attached to the viewport), *Skeumorphism* (metaphorical representations, such as a clock face for an alarm), *3D Pop-ups* (notification blocks placed in 3D space, on a wall or other surface), and *Flash and Pause* (in dangerous situations, halting the immersion and allowing the user to reorient with the real-world space after stopping).

In the case of AR, there have only been a select number of studies regarding displaying notifications. Although AR has been researched in academic and industrial settings for specific use cases, there has been a movement by companies like Facebook [38], Apple [39] and Xiaomi [40] to bring AR to the general public. Lucero et al. conducted a study where participants were asked to walk down a busy road while keeping track of the notifications that popped up in their view frame called NotifEye [41]. They were provided a rubbing pad for simple interactions like opening and closing notifications. The participants reported that a combination of minimalistic information and rub pad created a non-distracting system for receiving notifications on the display. The main concern for the participants was social acceptability (size, comfort, fashion). All the current AR technologies that provide spatial tracking and interaction are bulky and make the user look like an alien or a bug.

2.2 Positioning Information

After learning about cues/notifications now we must check where to put these small bits of information in 3D space. As technology progresses towards all-day wearable displays, the need to study proper positioning becomes more prevalent. Rzayev et al. studied the effect of the position of text on a user's reading speed and comprehension while sitting or walking. [42] They compared 3 positions on the AR HMD: top right, center, and bottom center, and observed that the top right position increases workload and reduces comprehension. They also noticed that there is a decrease in reading speed when the participant is walking as opposed to sitting. The study did not test how the participants' state (busy with another task) while sitting affects the results. Similarly, Ku et al. conducted a study to check what is the preferred format of text while displaying it in peripheral vision [43]. They [43] found that title case capitalization and sans-serif font increased the efficiency. They also tested 2 retinal eccentricities and reported a preference of 5°over 8°.

In the case of the display itself, Jaschinski et al. performed a study with 38 participants, where the participants tried different positions of the screen (near vs distant, high vs low) [44]. Participants used the system for a full working day and experienced all the advantages and discomforts of different positions. The authors observed that when the participants were allowed to freely select their display position, they adjusted the display, thus signifying that the screen at about 66cm induced more strain than screens at about 98 cm. This study was conducted in or before 1999, and the display screen has improved drastically since then, so results may not hold true to this extreme amount now. The size of the display is also an important factor to consider when looking at user preference. As observed by Polys et al., a bigger display field is preferred [45]. In the future, the current limitation of small FoV in AR HMDs will improve, but at the moment, there are 2 ways to go around this limitation. Using devices like Varjo XR-3 [46], which use video pass-through technology to sacrifice lightweight devices for FoV of 115°. The second way is to simulate an AR environment by recreating the room in VR. This is called Mixed Reality Simulation. [47–49]

Imamov et al. created a Mixed Reality Simulation of a living room with all the basic furniture and decorations like a couch, lamp, paintings, etc. [50] The experiment aimed to determine the optimal position for world-stabilized glanceable content. It was discovered that top-left display positions were the slowest, whereas center and central-low display positions had the fastest interaction times and were the preferred positions. Regarding distance from the user, a two-meter position was chosen over a three-meter. It's interesting to note that participants claimed that looking down at the interface felt more comfortable and natural, which may have something to do with the headset's weight distribution and habits from using phones. It's crucial to remember that this study only looked at basic tasks; further research is needed to see how placement affects more difficult tasks that demand greater mental effort. Furthermore, there is a gap in our understanding of user behavior, such as possible head reorientation to view the interface, because eye gaze data was not recorded.

According to Plabst et al., notifications on AR HMDs work best when positioned in the real world or at the bottom center of the FoV, particularly when performing sustained concentration tasks [51]. Plabst et al. noted that users also favored these positions. Similarly, during an AR walking exercise, Lee et al. found that bottom FoV placement led to higher noticeability and comprehension for both icon- and text-type messages compared to top placement [52]. Lazaro et al. suggested employing both audio and visual cues for augmented reality notifications and proposed more investigation into the optimal location of notifications [53].

Chapter 3

ARtisan Bistro [1]

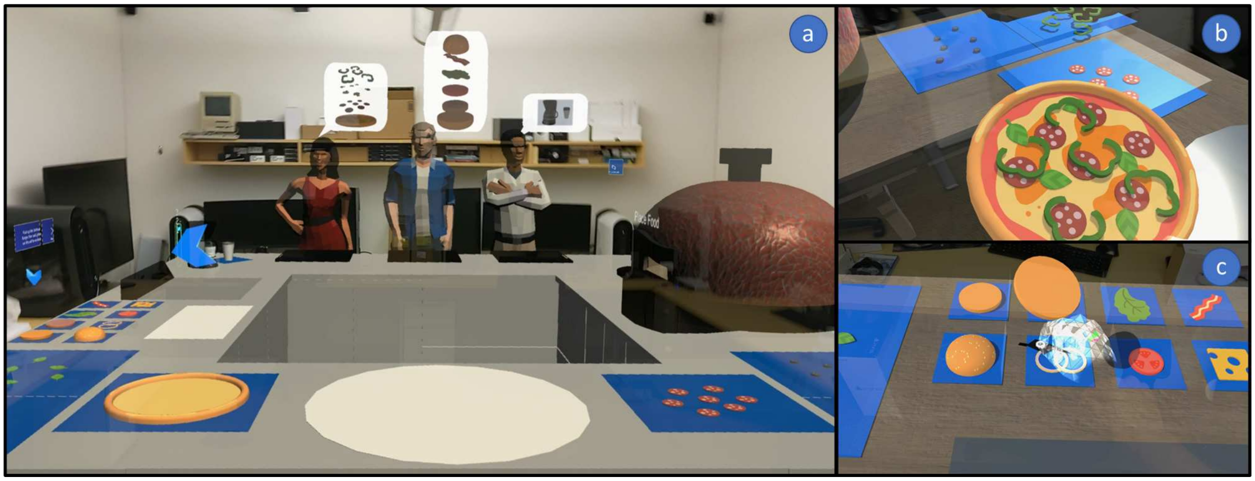

Figure 3.1: (a) Complete Setup of ARtisan Bistro; (b) Perfectly Cooked Pizza Taken out of the Oven; (c) Making Burger, Picking up Bottom Bun by Hand

This chapter explores the design and features of an AR cooking environment. It showcases and adds further details to a previous publication. Aditya Raikwar, Lucas Plabst, and Francisco R. Ortega introduced "ARtisan Bistro: A Cooking Task Environment to Conduct Studies in Augmented Reality" in the proceedings of the 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) [54].

3.1 Introduction

ARtisan Bistro is an AR environment that resembles a fast food restaurant where users may prepare burgers, pizza, and coffee upon customers' requests. It was created using the Unity engine v2022.1.10f1 and the Mixed Reality Toolkit (MRTK). The researchers can

[1] The content of this Chapter is identical to that presented in: A. Raikwar, L. Plabst and F. R. Ortega, "ARtisan Bistro: A Cooking Task Environment to Conduct Studies in Augmented Reality," 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Singapore, Singapore, 2022, pp. 909-910, doi: 10.1109/ISMAR-Adjunct57072.2022.00200.

use this environment to evaluate different user interfaces in a general setting. More time can be dedicated to research if researchers are not required to build and execute a testing environment, in turn saving time and resources. Currently, many experiments employ specialized environments for highly specific use cases. This environment can be utilized by researchers who do not need a specific goal but rather a context in which they can test their interfaces.

With flexibility in mind, ARtisan Bistro was created as a sandbox setting for AR research. Most people can relate to cooking, therefore it requires little to no explanation and thus makes it simple for the participants to perform the appropriate tasks. There is no need to learn new interaction techniques because every object in the environment can be interacted with almost as if it were a real physical object. Additionally, the virtual environment mitigates potential risks of injury during the experiment. While cooking is a straightforward and accessible task, it may also serve as a metaphor for more complex work, particularly in high-pressure environments. For example, a busy restaurant kitchen's fast-paced and demanding nature parallels the stress and time constraints experienced in an operating room during surgery. However, simulating a complex medical procedure can be challenging for participants unfamiliar with the medical field. The stressful circumstances of a surgery room can be translated into a more familiar and relatable atmosphere for experiment participants using the culinary task.

3.2 Features

3.2.1 Components and Interactions

ARtisan Bistro consisted of three primary cooking tasks: a Burger Station, a Pizza Station, and a Coffee Station. Each station simulated an essential food preparation task, allowing users to interact with virtual ingredients and tools.

Burger Station: The ARtisan Bistro environment includes a dedicated burger station for user interaction. This station consists of two key components:

(1) Ingredients: The virtual burger assembly task provides users with eight selectable components: bottom bun, top bun, burger patty, tomato, lettuce leaf, bacon, cheese, and onion. Each component can be spawned in unlimited quantities. Notably, the bottom and top buns are designated as essential elements for burger completion. Users can manipulate these virtual ingredients using intuitive techniques, similar to interacting with physical objects. Ingredient stacking functionality allows users to build their burgers by placing items on top of the bottom bun. These stacked components can be lifted and moved as a unit while maintaining relative positions. The top bun serves as the final placement layer, signifying burger completion and preventing further additions on top.

(2) Grill: A virtual grill is present for cooking burger patties. Users can place multiple patties on the grill simultaneously for parallel cooking. Each patty undergoes a visually represented cooking process with five distinct stages, ranging from uncooked to burnt, depicted through changes in surface texture. The system allows for adjustable cooking speed, enabling researchers to control the cooking duration for experimental purposes.

Pizza Station: The ARtisan Bistro environment includes a dedicated pizza station for user interaction. This station consists of two key components:

(1) Ingredients: Users can choose from five main ingredients for their virtual pizza: pepperoni, mushrooms, bell peppers, olives, and basil leaves. Each ingredient has an unlimited spawning capacity, allowing for customization. The necessary pizza base comes pre-topped with sauce and cheese. Users can place the additional ingredients on top of the base, with a snapping mechanism ensuring clear visual distinction of each element.

(2) Oven: Once users are satisfied with their pizza creation, they can place it in the virtual oven. Each pizza undergoes a visually represented cooking process with five distinct stages, ranging from uncooked to burnt. Similar to the grill, the system allows for adjustable cooking speed, enabling researchers to control the cooking duration for experimentation.

Coffee Station: The ARtisan Bistro environment offers a virtual coffee station for user interaction. This station features:

(1) Coffee Machine and Cup: Users can interact with a virtual coffee machine and cup to brew coffee. The virtual cup visually depicts a rising liquid level as the brewing process progresses. A lid appears on top when the cup is full to indicate completion.

(2) Coffee Pot and Level Indicator: A gauge on the side of the coffee machine displays the remaining coffee in the pot. This gauge utilizes a three-tier system (levels 1, 2, and 3) where each level corresponds to the number of cups that can be filled with the remaining coffee. Users initiate brewing by pressing a virtual button at the base of the coffee machine. The process pauses automatically when the pot is removed and resumes when replaced. Similar to other cooking elements, the system allows for adjustable brewing speed in the code, enabling researchers to control the duration for experimental purposes.

Following are the references to the assets used to create the virtual environment for the experiments:

1. Fast food [55]: This set includes 3D models of pizzas and burgers with all their ingredients.

2. Pizza Oven [56]: A 3D model of a pizza oven.

3. Coffee Maker [57]: A 3D model of a coffee maker. Some model changes were made in Blender 3D modeling software [58] to allow dynamic coffee levels.

4. Coffee Cup [59]: A 3D model of a Starbucks coffee cup. Some model changes had to be made in Blender 3D modeling software [58] to allow dynamic coffee levels.

5. Customer Tray [60]: A simple 3D model of a tray for carrying food items.

6. Customers [Citation not available anymore]: Rigged character models representing the virtual customers within the restaurant environment.

3.2.2 Additional Features

Customers: There are 12 meshes available for buyers (6 female and 6 male). There are two animations associated with each consumer. Three clients at once are the most that can be accommodated. When the client is spawned, the animation and mesh are selected at random. The number of customers, duration/level of their patience, and the length of time between new customers are easily modifiable.

Feedback Mechanisms: The ARtisan Bistro environment incorporated visual and auditory cues to guide user actions and provide feedback. For instance, sizzling sounds indicate real-time feedback at the start of the grilling process of the burger patties, and a textual prompt, "Wrong Item," appears when the user serves food that does not match the customer's request, offering immediate visual feedback. This multimodal approach enhances user experience by creating a more intuitive and informative virtual environment.

3.3 Customization and Control

ARtisan Bistro was designed with customization in mind. Various aspects can be manipulated through code. To facilitate this, the system incorporates a range of configurable parameters, which are categorized into three primary types:

Time-Based Variables: These variables allow researchers to manipulate the duration of specific tasks. This includes adjusting the cooking time for burger patties and pizzas, the rate at which coffee fills the cup and pot, the wait time for customer orders, and the interval between new customer arrivals. This level of control enables researchers to tailor the pace of the experiment to their specific needs.

Quantity-Based Variables: These variables govern the number of ingredients included in customer orders for both pizzas and burgers. While the base ingredients remain consistent, the quantity of additional toppings can be adjusted. This allows researchers to explore how varying ingredient complexity impacts user performance.

Appearance-Based Variables: As previously mentioned, the system utilizes a set of twelve available meshes. These variables enable researchers to introduce a selection of additional meshes into the environment beyond the base set. This functionality can be used to investigate potential biases in user behavior, such as implicit biases towards certain customer types based on visual cues associated with the additional meshes.

By providing these configurable parameters, ARtisan Bistro allows researchers to conduct controlled experiments within a customizable virtual restaurant environment, allowing them to explore various factors that might influence user behavior. This level of control allowed me to create various scenarios and test the effectiveness of notifications in different contexts.

3.4 Limitations

While ARtisan Bistro offered a valuable controlled environment for our research, it is essential to acknowledge some limitations:

Transfer of Skills: While the environment was designed to metaphorically represent high-pressure kitchens and similar task-oriented fields, the skills learned within ARtisan Bistro (e.g., ingredient selection, virtual cooking) do not directly translate to real-world culinary expertise.

Field of View: The Microsoft HoloLens 2, has a smaller field of view (FOV) compared to the human eye. This limitation restricts the amount of visual information that can be displayed or tracked within the user's view. This may influence the users to interact with virtual objects differently compared to real life.

Hand Tracking: The HoloLens 2 employs inside-out hand-tracking technology. This means that user hand movements are only tracked when the user is actively looking at their hands. This could have potentially impacted the natural flow of tasks and potentially user frustration, particularly when trying to interacte with ingredients or tools outside the direct line of sight.

Despite these limitations, ARtisan Bistro provided a strong foundation for investigating user interaction with notifications in an AR context. The controlled environment allowed us to isolate variables and manipulate notification delivery to observe user behavior and gauge the effectiveness of different notification designs.

Chapter 4

Ping! Your Food is Ready: Comparing Different World-Fixed Notification Techniques in 3D AR Cooking Environment

This chapter explores the design and integration of notifications in an AR cooking environment. Effective notifications serve as essential communication channels, alerting users to important updates, task instructions, or changes in their virtual environment. This chapter investigates how the design of these notifications can be optimized to ensure users are aware of critical information while minimizing disruption to the cooking workflow.

The research presented in Chapter 4 has been accepted for publication in the prestigious IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2024 (Reference to arxiv [61]). I have chosen to include the full details of this study within the dissertation to provide a comprehensive understanding of the factors influencing user experience in AR environments. By focusing on notification design in Chapter 4, we establish a foundation for understanding how notification modality, placement, and user preferences can influence user performance and overall satisfaction within AR applications.

4.1 Introduction

Intuitive and informative notifications are crucial for seamless interaction with Augmented Reality (AR) interfaces. Unlike traditional interfaces with dedicated screen space, AR displays information directly within the user's FoV. This unique environment presents challenges in designing notifications that effectively guide users without disrupting their

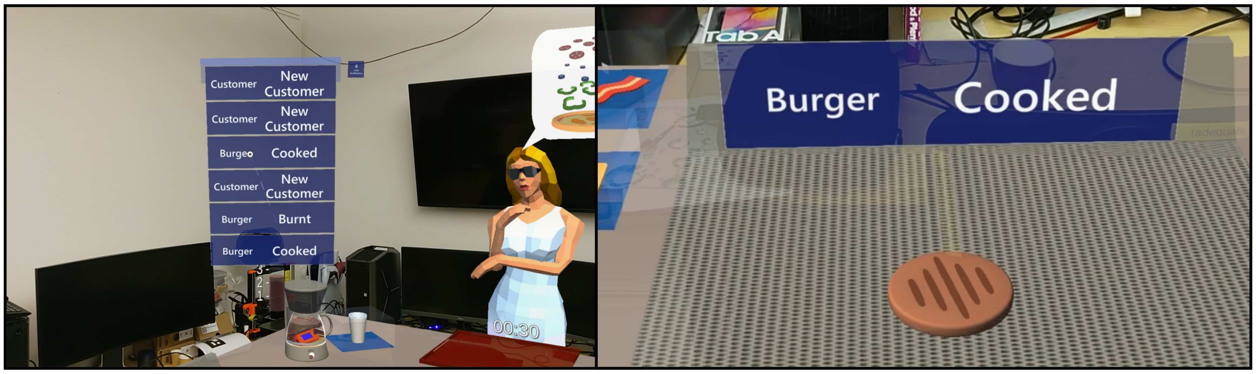

Figure 4.1: Left: Notifications on Dock; Right: Notification on Object

primary tasks. This chapter explores how visual and audio notifications, strategically placed within the AR environment, can enhance user experience. Taking advantage of the additional dimension the AR environment provides, the notifications are designed to be displayed in the 3D space and attached to the environment (fixed to locations and objects) rather than the user. This approach is known as world-fixed notifications [62].

I designed two types of world-fixed visual notifications in this study: "Notifications on Dock" and "Notifications on Object" (illustrated in Figure 4.1). I further explored the effectiveness of accompanying these visual notifications with localized audio notifications. A spatially-positioned bubble-popping sound effect drew attention to the visual notification's location in 3D space.

The ARtisan Bistro environment [54] served as the foundation for this user study. The parameters, such as customer arrival frequency and cooking speed, were intentionally manipulated within ARtisan Bistro to create a simulated high-pressure environment that demanded focused attention from the participants. Notifications were then introduced at specific moments to assist participants in keeping track of working on multiple stations simultaneously.

The findings demonstrate that strategically placing notifications directly on relevant objects within the AR environment can significantly improve task performance and user preference. Additionally, the study highlights the crucial role of audio notifications in capturing user attention. The introduction of audio notifications increased the noticeability of visual notifications, ultimately leading to improved performance and higher user satisfaction. These results underscore the importance of combining well-designed visual and audio notifications to optimize user experience in AR applications.

4.2 ARtisan Bistro

ARtisan Bistro is an AR environment designed to simulate a fast food restaurant where users can prepare items like burgers, pizza, and coffee based on customer requests [54]. Developed using Unity engine v2022.1.10f1 and the Mixed Reality Toolkit (MRTK), the application serves as a tool for researchers to evaluate user interfaces [63]. The application was designed for researchers to evaluate user interfaces in general. We deliberately chose this open-source solution since it provides a standardized testbed for comparison of future results and replicates a familiar cooking scenario that is relatable to a wide audience. 3.1 shows different stills from the ARtisan Bistro environment.

We also kept the entire environment in (AR) for two reasons: (1) to enable accurate and meaningful comparisons with future experiments that will use real objects, thereby assessing the impact of real objects on participant behavior and outcomes; and (2) to enhance the study's external validity through random visual clutter.

4.2.1 Environment Settings

It is possible to change and set the environment in ARtisan Bistro. In this paper, we selected settings that simulate a high-stress environment, necessitating notifications to assist in task completion and promote multitasking across multiple food items concurrently.

Burger Station: Participants were instructed to cook patties and assemble burgers with six layers: bottom bun, top bun, patty, and three random ingredients. The patty must be placed directly above the bottom bun, but the order of the other ingredients is flexible. The different cooking statuses for the patty were Uncooked, Cooked, and Burnt, with each status taking ten seconds to achieve.

Pizza Station: Participants were instructed to assemble and cook pizzas that consisted of a base and three random ingredients, with the order of ingredients being non-essential. Unlike checking the cooking status of a patty, which is straightforward, determining the status of a pizza requires participants to stand in front of the oven and look through a small window. The cooking status of the pizza could be Uncooked, Cooked, or Burnt, with each stage taking ten seconds to achieve.

Coffee Station: The coffee maker was mostly set to default settings, except participants did not need to press the start button. At the beginning of the level, the coffee pot's level resets to zero, and the coffee maker starts filling it automatically. The time to fill a cup was set to ten seconds of continuous pour.

Customers: Among the twelve meshes available for customers, three females and three males were selected at random. Except for the tutorial level, all levels consisted of 6 customers who waited for their requested food for only two minutes. The time remaining was displayed right in front of the respective customer.

The environment was engineered to subject participants to stringent time constraints while incorporating slight physical demand elements. This can be confirmed using the overall Raw NASA-TLX scores [64]. The average overall workload experienced by participants was 60.64 (SD = 10.84). This is above average global workload score reported in [65]. The average raw scores are listed in 4.1.

4.3 Notification Design

This section dives into the design of the notification system implemented in our study. I explored two modalities for notifications: visual and audio. Depending on the designated condition, participants were exposed to notifications that were either visual, audio, both visual and audio, or none at all.

Table 4.1: Raw NASA-TLX scores

Dimensions	Mean	STD
Mental Demand	67.31	19.52
Physical Demand	51.92	23.86
Temporal Demand	75	19.86
Performance	65.38	20.8
Effort	71.56	15.86
Frustration	32.69	17.22

4.3.1 Visual Notifications

Figure 4.2: Notification when a new customer arrives.

AR-HMDs provide a unique platform for notification delivery compared to traditional interfaces. Unlike fixed-screen displays, AR-HMDs offer a dynamic canvas that can leverage the 3D environment for notification presentation. Zenner et al. proposed integrating notifications into the scenario itself to maintain immersion [36]. However, this approach might limit notification types and rely heavily on manual prompts to send messages. Additionally, subjectivity in message priority can create a burden on the sender to assign appropriate levels, leading to user frustration if notifications are not deemed important by the receiver.

To address these limitations, I designed two types of visual notifications suitable for handling a broader range of notification scenarios.

Visual Notification Design and Properties

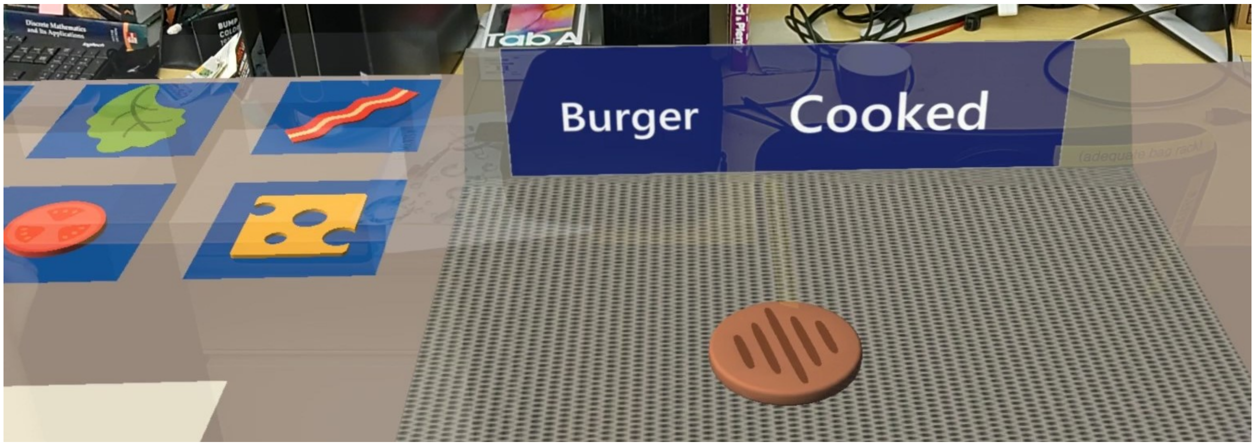

Figure 4.3: Notification Button.

The implemented visual notifications, illustrated in Figure 4.3, displayed information in a clear "Station Message" format. The "Station" field indicated the source station's name, while the "Message" field conveyed relevant information specific to that station. These cuboid-shaped notifications featured navy blue backgrounds with white text, as per the design recommendations for buttons outlined by Hololens 2 developers. To ensure users had adequate time to process the notifications, they remained visible for seven seconds before automatically disappearing. This duration, determined during the experiment phase, proved sufficient for users to pause their primary tasks, read the notification content, and grasp its message. Any notifications that remained unseen after this time frame were considered disregarded by the user.

The interval between notifications was dynamic and influenced by user actions and the virtual environment. For example, a customer might wait for 120 seconds (2 minutes) after placing their order, with a new customer arriving 5 seconds after the previous customer departs. In this scenario, the notification interval between these customers would be 120 seconds + 5 seconds = 125 seconds. However, if the participant served the first customer within 65 seconds, the notification interval for the next customer would be reduced. In this case, the interval would be calculated as 65 seconds + 5 seconds = 70 seconds.

Four notification messages were implemented to inform participants about various status changes in the AR environment:

1. **New Customer:** When a new customer arrives and requests food.
2. **Cooked:** When either the burger patty on the grill or the pizza in the oven reaches the desired cooking state based on customer preference.
3. **Burnt:** When either the burger patty on the grill or the pizza in the oven becomes overcooked and exceeds the customer's desired cooking state.
4. **Coffee cup added:** When the coffee level in the coffee maker reaches a threshold sufficient to fill a complete cup.

Notification on Object (NoO): Context-Aware Notifications

Figure 4.4: Notification on Object

Inspired by Zenner et al.'s concept of immersive notifications [36], I created Notification on Object (NoO) to display notifications. This approach presents notifications directly within the 3D space, adhering to the context of the surrounding environment. This approach aimed to provide users with relevant information without overwhelming them, particularly when focusing on specific tasks. For example, during food preparation, notifications about the pizza's cooking status could be displayed directly above the oven, allowing users to respond quickly without breaking immersion (Figure 4.4).

In the experiment, object-based notifications were displayed in a single-box format. If a new notification from the same station was prompted before the current one disappeared, the existing notification would update its text with the new information, and the timer to dismiss the notification would reset. While this approach reduces clutter and potentially improves response times, it also increases the risk of missed notifications. This could occur if the notification falls outside the user's field of view or if they fail to glance at it in time. Potential solutions include using audio alerts for new notifications or offering a centralized list of all notifications for periodic review.

Notification on Dock

The Notification on Dock (NoD) concept draws inspiration from smartphone notifications, where all notifications are displayed in a readily accessible location by swiping down on the screen [31, 32]. This approach presents notifications in a centralized list format, simplifying information access and reducing the cognitive load associated with locating individual notifications scattered throughout the environment. Although users might miss notifications if they are scattered throughout the environment, a centralized list allows for periodic review. Unlike object-based notifications, NoD cannot provide additional information beyond the text itself (like the location of the object that triggered the notification). Therefore, it becomes crucial to clearly indicate the station that generated the notification within the message itself.

Figure 4.5: Notification on Dock

Building on this concept, NoD is visualized as a drop-down list displaying notifications in reverse chronological order. As notifications disappeared, subsequent ones would automatically move up within the list to fill the vacant space. Users could position the dock anywhere within the 3D space, allowing for customization and optimal placement based on comfort or task demands (Figure 4.5).

However, Notification on Dock also has limitations. While it simplifies locating notifications, users need to find an optimal position within the 3D space. Inefficient placement can negatively impact the time it takes to find visual objects in a 3D environment [66]. Additionally, the dock itself doesn't inherently provide location information for the notification source, necessitating extra details within the message text.

4.3.2 Audio Notification

Audio notifications offer a valuable solution to address the limitations of visual notifications, particularly in scenarios where visual cues might not be the most effective primary notification method [35]. Similar to how smartphones allow users to set distinct notification tones for various applications, AR environments can utilize different sounds for different notifications. This auditory distinction allows users to identify and prioritize notifications without needing to glance at the application name or notification content, enhancing the overall user experience.

A single, localized sound clip for notifications was employed to avoid over-complication and increase the learning curve in the experiment. This approach enabled participants to leverage spatial audio cues and determine the direction of the sound source. While the Hololens 2 speakers, unlike true surround sound systems, lack the capability for full 360-degree sound projection, they provided sufficient left-right differentiation for our purposes. Since the notifications were confined to the XZ plane (horizontal and depth), the Hololens 2 speakers were adequate for this specific scenario. This localized audio played a crucial role in helping users identify the direction and source of the notification, such as specific cooking stations, ultimately leading to faster response times.

The origin of the sound directly corresponded to the notification's spawn location:

Notification on Dock (NoD): The sound originated from the dock itself.

Notification on Object (NoO): The sound emanated from the location above the object that prompted the notification.

Sound-Only Condition: In the absence of visual notifications, the sound played as a standard audio clip, seemingly originating from within the user's head.

This single sound notification informed participants without introducing complexity through various notification sounds. However, in a real-world application, designers could leverage diverse sound designs to create a richer notification experience, allowing users to distinguish between different notification types based on sound alone.

4.4 Methodology

4.4.1 Research Questions

Building upon prior research, this study investigated the following research questions (RQ) to gain insights into user experience with AR-HMD notification systems:

RQ1: ***Impact of Multimodal Notifications on User Performance (Serving Customers)*** Do world-fixed notifications, compared to a no-notification baseline, improve user performance in terms of serving customers efficiently and accurately within a time limit?

Building on the understanding that visual and audio notifications can enhance user attention and reduce task load [13, 26], this study investigated the effectiveness of world-fixed notification designs in supporting user performance and satisfaction within an AR environment. Specifically, I wanted to explore how these notifications, presented at different locations within the AR space, can influence users' ability to manage tasks and interruptions in the context of serving customers, using MRT [67].

RQ2: ***Identifying the Most Noticeable Notification Types*** Among the various notification designs (visual only, audio only, combined visual and audio), which types are most successful in capturing users' attention and ensuring they don't miss critical information about customer requests?

In a virtual environment where users can become engrossed in tasks, effective notifications are crucial for ensuring the timely delivery of critical information. This research question seeks to determine which notification types are most prominent and readily noticeable to users, minimizing the risk of missed information.

RQ3: ***User Preferences for Visual and Audio Notifications*** How do users perceive the different notification design approaches in terms of ease of use, clarity, and overall user experience? Do they have preferences for specific notification types that might influence their long-term adoption and satisfaction within AR environments?

Previous studies have shown that user preferences can significantly impact the perceived effectiveness and user acceptance of notification systems [34]. By understanding user preferences, more user-friendly and accepted systems can be designed in the context of AR-HMD notifications.

By addressing these questions, I hope to gain valuable insights into how world-fixed notifications can be designed to optimize user experience within AR workflows. The goal is to create notification systems that effectively support task management while minimizing disruption and promoting user comfort and acceptance of this technology.

4.4.2 Experiment Design

This study employed a within-subjects design with two independent factors to evaluate the effectiveness of world-fixed notifications:

Visual Notification (V_3): This factor had three levels - Notification on Object (NoO), Notification on Dock (NoD), and no visual notification (Control).

Auditory Notifications (A_2): This factor had two levels - with audio notification (WS) and without audio notification (WoS).

The combination of these factors resulted in six experimental conditions. NoO WoS - Participants received visual notifications displayed directly on the virtual object requiring attention. No audio notifications were presented. NoO WS - Participants received both visual notifications on the object and an accompanying audio notification. NoD WoS - Participants received visual notifications displayed on a fixed dock within the AR environment. No audio notifications were presented. NoD WS - Participants received both visual notifications on the dock and an accompanying audio notification. Control WoS - This baseline condition presented no visual notifications. Participants relied solely on other cues and changes in the environment to manage tasks. Control WS - This baseline condition included only the audio notification without any visual notifications.

Figure 4.6 illustrates the overall flow of the experiment. Each participant completed all six conditions in a randomized order, ensuring a balanced evaluation across notification designs.

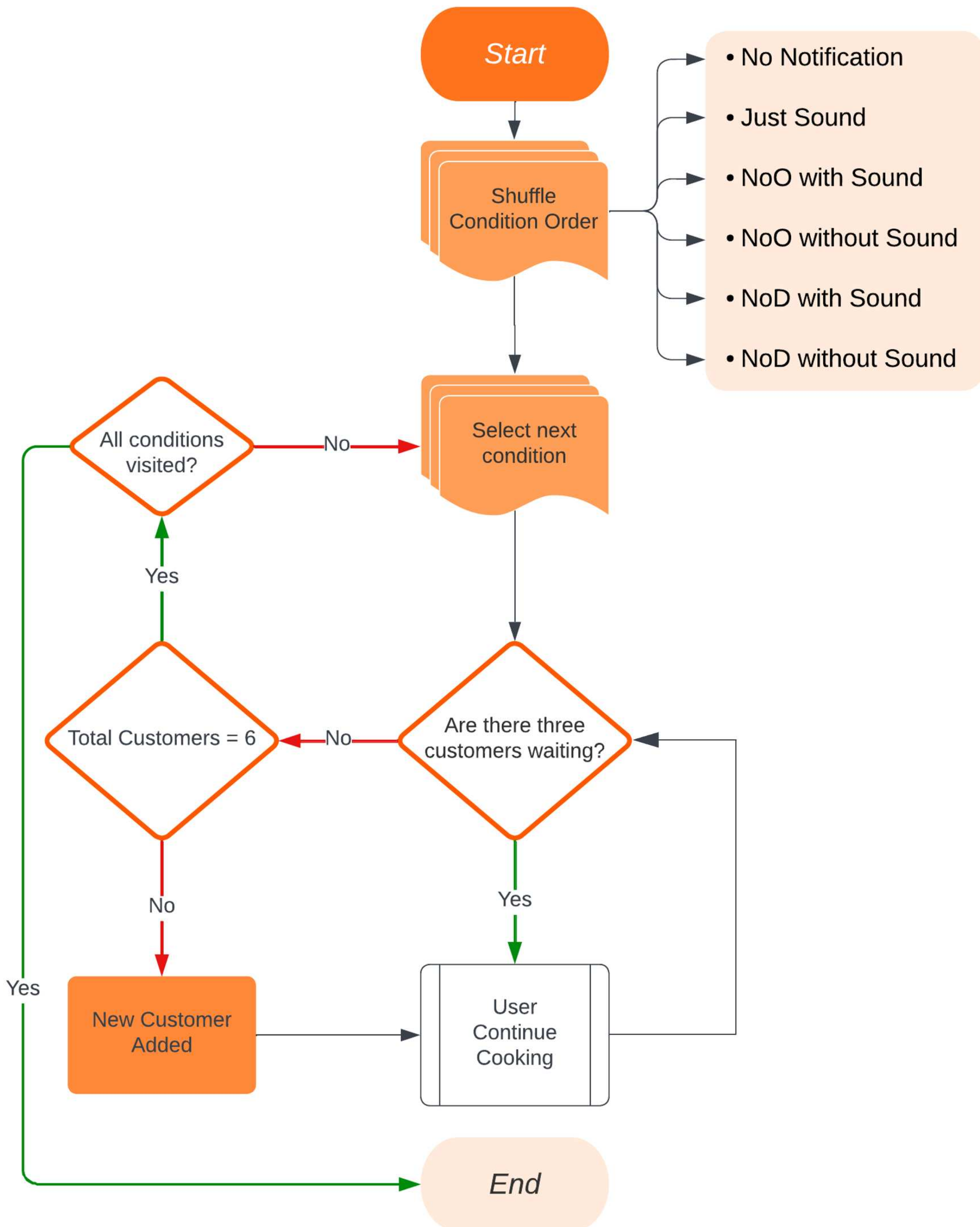

Figure 4.6: Experiment Flowchart

The experiment followed these steps:

1. **Arrival and Introduction:** Participants arrived at the research lab, provided informed consent, and received a detailed explanation of the study.

2. **Familiarization:** Participants watched a short video introducing the AR environment and the different notification types they would encounter throughout the experiment.

3. **Training Level:** Participants completed a practice level using a Hololens 2 headset. This level involved making food items for three virtual customers without a time limit, allowing them to familiarize themselves with the controls and virtual kitchen environment. During this level, participants were allowed to ask questions and guided through the steps to prepare food if needed.

4. **Experimental Levels:** Participants proceeded through the six experimental conditions after completing the training level. The order of conditions was randomized for each participant using a chi-square model.

    (a) At the beginning of each level, participants received instructions regarding the specific notification design they would experience (e.g., Notification on Object with Sound).

    (b) Each level presented a simulated customer service scenario where virtual customers requested food items within a set time limit. Participants used virtual objects to prepare the food while receiving notifications about cooking status and new customers.

    (c) The chosen notification sound was a subtle bubble-popping sound to minimize distraction and maintain neutrality. Visual notifications remained active for 7 seconds in NoO and NoD conditions.

(d) Participants acknowledged notifications by calling out a randomly generated number in the notification text (to verify that they have indeed read the notification).

5. **Post-Surveys:** Upon completing all six levels, participants answered post-experiment surveys to convey their subjective experience with the different notification designs. These surveys assessed ease of use, workload, clarity, and user preference. Due to time constraints, no verbal interview was conducted.

The experiment recorded the following data to assess user performance and experience:

Table 4.2: Data measures recorded

User Actions		
Experiment start	Experiment end	New notification
New ingredient spawn	Ingredient deleted	Notification viewed
Food presented to a customer	Pizza put in the oven	New Customer
Cooking Status Change	Coffee cup added	Customer leaves
Performance Metrics		
Customer Served	Customer Departure	Customer Waiting Time
Notification Recognition Frequency	Reaction Time	Post-Survey Responses

This comprehensive data collection strategy allowed for a detailed analysis of user behavior, performance metrics, and subjective experience across the different notification design conditions within the AR customer service scenario. The data was later preprocessed to prepare it for further statistical analysis.

4.4.3 Participants

The experiment involved 26 participants, equally divided between males and females, aged 23 to 70 (M = 32.73, STD = 11.9). All participants identified as either male or female. Participants consisted of students and staff from Colorado State University and people who were not affiliated with the university. 84.6% of participants were either students or faculty. 69.23% of the participants reported using at least one AR or VR device. Participants received compensation in the form of $20. Among the participants, 47.6% reported that they have worked in a restaurant in some capacity (this includes as a cook or a waiter). This study was approved by the university's internal review board.

4.4.4 User Feedback During Design Process

User feedback obtained throughout the design process yielded several valuable insights. While participants initially found the interaction challenging, they reported a gradual increase in ease of use over time. As anticipated, Notification on Object with Sound (NoO WS) was the preferred notification configuration. Participants noted that the accompanying sound effectively captured their attention, while the localized nature of the audio further guided their focus toward the relevant object. The visual notification then confirmed the information conveyed through the audio notification. Overall, participants viewed the notifications as a positive addition to the AR experience.

One suggestion offered by a participant involved presenting a brief introductory video showcasing the virtual kitchen environment, the tasks, and the notifications before the start of the experiment. The video may potentially familiarize users and reduce initial learning time. Additionally, observations revealed that some tasks took longer than anticipated due to limitations with the Hololens 2 hand-tracking capabilities. For instance, participants encountered difficulties starting the coffee maker due to unreliable hand registration, preventing the coffee from brewing. To address this challenge, two potential solutions were considered:

Key-Binding Control: Introduce a key-binding to an external keyboard, allowing the experiment supervisor to remotely control the start and stop functions of the coffee maker.

Pre-Activated Coffee Maker: Eliminate the need for user interaction by keeping the coffee maker pre-activated throughout the experiment, similar to the approach used with the pizza oven.

Ultimately, the chosen solution involved keeping the coffee maker continuously active, removing the need for user intervention, and mitigating the potential impact of hand-tracking limitations.

4.5 Results

After pre-processing the data, a two-way analysis of variance (ANOVA) was used to analyze the data collected during the experiment. In cases where the data violated the assumptions of ANOVA, we applied the Aligned Rank Transform (ART) procedure for non-parametric analysis, as described by Wobbrock et al. (2011) [68]. If significant main or interaction effects were identified, we conducted post hoc pairwise comparisons using Tukey's Honestly Significant Difference (TukeyHSD) test. However, for data transformed with ART, the appropriate post hoc analysis procedure involved the ART-C method proposed by Elkin et al. [69].

4.5.1 Performance

The analysis focused on the number of customers served successfully within the time limit as a measure of user performance. Figure 4.7 depicts the average number of customers served across the different notification design conditions. The x-axis represents the different conditions (3 conditions of visual notification and 2 conditions of audio notification). The y-axis represents how many customers were served on average in each condition (maximum is 6). Error bars represent the standard error of the mean. Detailed information on individual means, standard deviations, and standard errors can be found in Table 4.3.

The results revealed a statistically significant interaction effect between visual and audio notifications ($F(2,125) = 4.14$, $p < .05$). Further analysis using simple main effects tests indicated that neither visual notifications ($F(2,125) = 2.01$, $p = .14$) nor audio notifi-

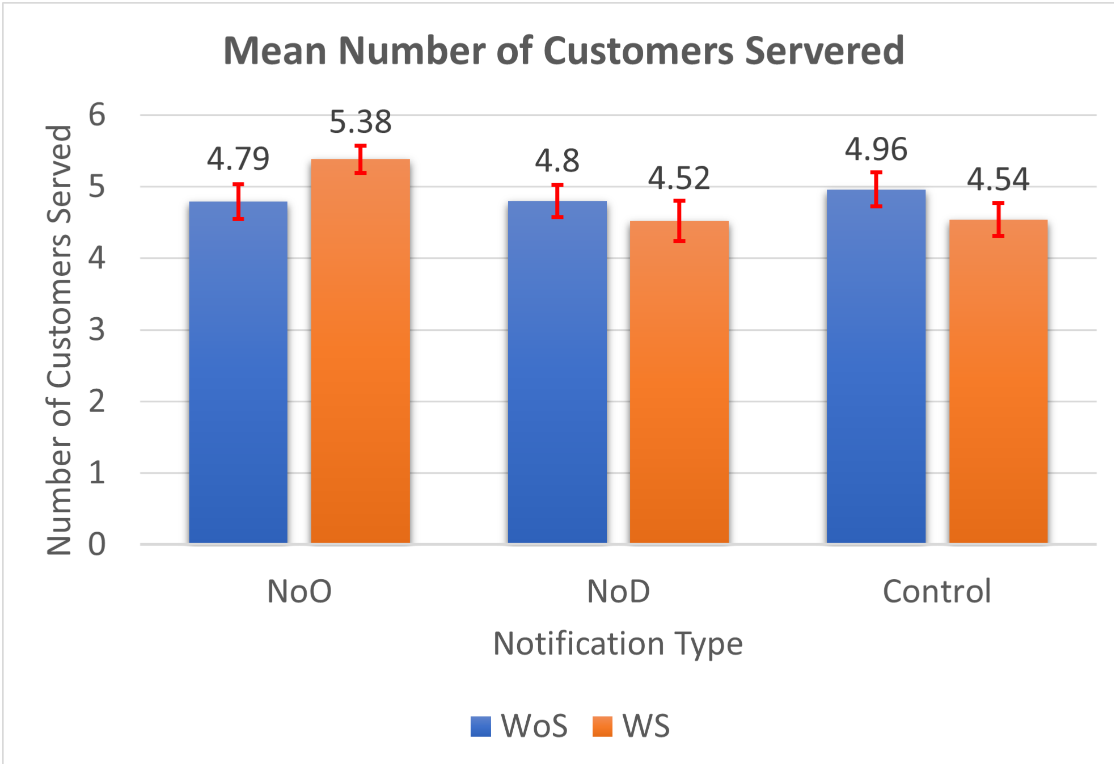

Figure 4.7: Performance of participants in different notification conditions based on how many customers they served.

Table 4.3: Mean performance of participants in different notification conditions

Notification Techniques	Mean	STD	SE
NoO WoS	4.79	1.24	0.24
NoO WS	5.38	0.96	0.19
NoD WoS	4.8	1.14	0.22
NoD WS	4.52	1.44	0.28
Control WoS	4.96	1.19	0.23
Control WS	4.54	1.18	0.23

cations ($F(1,125) = 0.27$, $p = .61$) had a significant main effect on the number of customers served on their own.

However, in post hoc contrast tests using the ART-C procedure, participants in the NoO WS condition served significantly more customers compared to both NoD WS ($t(125) = 3.52$, $p < .01$) and Control WS ($t(125) = 3.16$, $p < .05$) conditions. There were no statistically significant differences observed between any other notification design combinations. During the ART-C procedure, the degrees of freedom are calculated using the Kenward-Roger Method [70].

These findings suggest that while overall notification type (visual, audio, or both) might not have a major impact on serving customers, the specific location of visual notifications in conjunction with audio notifications can influence performance. Combining visual notifications directly displayed on the served object with accompanying audio notifications appears to be most beneficial for user performance in this task.

I investigated whether prior restaurant experience influenced user performance and perceived task load. An independent samples t-test revealed no statistically significant effect of restaurant experience on the number of customers served ($t(24) = 1.6$, $p = .12$). While participants with restaurant experience served slightly more customers ($M = 5.12, SD = 0.72$) compared to those without experience ($M = 4.57, SD = 1.01$), the difference did not reach statistical significance. Similarly, the NASA-TLX scores indicated no significant difference in overall perceived task load between participants with ($M = 50.15, SD = 13.33$) and without ($M = 49.62, SD = 11.56$) restaurant experience ($t(24) = 0.11$, $p = .91$). A further breakdown of the NASA-TLX sub-scales revealed no significant difference in perceived temporal demand between the two groups ($t(24) = -1.07$, $p = .29$). Participants with restaurant experience reported a temporal demand score of ($M = 7.08, SD = 2.36$), while those without experience reported ($M = 7.92, SD = 1.61$). These findings suggest that prior restaurant experience did not exert a significant influence on either user performance or perceived task load within the context of this experiment.

4.5.2 Visual Notifications Called

The analysis focused on the number of times participants actively acknowledged visual notifications by verbally calling out the confirmation number displayed within the notification text. The average number of visual notifications called out across the different notification design conditions is illustrated in Figure 4.8. The x-axis represents the different conditions (2 conditions of visual notifications and two conditions of audio notifications). The y-axis represents how many notifications were called out on average in each condition. Error bars represent the standard error of the mean. Detailed information on individual means, standard deviations, and standard errors can be found in Table 4.4

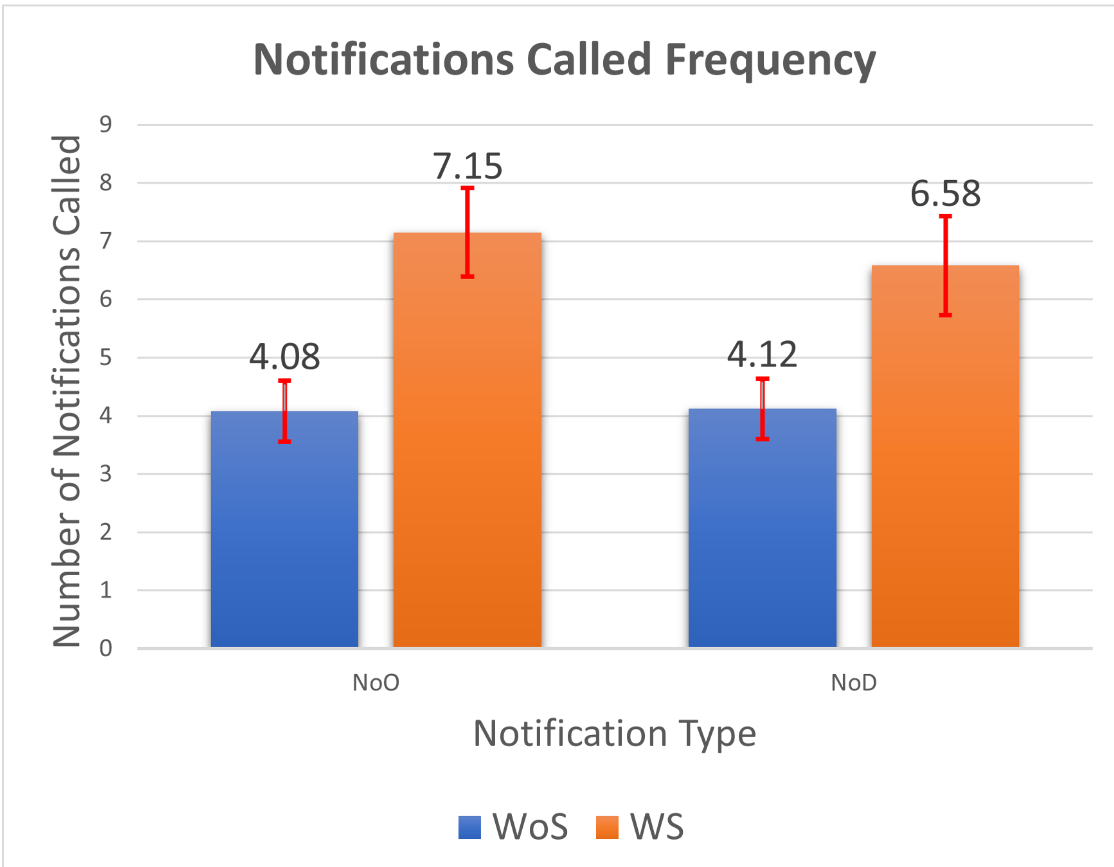

Figure 4.8: Average number of notifications called in different notification conditions

Table 4.4: Mean number of notifications called in different notification conditions

Notification Techniques	Mean	STD	SE
NoO WoS	4.08	2.66	0.52
NoO WS	7.15	3.88	0.76
NoD WoS	4.12	2.65	0.52
NoD WS	6.58	4.32	0.85

The analysis revealed no statistically significant interaction effect between visual and audio notifications ($F(1,100) = 0.2, p = .66$). Looking at the analysis of simple main effects showed that the presence of visual notifications alone (regardless of accompanying audio) did not exert a significant main effect on the number of notifications called out ($p = .7$). However, audio notifications played a crucial role ($F(1,100) = 22.01, p < .001$). Participants called out significantly more visual notifications when accompanied by audio notifications ($M = 6.87, SD = 4.18$) compared to conditions with only visual notifications ($M = 4.1, SD = 2.71$).

These findings suggest that while the presence of visual notifications alone might not guarantee user attention, strategically combining them with audio notifications can significantly improve user awareness of these notifications within the AR environment. This highlights the potential benefits of multimodal notification design for enhancing user experience in AR tasks requiring divided attention.

4.5.3 Visual Notifications Called Reaction Time

The analysis focused on the time participants took to verbally call out the confirmation number displayed in a visual notification. The average reaction time across the different notification design conditions is illustrated in Figure 4.9. The x-axis represents the different conditions (2 conditions of visual notifications and two conditions of audio notifications). The y-axis represents the number of seconds the participants took to shout out the number on the notification. Error bars represent the standard error of the mean. Detailed information on individual means, standard deviations, and standard errors can be found in Table 4.5 (to be inserted later in the document).

The analysis revealed no statistically significant interaction effect between visual and audio notifications ($F(1,68.08) = 3.28, p = .07$). The analysis of simple main effects showed that neither the presence of visual notifications alone ($p = .36$) nor the presence of audio notifications ($p = .34$) exerted a significant main effect on reaction time.

In simpler terms, these findings suggest that, within the context of this experiment, neither the presence or absence of visual notifications nor the presence or absence of audio notifications significantly influenced how quickly participants reacted to visual notifications.

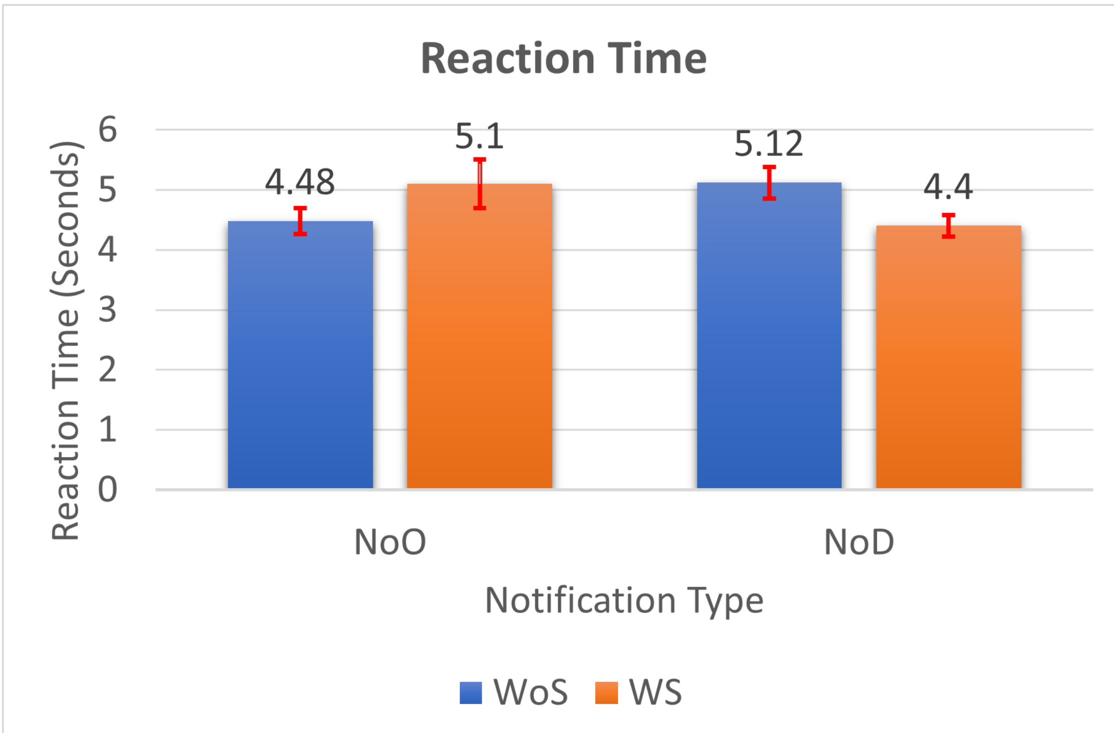

Figure 4.9: Average reaction time to notice the notification in different notification conditions in seconds

Table 4.5: Average reaction times to notice the notification in different notification conditions in seconds

Notification Techniques	Mean	STD	SE
NoO WoS	4.48	1.1	0.21
NoO WS	5.1	2.07	0.41
NoD WoS	5.12	1.34	0.26
NoD WS	4.4	0.93	0.18

4.5.4 User Preference and Usability

The system's usability was evaluated using the well-established Post-Study System Usability Scale (PSSUQ) [71]. Participants completed the PSSUQ after experiencing all conditions within the AR environment. The PSSUQ consists of 18 items rated on a 7-point Likert scale, ranging from 1 (strongly agree) to 7 (strongly disagree). In addition to the standard PSSUQ items, we included a custom question to gauge user preference for notification types.

Participants rated the system favorably on the PSSUQ, with a mean score of 1.23 (SD = 0.44), indicating generally positive perceptions of usability. Next, I analyzed individual PSSUQ items to assess agreement levels. Considering a mean score below 1.75 (25%) as indicative of strong agreement, most statements received strong agreement from participants. However, items regarding task completion time (item 4) ($M = 1.85, SD = 1.38$) and error messages (item 8) ($M = 2.38, SD = 1.88$) showed mixed agreement scores. The raw scores are listed in Section A.1.

To understand user preference for notification types, we included a question asking participants to choose their preferred option from six conditions. The results revealed a clear preference for "Notification on Object With Sound" (50%). This suggests that users found this notification type to be most helpful in the AR customer service scenario. Other user preferences included notifications on the object without sound (7.7%), notifications on the dock with sound (19.2%), and notifications with sound only (19.2%). Minimal interest was shown for notifications on the dock without sound (0%) and no notifications without sound (1%).

4.6 Discussion

This study investigated how user performance and preference are influenced by the design of visual and audio notification systems within an AR environment. We evaluated two visual notification placements (fixed dock vs. above relevant objects) and the presence or absence of accompanying audio notifications. The results demonstrate that the combination of visual notifications positioned directly above relevant objects and synchronized audio notifications outperformed other conditions. Notably, this combination achieved better noticeability and emerged as the preferred notification mode among participants. These findings highlight the importance of strategic design choices when using visual and audio notifications in AR interfaces. The findings align with the work of Plabst et al. [51], who also found that placing AR notifications in the world anchored to the task context (in their case, on the object) led to the quickest reaction times, even though their study did not involve audio cues. This reinforces the notion that strategically positioning visual notifications can significantly enhance user performance in AR tasks.

The study examined the number of customers served (out of six) as a performance metric. Interestingly, neither visual nor audio cues alone had a statistically significant effect on user performance. However, the combination of both modalities yielded promising results. We posit that audio notifications effectively capture user attention, while visual notifications positioned above objects aid users in locating and engaging with their primary tasks more swiftly.

The post hoc analysis revealed a noteworthy trend: the presence of audio notifications significantly enhanced the benefit of visual cues positioned above objects compared to visual notifications on a dock or no visual notification at all. Conversely, the absence of audio notifications did not reveal significant differences between the various visual cue configurations. This suggests that when solely relying on visual cues, participants relied less on visual notifications as a primary means of task assistance.

These observations align with the findings from our noticeability assessment. The results showed that the presence of audio notifications significantly improves the overall noticeability of notifications. In a 3D environment characterized by distributed information, there's a high probability of users missing notifications located outside their immediate FoV. Therefore, incorporating audio notifications, where feasible, offers an impactful way to augment notification noticeability within AR interfaces. Our finding that the combination of visual and audio notifications outperformed either modality alone is corroborated by Lazaro et al. [53]. Their study demonstrated that audio notifications outperform visual notifications, but combining both leads to significantly better results. This convergence of findings highlights the potential benefits of multimodal notification design for enhancing user performance in AR tasks requiring divided attention.

Investigating user preferences yielded results consistent with the performance findings. Approximately 50% of participants indicated a preference for "Notifications on objects with sound." The PSSUQ survey results revealed that participants generally found the system to be usable and satisfactory. The slightly lower scores on task completion time might be attributed to the intentional design of time-sensitive tasks aimed at encouraging multi-tasking. The low average score for the item discussing error messages could be due to the limited nature of the error message itself, which only displayed a brief text notification when presenting an incorrect food item. Providing more specific and timely feedback within the task flow might improve user experience and potentially lead to higher scores on this item.

While further research is necessary to fully understand the optimal design of AR HMD notifications, our study provides valuable insights into the effectiveness of strategic visual placement and multimodal notification design. We found that placing visual notifications directly on task-relevant objects enhanced user performance and noticeability, with the combination of visual and audio notifications being the most preferred by participants. These findings build upon existing research on notification design by demonstrating the benefits of a multimodal approach within an AR context.

4.6.1 Design Implications

The study found that while neither visual nor audio cues alone significantly affected user performance, the combination of both led to promising results. Developers should thus consider integrating both modalities into their AR applications to enhance noticeability and user engagement.

The study also highlighted the importance of the effective placement of visual notifications, particularly above relevant objects, to improve user task performance. This suggests that the developers should consider the placement of visual cues to improve users' ability to find and interact with objects in an AR environment. The sound cues also proved effective in capturing users' attention, and developers should explore incorporating sound cues where feasible, especially in 3D environments with distributed information.

4.6.2 Limitations

While our results favored the combination of visual and audio cues, it's important to underscore the significant role played by sound cues in guiding users' attention toward visual notifications. The design of a visual cue with attention-directing capabilities may yield intriguing outcomes.

Another limitation of our study relates to the absence of ambient audio noise in the AR environment, which could potentially diminish the effectiveness of sound cues. Despite the presence of visual clutter and noise in the AR setting, the absence of concurrent auditory noise represents a distinct environmental constraint worth acknowledging. Lastly, users had to call out notifications verbally when they saw them, for the experimenter to log the notice. This could introduce potential noise to the reaction times.

4.7 Conclusion and Future Work

This study investigated the effectiveness of visual and audio notifications in user-friendly AR interfaces. An empirical evaluation explored user performance and preference for different notification modalities within the immersive environment of ARtisan Bistro, a collaborative AR platform. Participants engaged in tasks related to cooking and customer service while we manipulated the placement of visual notifications (fixed vs. above objects) and the presence of localized audio notifications.

The results revealed a clear trend: the combination of visual notifications positioned directly on relevant objects with synchronized audio cues emerged as the most effective configuration within AR HMDs. This approach not only enhanced user performance but also captured user attention most effectively compared to alternative notification designs.

These findings underscore the importance of strategic design choices regarding visual and audio notifications in AR environments. Our research offers valuable insights for engineers, developers, and others who create 3D interfaces for AR applications. Ultimately, this fosters user-centered innovation for smoother everyday interaction with mixed reality technology.

Future research directions include exploring visual cueing techniques that notify users and guide their attention toward the notifications themselves. Another direction of research can include transitioning to a setting that blends physical and virtual elements, where notifications remain virtual while interacting with tangible objects. Investigating alternative visual notification techniques, such as color-coding and positioning notifications within the user's FoV, is another potential area for further investigation. These enhancements could potentially reduce cognitive load and improve user efficiency in managing multiple notifications within an AR environment.

Chapter 5

Beyond the Beep: Comparing Visual-Only and Multimodal Notification Design for Improved Noticeability in Augmented Reality

This chapter addresses a crucial aspect of AR user experience: designing and optimizing information delivery through notifications within AR environments. In a noise-filled environment, visual notifications may serve as the primary communication channel between the AR system and the user. This chapter investigates whether visual-only notification designs can be optimized to minimize disruption, maximize user performance, and assist users as effectively as multimodal notifications. Additionally, it explores whether visual-only notifications can become a preferred design choice.

This Chapter will be adapted and submitted to the 32nd IEEE Conference on Virtual Reality And 3d User Interfaces 2025.

5.1 Introduction

Augmented Reality (AR) is rapidly evolving as a critical technology in various domains, offering enhanced user experiences by overlaying digital information onto the real world. However, effective information delivery within AR environments presents significant challenges, particularly regarding notification design. Notifications must balance being informative and attention-grabbing without causing undue distraction or cognitive overload. This balance is especially important in environments where users must maintain high-performance levels and situational awareness, such as in customer service or industrial applications.

Notifications in AR can be delivered through various modalities, like visual, audio, or haptic. Prior research suggests that multimodal notifications, which integrate both visual and auditory cues, can improve user focus and reduce task load [13, 26]. However, the effectiveness of visual-only notifications, especially in environments where auditory cues may be impractical due to ambient noise or other constraints, remains underexplored. This study aims to fill this gap by investigating the effectiveness of different visual-only notification designs compared to combined visual-audio notifications in AR environments.

The primary objective of this research was to determine whether user performance could be maintained with more intrusive visual-only notifications compared to multimodal notifications (visual + audio). The secondary objective was to explore whether specific visual-only notification designs can capture user attention as effectively as notifications that combine visual and auditory modalities.

This study employed a simulated AR environment called ARtisan Bistro [54] to evaluate user performance, attention capture, and user experience across various notification designs: Notifications on Object With Sound (NoO WS), Notification on Viewport (NoV), Notification On Object With Arrow (NoO WA), and a Control condition with No Notifications.

The findings indicate that user performance remained consistent across all notification types, suggesting that visual-only notifications do not hinder performance compared to multimodal designs. Attention capture was significantly higher with combined audio and visual cues (NoO WS) compared to visual-only notifications guided by an arrow (NoO WA). FoV fixed visual notifications (NoV) showed similar attention rates to both NoO WS and NoO WA, suggesting their potential effectiveness in noisy environments.

While visual-only notifications are effective in maintaining user performance, combined visual-audio notifications are more efficient in capturing user attention and are preferred by users. These findings highlight the need for further research to explore more complex AR tasks, optimize visual notification designs, and understand the impact of ambient noise on visual-only notifications.

5.2 Notification Design

This section dives into the design of the notification system implemented in our study. The best-performing and the most preferred notification design (Notification on Object With Sound) from the last study was chosen as the baseline for comparison. I then introduced and evaluated two new visual-only notifications: Notification on Viewport and Notification on Object with Arrow. For more details about Notification on Objects With Sound, refer to Subsections 4.3.1 and 4.3.2.

5.2.1 Notification on Viewport

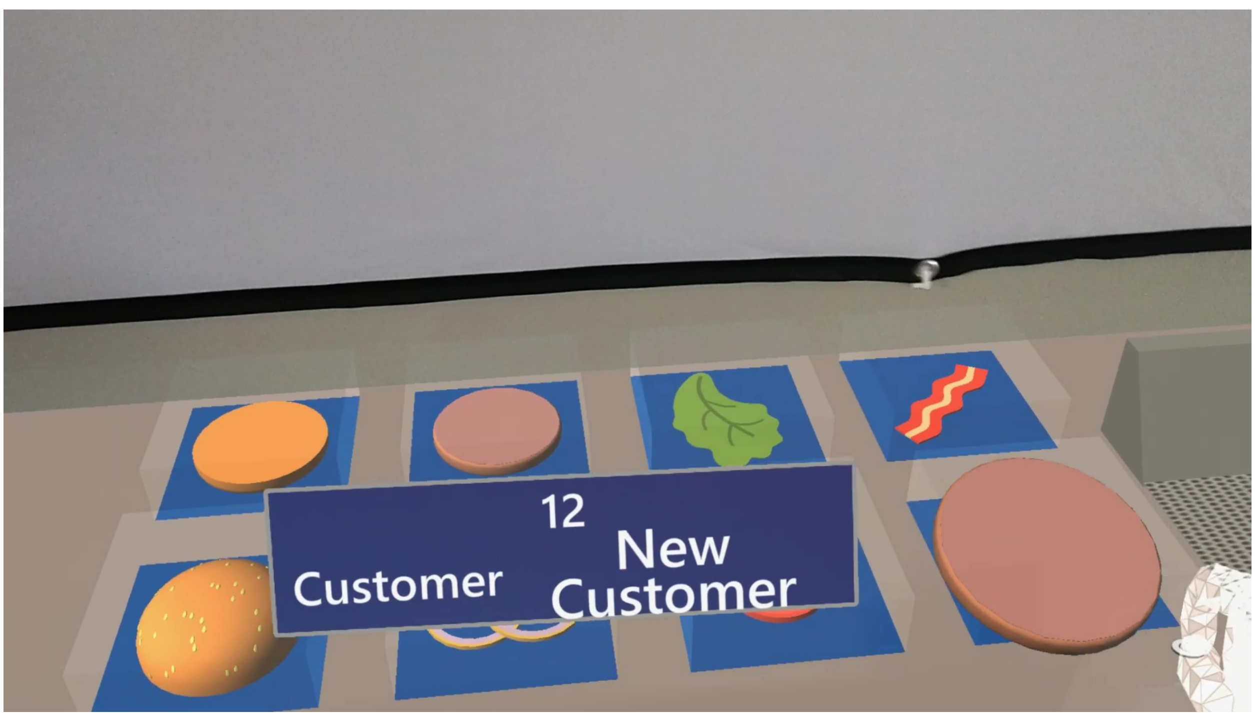

Figure 5.1: Notification on Viewport

Notification on Viewport displays notifications attached to a fixed position in the FoV. The notification's movement is synchronized with the movement of the HMD. Positioning the notification attached to the display will ensure that users are more likely to notice and read it. In cases of urgent notifications, it is crucial to promptly inform the user of the conveyed information. Additionally, by synchronizing the movement of notifications with the HMD, users can easily track and interact with notifications as they navigate their environment.

However, it is essential to consider potential challenges such as obstructing primary tasks, notification overload, and missed notifications due to limited FoV. The position of the notification on the display is an important factor to consider. [72] investigated the influence of AR notification design on user experience while walking. Their findings suggest that optimal design depends on notification type, with bottom placement preferable for high-priority notifications and right placement better for low-priority notifications. Similar results were observed by Plabst et al. [51] who stated, "bottom-center position should be used over top-right placement." Based on the previous studies, the bottom-center position on the FOV was selected (Figure 5.1).

5.2.2 Notification on Object with Arrow

Notifications on Object with Arrow improves on the perks presented by Notifications on Object by adding a guiding object. This guiding object, in this case, an arrow, can lead the user to the notification when it's out of view. To minimize obstruction of the FoV, the arrows were positioned on the periphery and aligned with the direction they were pointing (Figure 5.2).

Warden et al. examined how different cue designs and placement affect visual search tasks in AR [66]. Their study compared three cue types (arrow, icon image, minimap) and two placement locations (center, downward) within an AR HMD. They found that arrows provided the greatest search benefit. The addition of more information on the display can lead to obstructing primary tasks, information overload, and distract attention away from primary tasks or important environmental cues.

Figure 5.2: Notification on Object with Arrow

5.3 Methodology

5.3.1 Research Questions

By posing these research questions (RQs), I aim to gain valuable insights that can improve user experience (UX) within AR environments.:

RQ1: Can user performance remain the same when aided with more intrusive visual-only notifications compared to notifications with added sound?

Prior research suggests that visual and auditory notifications can improve user focus and reduce task load [13, 26]. This research question explores the effectiveness of visual-only notifications in supporting user performance and satisfaction within an AR environment. Specifically, I wanted to explore how these notifications, presented only as a visual modality, can influence users' ability to manage tasks and interruptions when serving customers.

RQ2: Among the two pure visual notification designs (NoV, NoO WA), are either of them as successful in capturing users' attention as notifications with both visual and audio modality?

Imagine a bustling AR environment where background noise might make it challenging to hear notification sounds. In such scenarios, visual notifications become the primary way to grab users' attention. This research question seeks to determine if visual-only notification types (NoV, NoO WA) are as noticeable and effective as notifications that combine visual cues with sounds (NoO WS).

RQ3: How do the different notification design techniques influence user experience? This user experience includes perception of ease of use, clarity, and overall satisfaction. Do users develop preferences for specific notification types?

These preferences could influence how readily users adopt and feel comfortable using notifications within AR environments in the long run. For example, some users might find visual-only notifications less intrusive, while others prefer the combined audio-visual approach for enhanced clarity. The perceived effectiveness and user acceptance of notification systems have been shown to be significantly impacted by user preferences in previous studies [34].

By addressing these questions, I hope to better understand the trade-offs between pure visual notifications and notifications that combine visual and auditory cues. This knowledge will inform the design of AR notification systems that optimize user experience. The ideal system should support task management effectively while minimizing distractions. It should also promote user comfort and acceptance by considering preferences for notification modality and intrusiveness.

5.3.2 Experiment Design

This study utilized a within-subjects design. The independent variable was the type of notification used, resulting in four experimental conditions:

(1) Notification on Object With Sound (NoO WS) - Participants received both visual notifications on the object and an accompanying audio notification.

(2) Notification on Viewport (NoV) - Visual notifications were displayed in a fixed location at the bottom center of the participant's FoV

(3) Notification on Object With Arrow (NoO WS) - Participants received visual notifications displayed directly on the virtual object requiring attention. An arrow was paired with each notification, guiding the participants to the respective notification.

(4) No Notifications (Control) - This baseline condition presented no visual notifications. Participants relied solely on their ability to keep track of changes within the environment to manage tasks.

Figure 5.3 illustrates the overall flow of the experiment. Each participant completed all four conditions in a randomized order, ensuring a balanced evaluation across notification designs. The experiment followed these steps:

1. **Arrival and Introduction:** Participants arrived at the research lab, provided informed consent, and received a detailed explanation of the study.

2. **Familiarization:** Participants watched a short video introducing the AR environment and the different notification types they would encounter throughout the experiment.

3. **Training Level:** Participants completed a practice level using a Hololens 2 headset. This level involved making food items for three virtual customers without a time limit, allowing them to familiarize themselves with the controls and virtual kitchen environment. During this level, participants were allowed to ask questions and guided through the steps to prepare food if needed.

4. **Experimental Levels:** Participants proceeded through the four experimental conditions after completing the training level. The order of conditions was randomized for each participant using a chi-square model.

   (a) At the beginning of each level, participants received instructions regarding the specific notification design they would experience (e.g., Notification on Object with Sound).

   (b) Each level presented a simulated customer service scenario where virtual customers requested food items within a set time limit. Participants used virtual objects to prepare the food while receiving notifications about cooking status and new customers.

   (c) The chosen notification sound was a subtle bubble-popping sound to minimize distraction and maintain neutrality. Visual notifications remained active for 7 seconds in NoO WS, NoV, and NoO WA conditions.

   (d) Participants acknowledged notifications by speaking out the word notification (to verify that they have indeed observed the notification).

   (e) After each condition, participants completed two surveys (System Usability Scale [73] and NASA-TLX [64]).

5. **Post-Interview:** Upon completing all four levels, participants answered post-experiment interview questions to convey their subjective experience with the different notification designs.

The experiment recorded the following data to assess user performance and experience:

This comprehensive data collection strategy allowed for a detailed analysis of user behavior, performance metrics, and subjective experience across the different notification design conditions within the AR customer service scenario. The data was later preprocessed to prepare it for further statistical analysis.

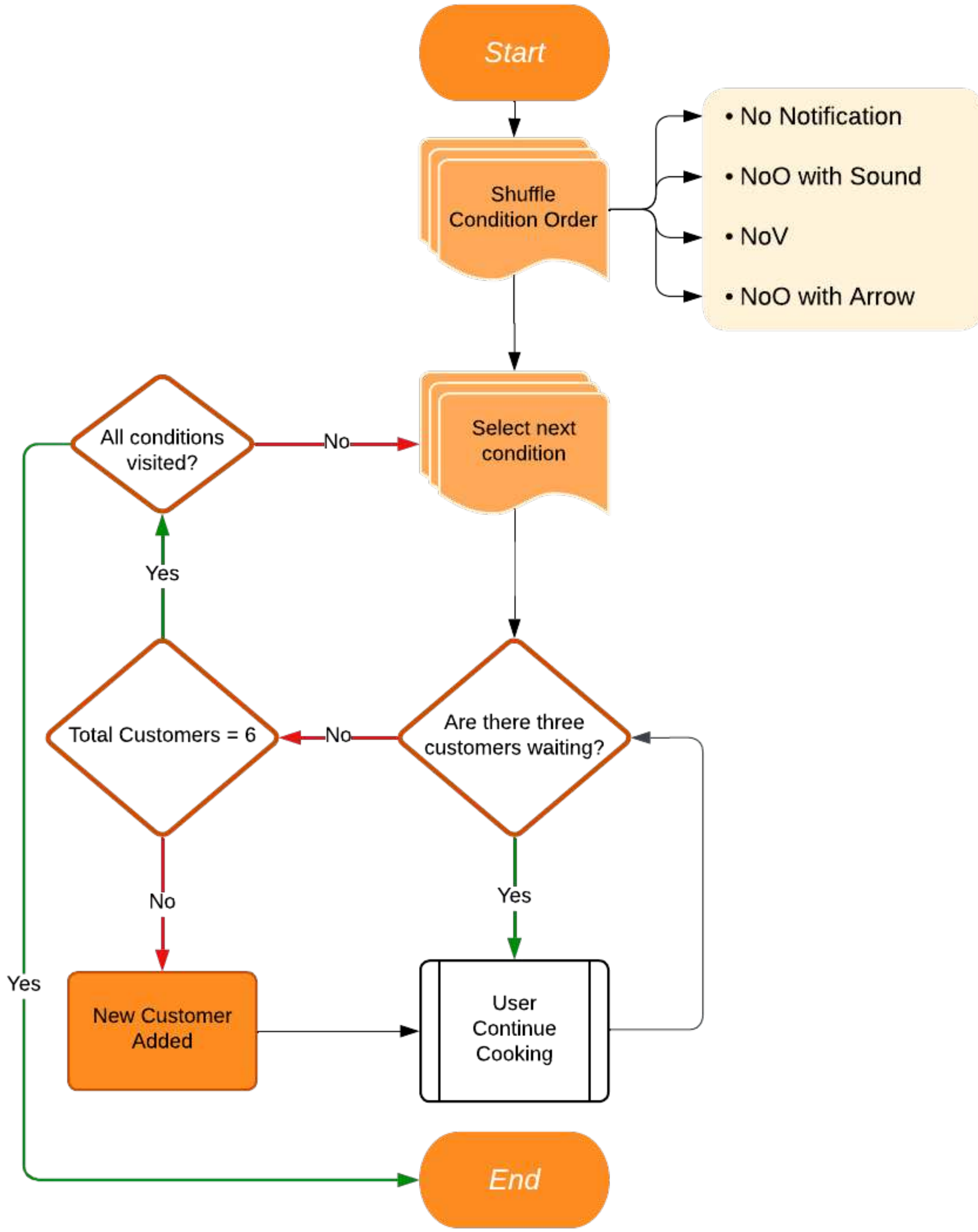

Figure 5.3: Experiment Flowchart

Table 5.1: Data measures recorded

User Actions		
Experiment start	Experiment end	New notification
New ingredient spawn	Ingredient deleted	Notification observed
Food presented to a customer	Pizza put in the oven	New Customer
Cooking Status Change	Coffee cup added	Customer leaves
Performance Metrics		
Customer Served	Customer Departure	Customer Waiting Time
Notification Observed Frequency	Reaction Time	Post-Survey Responses

5.3.3 Participants

The experiment involved 24 participants. All participants identified as either male or female, with a breakdown of 13 male and 11 female participants. Their age ranged from 19 to 56 (M = 27.88, STD = 8.99). This age range represents a mix of young adults and individuals in their mid-fifties. Participants were recruited from a combination of sources: students and staff from Colorado State University and individuals not affiliated with the university. The majority of participants were either students or faculty.

Prior experience with AR and VR technology was assessed to gauge the participants' familiarity with immersive environments. Over 90% of the participants (91.67%) reported having used at least one AR or VR device in the past. This indicates a relatively high level of comfort and a basic understanding of interacting within these environments. To understand the participants' potential familiarity with the task context (restaurant service), I inquired about previous work experience in a restaurant setting (e.g., cook, waiter). Approximately one-third of the participants (33.3%) reported having some form of restaurant work experience. While not a prerequisite for successful task completion, this information provides context for interpreting any potential influence of prior experience on user performance. As a token of appreciation for their time and participation, each participant received a compensation of $20. This study was approved by the university's internal review board.

5.4 Results

5.4.1 Performance

The analysis focused on the number of customers served successfully within the time limit as a measure of user performance. Figure 5.4 depicts the average number of customers served across the different notification design conditions. The x-axis represents the various conditions (4 notification types). The y-axis represents how many customers were served on average in each condition (maximum is 6). Error bars represent the standard error of the mean. Detailed information on individual means, standard deviations, and standard errors can be found in Table 5.2.

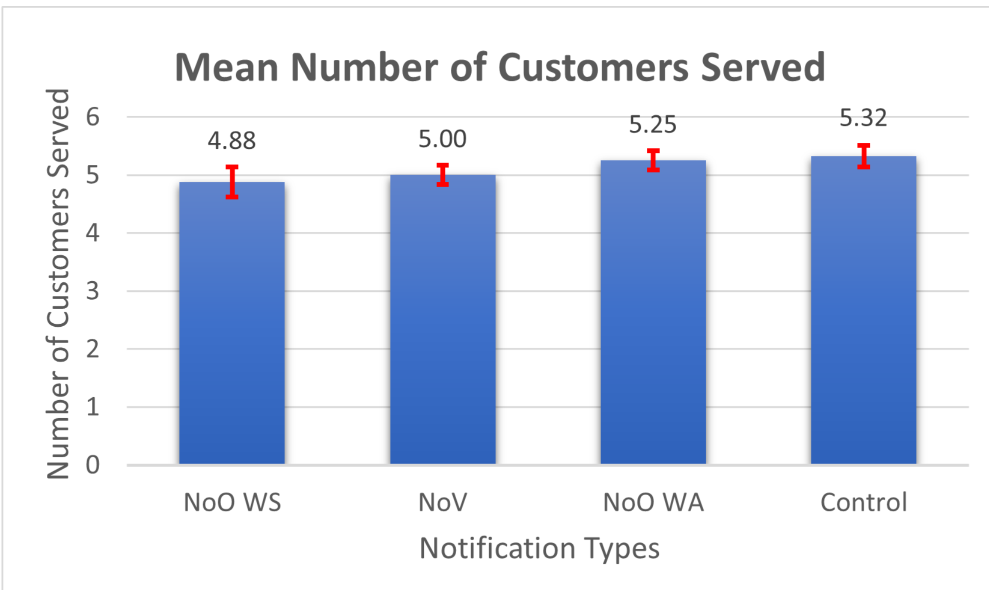

Figure 5.4: Performance of participants in different notification conditions based on how many customers they served.

A Shapiro-Wilk test [74] was performed and showed that the distribution of the data departed significantly from normality ($W = 0.81, p < .001$). Levene's test [75] was conducted to assess the homogeneity of variances. The results showed variances were unequal across groups ($F(3,92) = 3.17, p < .05$). Based on this outcome, a non-parametric test called the Kruskal-Wallis H test [76] was used. It showed no significant difference

Table 5.2: Mean performance of participants in different notification conditions

Notification Techniques	Mean	STD	SE
NoO WS	4.88	1.27	0.26
NoV	5.0	0.82	0.17
NoO WA	5.25	0.83	0.17
Control	5.32	0.87	0.18

in pain score between the different drug treatments, $\chi^2(3) = 1.48, p = 0.69$. This suggests that the type of notification participants received did not significantly impact their ability to complete customer service tasks within the simulated environment.

We further investigated whether prior restaurant experience influenced user performance. A one-way ANOVA was performed to compare the effect of prior work experience in a restaurant on the number of customers served. A Shapiro-Wilk test was performed and did not show evidence of non-normality ($W = 0.92, p = 0.388$). Levene's test was conducted to assess the homogeneity of variances. The results confirmed equal variances across groups ($F(1,22) = 0.74, p = 0.4$). The ANOVA revealed no statistically significant difference in the number of customers served between the two groups ($F(1,22) = 0.174, p = 0.68$). This suggests that any prior experience in the culinary field did not significantly impact the user's ability to complete customer service tasks within the simulated environment.

5.4.2 Notifications Observed

The analysis focused on the number of times participants actively acknowledged notifications by verbally calling out the word "notification" every time they observed a new notification (this can be visual or auditory observation). The average percent of notifications called out across the different notification design conditions is illustrated in Figure 5.5. The x-axis represents the different conditions (notification types). The y-axis represents the percentage of notifications that were observed on average in each condition.

Error bars represent the standard error of the mean. Detailed information on individual means, standard deviations, and standard errors can be found in Table 5.3

A Shapiro-Wilk test was performed and did not show evidence of non-normality ($W = 0.99, p = 0.695$). Levene's test was conducted to assess the homogeneity of variances. The results confirmed equal variances across groups ($F(2,57) = 2.22, p = 0.118$). A one-way ANOVA was conducted to assess potential differences in notification observation rates across the three experimental conditions (NoO WS, NoV, and NoO WA). The ANOVA results revealed a statistically significant main effect of notification type on the percentage of notifications observed ($F(2,56) = 4.32, p < 0.05$). This suggests that at least one of the notification types differed from the others in terms of how many notifications participants noticed.

Further analysis using Tukey's Honestly Significant Difference (HSD) test for multiple comparisons identified a significant difference in notification observation rates between the NoO WS and NoO WA conditions ($p < 0.05, 95\% \ CI = [1.70, 19.79]$). This indicates that participants observed a significantly higher percentage of notifications when presented with both visual cues on the object and accompanying sound cues (NoO WS) compared to the condition with only visual cues on the object and an arrow (NoO WA).

Interestingly, the ANOVA results did not reveal any statistically significant differences in notification observation rates between the NoO WS and NoV conditions ($p = 0.124$), nor between NoV and NoO WA ($p = 0.567$). This suggests that participants observed a similar percentage of notifications when presented with visual and audio cues on the object (NoO WS) compared to only visual cues displayed at a fixed position in the user's FoV (NoV) and with visual cues on the object with an arrow (NoO WA) compared to visual cues displayed at a fixed position in the user's FoV (NoV).

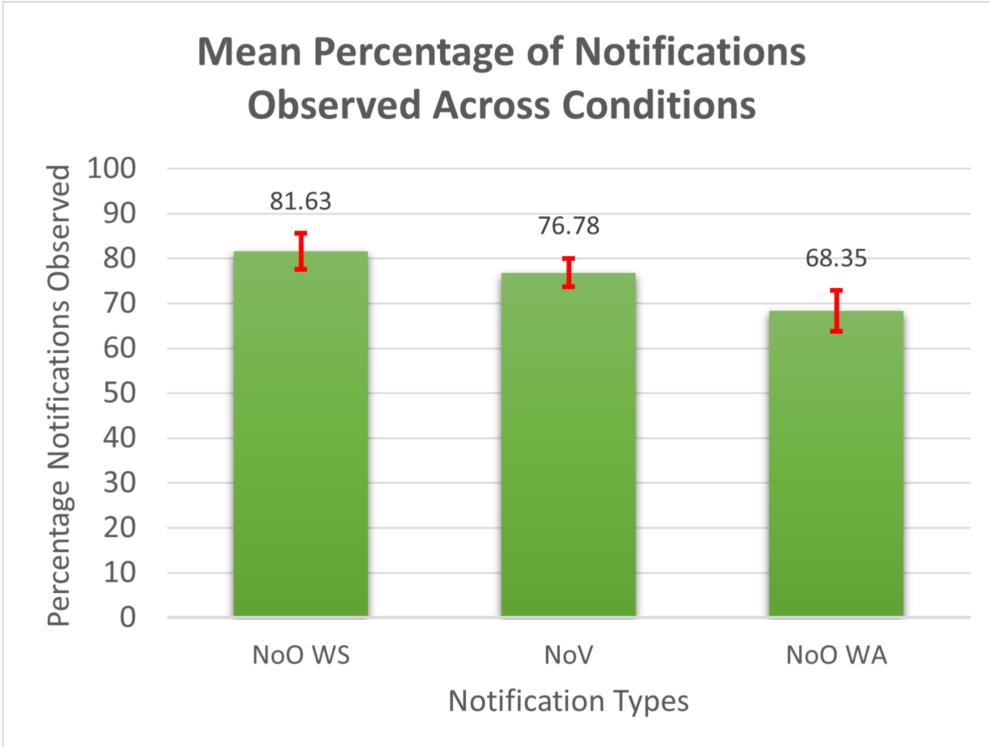

Figure 5.5: Average number of notifications observed in different notification conditions

Table 5.3: Mean number of notifications observed in different notification conditions

Notification Techniques	Mean	STD	SE
NoO WS	81.63%	18.90	4.03
NoV	76.78%	15.03	3.13
NoO WA	68.53%	20.26	4.53

5.4.3 Notifications Observed Reaction Time

The analysis focused on the time participants took to verbally call out the confirmation number, which was displayed in a visual notification. The average reaction time across the different notification design conditions is illustrated in Figure 5.6. The x-axis represents the different conditions (notification types). The y-axis represents the number of seconds the participants took to shout out the word "notification". Error bars represent the standard error of the mean. Detailed information on individual means, standard deviations, and standard errors can be found in Table 5.4.

A Shapiro-Wilk test was performed and did not show evidence of non-normality ($W = 0.98, p = 0.526$). Levene's test was conducted to assess the homogeneity of variances. The results confirmed equal variances across groups ($F(2,57) = 0.75, p = 0.477$). A one-way ANOVA was conducted to assess potential differences in reaction times (time taken to notice a new notification) across the three experimental conditions (NoO WS, NoV, and NoO WA). The ANOVA results revealed a statistically significant main effect of notification type on reaction time for observed notifications ($F(2,58) = 7.13, p < 0.01$). This indicates that at least one of the notification types led to significantly faster or slower reaction times compared to the others.

Further analysis using Tukey's HSD test for multiple comparisons identified significant differences in notification response time between the following conditions:

NoO WS vs NoO WA: Participants responded significantly faster to notifications presented with both visual cues on the object and accompanying sound cues (NoO WS) compared to the condition with only visual cues on the object and an arrow (NoO WA) ($p < 0.01, 95\%\ CI = [-1.195, -0.225]$).

NoV vs NoO WA: Participants also responded significantly faster to visual cues displayed at a fixed position in the user's FoV (NoV) compared to the condition with visual cues on the object and an arrow (NoO WA) ($p < 0.01, 95\%\ CI = [-1.113, -0.133]$).

Interestingly, the ANOVA results did not reveal any statistically significant difference in notification response time between the NoO WS and NoV conditions ($p = 0.896$). This suggests that participants responded with similar speeds to notifications presented with both sound and visual cues on the object (NoO WS) compared to only visual cues displayed at a fixed position in the user's FoV (NoV).

5.4.4 User Preference, Usability, and Workload

The system's (notifications) usability was evaluated using the well-established System Usability Scale (SUS) [73]. Participants completed the 10-item SUS survey after each

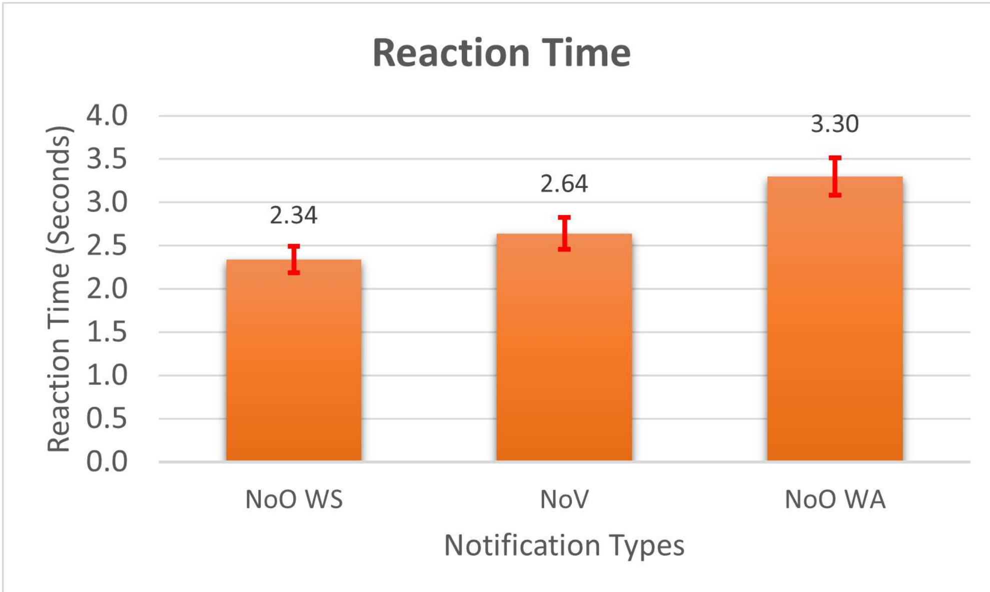

Figure 5.6: Average reaction time to register the notification in different notification conditions in seconds

Table 5.4: Average reaction times to notice the notification in different notification conditions in seconds

Notification Techniques	Mean	STD	SE
NoO WS	2.34	0.70	0.15
NoV	2.64	0.88	0.18
NoO WA	3.30	0.96	0.21

condition, recording their subjective experience. The SUS consists of 10 items rated on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). In addition to the standard SUS items, a custom question was included to gauge user preference orders for notification types.

Participants rated the system favorably on the SUS. The generally accepted rating is $> 80.3 = Excellent$; $68 - 80.3 = Good$; $68 = Okay$; $51 - 68 = Poor$; $< 51 = Awful$ [77]. The results showed variations in perceived usability across conditions. Participants' perceived usability scores for each condition are reported in Table 5.5.

A Shapiro-Wilk test was performed and did not show evidence of non-normality ($W = 0.98, p = 0.172$). Levene's test was conducted to assess the homogeneity of vari-

Table 5.5: Average SUS scores for different notification conditions

Notification Techniques	Mean	STD	SE
NoO WS	80	15.28	3.12
NoV	77.5	14.49	2.96
NoO WA	66.67	18.34	3.74
Control	75	17.14	3.50

ances. The results confirmed equal variances across groups ($F(3{,}92) = 0.24, p = 0.866$). A one-way ANOVA was conducted to assess potential differences in SUS scores across the four experimental conditions (NoO WS, NoV, NoO WA, and Control). The ANOVA results revealed a statistically significant main effect of notification type on the SUS scores ($F(3{,}92) = 2.87, p < 0.05$). This suggests that at least one of the notification types differed from the others in terms of SUS score.

Further analysis using Tukey's HSD test for multiple comparisons identified a significant difference in SUS scores between the NoO WS and NoO WA conditions ($p < 0.05, 95\%\ CI = [0.694, 25.972]$). This indicates that participants experienced better system usability when presented with both visual cues on the object and accompanying sound cues (NoO WS) compared to the condition with only visual cues on the object and an arrow (NoO WA).

Interestingly, the ANOVA results did not reveal any statistically significant differences in SUS scores between the NoO WS and NoV conditions ($p = 0.955$), the NoO WS and Control ($p = 0.729$), the NoV and NoO WA conditions ($p = 0.120$), the NoV and Control ($p = 0.955$), and the NoO WA and Control ($p = 0.317$).

The NASA Task Load Index (TLX) was used to assess participants' perceived workload while interacting with the system (notifications) under different conditions. After completing each condition, participants provided ratings on six subscales. Each subscale ranged from 0.5 to 10 with intervals of 0.5. The overall workload was calculated by using Raw NASA-TLX [64].

Participants' perceived workload scores for each condition are reported in Table 5.6.

Table 5.6: Average workload scores for different notification conditions

Notification Techniques	Mean	STD	SE
NoO WS	53.02	18.53	3.78
NoV	52.29	16.82	3.43
NoO WA	54.97	17.69	3.61
Control	53.19	18.80	3.84

A Shapiro-Wilk test was performed and did not show evidence of non-normality ($W = 0.98, p = 0.098$). Levene's test was conducted to assess the homogeneity of variances. The results confirmed equal variances across groups ($F(3,92) = 0.056, p = 0.98$). A one-way ANOVA was conducted to assess potential differences in NASA TLX scores across the four experimental conditions (NoO WS, NoV, NoO WA, and Control). The ANOVA results did not reveal a statistically significant main effect of notification type on the NASA TLX scores ($F(3,92) = 0.092, p = 0.97$). This suggests that participants' perceived workload, as measured by the NASA TLX, did not differ significantly between the four notification conditions. In other words, regardless of the notification type presented, participants reported similar workload levels while performing the tasks within the AR environment.

To gain insights into user preferences for notification design, participants were asked to rank the notification conditions from "Best" to "Worst". The results revealed a clear order of preference. Participants overwhelmingly favored the NoO WS condition, followed by NoV, Control, and lastly, NoO WA.

5.5 Discussion

5.5.1 Performance

Building upon the established value of notification cues for user experience, RQ1 delved deeper into the effectiveness of visual-only notifications for user performance in AR environments. While past research suggests the benefits of both visual and auditory notifications for focus and task load reduction [13, 26, 61], this study explored if visual-only notifications could maintain similar performance levels. Maintaining or even surpassing the performance with vision-only notifications is important for environments where sound modality may be impractical.

The analysis of the number of customers served revealed no significant difference between the three notification conditions (NoO WS, NoV, NoO WA). This suggests that user performance in serving customers remained comparable regardless of whether they received visual-only notifications or notifications with sound within the simulated restaurant environment. These findings indicate that visual-only notifications might not necessarily hinder user performance compared to multimodal notification designs.

However, it's important to acknowledge that these results also indicate no significant difference in performance between conditions where participants received notifications and those where they did not (Control condition). This lack of distinction between notification and control groups is unexpected and warrants further investigation. Notably, these findings diverge from the results obtained in the previous studies [13, 26].

One potential explanation for these observations is the cognitive demand for customer service tasks within the AR environment. The tasks may not have been complex enough for visual-only notifications to be a significant disadvantage compared to notifications with sound. This could also explain the lack of difference between participants who received notifications and those who did not. More complex tasks might benefit more from the combined support of visual and auditory cues. While visual-only notifications did not hinder performance in this specific scenario, further research is needed to understand their effectiveness in a broader range of AR tasks and environments with varying levels of cognitive demand.

Future studies could explore the impact of notification type on user performance in more complex AR tasks that require a higher level of cognitive processing.

5.5.2 Notifications Observed

RQ2 focused on the effectiveness of visual-only notifications (NoV, NoO WA) in capturing user attention, particularly in scenarios with potential ambient noise. The findings highlight the importance of considering the specific design of visual notifications for optimal attention capture.

The analysis of notification observation rates revealed that participants acknowledged a significantly higher percentage of notifications when presented with the combined audio and visual cues (NoO WS) compared to the condition with only an arrow guiding to the visual notifications (NoO WA). This aligns with previous research suggesting the benefits of multimodal notifications for enhanced user awareness [13, 26].

Interestingly, the observation rates for FoV fixed visual notifications (NoV) were statistically similar to both the combined modality condition (NoO WS) and the visual-only with arrow condition (NoO WA). This suggests that in this specific scenario, the NoV notification type might have been sufficient to capture user attention to a similar degree as NoO WS. With FoV fixed visual notifications (NoV) achieving similar attention rates compared to the combined modality condition (NoO WS) in this specific task, it may prove more advantageous in environments with significant background noise. In such scenarios, NoV notifications could ensure users don't miss important information while minimizing auditory disruption.

However, it's important to consider potential limitations. The tasks in this study might not require as much attention to necessitate a clear advantage for the combined modality condition. More demanding tasks might show a greater difference. Building on these findings, future research could explore how visual notification design can be optimized to improve attention capture in AR environments, like experimenting with visual design elements like size, color, animation, and location within the AR space to maximize their attention-grabbing potential. Another avenue can be to explore the effectiveness of NoV notifications in various noise environments and with different task complexities.

The observation rates for purely visual notifications (NoV) were statistically similar to the combined modality condition (NoO WS). However, the analysis of response times (time taken to verbally confirm a notification) paints a different picture. Participants responded significantly faster to both NoV and NoO WS notifications compared to NoO WA. This suggests that while purely visual notifications might be noticed with similar frequency as combined notifications in this specific task, the presence of an arrow pointing to the object wasn't fast enough to capture the attention away from the primary task.

One reason for slow response time may be that arrows are located at the periphery. If the user was too focused on the primary task, they might have missed the appearance of a new arrow. Another reason might be information overload: The additional visual element (arrow) might have caused a momentary information overload, especially compared to the simpler cues in NoV and NoO WS.

This research question highlights a trade-off between notification design and user experience. While visual-only notifications can be effective in some situations, adding auditory cues can enhance attention capture. Future research can further refine visual notification design for optimal user experience in diverse AR contexts.

5.5.3 User Preference, Usability, and Workload

The effects of various notification design strategies on user experience were assessed using metrics for usability, workload, and user preferences to get answers for RQ3. These metrics provided crucial new information about how participants saw and interacted with the various notification situations.

Participants rated the overall system favorably based on the SUS scores (except NoO WA). However, these scores varied across notification conditions. The highest SUS score was observed with the NoO WS condition. This finding aligns with the user-reported preference ranking, where NoO WS was the most favored condition.

A one-way ANOVA confirmed a statistically significant main effect of notification type on SUS scores. Further analysis revealed that the NoO WS condition led to significantly better-perceived usability compared to the NoO WA condition. The SUS scores for the NoV and Control conditions were not statistically different from the NoO WS condition. This might indicate that while the combined visual and auditory approach offered the most optimal usability, visual-only notifications might be a viable alternative. No other pair of conditions had a statistically significant difference. This could be because the SUS scores are not sensitive enough to detect subtle differences in usability, or a learning curve could have masked initial usability differences.

The NASA TLX scores did not reveal any statistically significant differences between the notification conditions, suggesting that participants' perceived workload remained comparable regardless of the notification type received. This could indicate that the customer service tasks in this study were not mentally and physically demanding enough to elicit significant workload variations between different notification designs.

The user-reported preference rankings mirrored the usability findings. Participants overwhelmingly favored the NoO WS condition, followed by NoV, Control, and NoO WA. This alignment suggests that users found the most usable notification design (combined visual and auditory) to be the most preferable. These findings highlight the potential benefits of considering both visual and auditory cues when designing notification systems for AR environments, particularly for tasks where high usability and user satisfaction are crucial.

While both usability and user preference favored the combined visual and auditory cues (NoO WS), even simpler notification designs (NoV and Control) achieved comparable task completion. Perceived workload remained similar across conditions.

5.5.4 Interviews

Participants were interviewed in follow-up sessions to better understand the user rationale and specific design aspects impacting user experience. A deductive method for analyzing qualitative interviews was used to structure and analyze the interview.

The interviews corroborated the survey results. Participants overwhelmingly favored the NoO WS condition due to its perceived clarity and efficiency. They explained that the combination of visual notification on the object and the accompanying sound subtly drew their attention away from the task at hand, minimizing the distraction from the primary task. Participant 8 (P8) stated, *"I'll be focusing on some object, but when it is a sound, I will sense that - Okay, there's a new notification."* P16 echoed this sentiment, saying, *"It let me know what's going on behind the scenes without intruding on what I was doing at the moment."*

In contrast, participants found the NoO WA condition to be the least favorable. Several users reported that the arrow pointing to the object was distracting and blocked their view momentarily, hindering their ability to see important details on the served object itself. This aligns with the finding that NoO WA had a statistically significant slower response time compared to NoV. P4 commented, *"The arrows felt a bit overwhelming. It felt like it was telling you to do a lot of stuff."* P12 offered a similar perspective, saying, *"I think they (arrows) were too big, but I think the main thing is when they piled on top of one another, I'm like, I don't even know where you are pointing anymore."*

The interviews revealed a divide among participants regarding the NoV visual notifications and the arrows in the NoO WA condition. Some participants were able to completely ignore or work around these notifications. P22 said, *"I will just bend down so that the notification went down, and then I would do the task. And it went away after 7 seconds."* However, others found themselves easily distracted by the new information appearing on the screen. P17 stated, *"It's (NoV notification) obstructing me from doing other things."*

Participants also offered valuable suggestions for improving the AR notification system. They indicated that notifications could be made more dynamic, such as by incorporating pulsating/ gradual change effects (P2, P16) or different color/metaphorical sounds to enhance their ability to capture attention (P6, P17, P18, P26), especially in visually cluttered environments. Additionally, participants expressed a desire for customization options (P8, P19), including the ability to adjust notification frequency (P2), select which stations trigger notifications (P15), and set notification priorities based on their individual preferences and specific work contexts (P14).

These results can inform the development of future AR notification systems that are not only usable and efficient but also customizable and adaptable to individual user preferences and environmental factors, ultimately leading to a more satisfying user experience in AR environments.

5.5.5 Limitations

Although this chapter provides insightful information on the design of notifications in augmented reality, it's crucial to take into account the following limitations when interpreting the results. A simplified customer service scenario was used in the study. The conclusion that multimodal notifications outperformed visual-only messages in terms of performance might not apply to more difficult augmented reality tasks. Further research needs to examine the influence of notification modality on user performance in increasingly complex augmented reality settings. The lack of background noise in the augmented reality environment could not accurately represent actual circumstances. Subsequent research efforts might look into the impact of ambient noise levels on the auditory notification experience of users and their overall work performance. Finally, there's a chance that measures of response times might have been affected due to the use of vocal notification confirmation for data collection. Future studies may investigate alternate techniques for gathering data, such as physiological assessments or eye tracking, in order to offer more refined perspectives on how users interact with alerts.

5.6 Conclusion and Future Work

This research sought to understand how notification design influences user experience within AR environments. By investigating the effectiveness of different notification designs (Notifications on Object With Sound, Notification on Viewport, Notification On Object With Arrow, and No Notifications), the study aimed to identify designs that promote efficient information delivery, minimize distraction, and enhance user satisfaction. The research questions aimed to determine whether user performance could be maintained with visual-only notifications (RQ1), compare the attention-capturing capabilities of different visual notification designs (RQ2) and evaluate user experience and preferences across these designs (RQ3).

The findings indicate that user performance remained consistent across all notification types. This suggests that visual-only notifications do not hinder user performance compared to multimodal notifications, aligning with the initial hypothesis. However, the lack of significant difference between notification and no notification conditions highlights the potential need for more cognitively demanding tasks to reveal the performance benefits of notifications.

In terms of attention capture, participants acknowledged a significantly higher percentage of notifications with combined audio and visual cues (NoO WS) compared to visual-only notifications guided by an arrow (NoO WA). Interestingly, fixed field-of-view visual notifications (NoV) showed similar attention rates to both NoO WS and NoO WA, suggesting their potential effectiveness in noisy environments where auditory cues might be less effective. Response times were faster for NoV and NoO WS compared to NoO WA, indicating that simpler visual notifications or those with audio support are more effective in quickly capturing user attention.

As measured by SUS scores, user preference, and usability favored the NoO WS condition, which also received the highest user preference ranking. The NASA TLX scores showed no significant differences in perceived workload across notification types, suggesting that the tasks were not sufficiently demanding to differentiate workload impacts.

The study's findings open several avenues for future research. Future studies could explore more complex AR tasks that demand higher cognitive processing. This will help determine if the benefits of visual-only versus multi-modal notifications become more pronounced under greater cognitive loads. Further research could investigate optimizing visual notification designs to enhance attention capture. Variables such as size, color, animation, and placement within the AR environment should be explored to identify the most effective configurations. Given the potential effectiveness of NoV notifications in noisy environments, additional studies should examine their performance across various noise levels and types. This will provide a deeper understanding of how ambient noise influences the effectiveness of visual-only notifications. And lastly, moving from a purely virtual world with all virtual objects to real-world object interactions is the most logical next step. This can confirm if these results can be transferred to the real world.

Chapter 6

Beyond the Wizard of Oz: Negative Effects of Imperfect Machine Learning to Examine the Impact of Reliability of Augmented Reality Cues on Visual Search Performance [2]

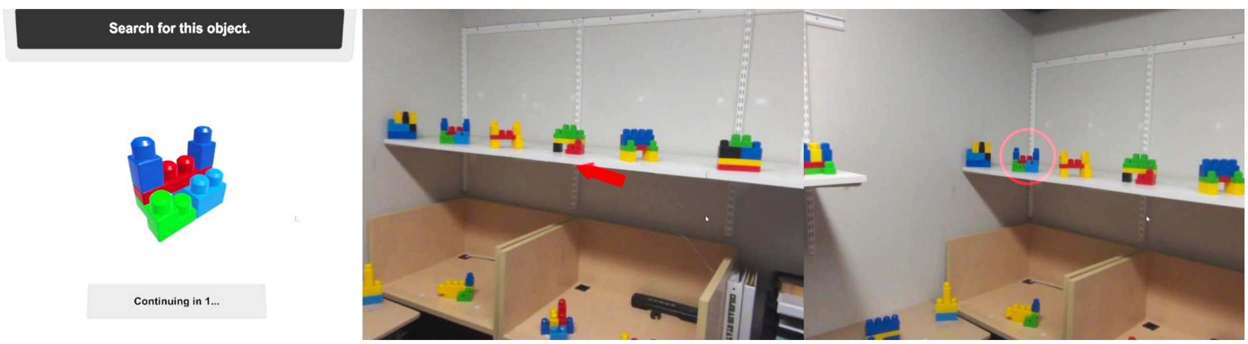

Figure 6.1: Snapshots of one trial (Left panel: Target image is shown to the participants to remember; Center panel: Participants search the 3D environment for the designated object, there may be a cue or no cue depending on the trial condition; Right panel: selecting or rejecting the suggested object

The research presented in Chapters 4 and 5 of this dissertation has established the importance of well-designed notifications in AR environments. By carefully considering factors like placement, modality, and user preferences, we can create notification systems that effectively deliver information while minimizing disruption to the user's primary task. However, a crucial element of the user experience in AR goes beyond simply presenting information – it's about ensuring the information itself is reliable and trustwor-

[2]The content of this Chapter is identical to that presented in: A. Raikwar et al., "Beyond the Wizard of Oz: Using Imperfect Machine Learning to Examine the Impact of Reliability of Augmented Reality Cues on Visual Search Performance," 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Sydney, Australia, 2023, pp. 437-438, doi: 10.1109/ISMAR-Adjunct60411.2023.00093

thy. This is particularly true for AR applications that utilize visual search functionalities to guide users towards specific objects within their environment.

This chapter delves into this critical aspect of AR user experience: the impact of imperfect cues on visual search performance. The research presented in this chapter has already been published in the proceedings of the 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) titled "Beyond the Wizard of Oz: Using Imperfect Machine Learning to Examine the Impact of Reliability of Augmented Reality Cues on Visual Search Performance" by the authors Aditya Raiwkwar, Domenick Mifsud, Christopher D. Wickens, Brendan Kelly, Amelia C. Warden, Benjamin A. Clegg, and Francisco R. Ortega [1].

By including this published research within the dissertation, I aim to emphasize the importance of understanding how imperfect cues, a common challenge in real-world AR applications, can influence user performance and potentially lead to automation bias. By acknowledging the limitations of current AR systems and the potential for imperfect cues, we can develop more robust and reliable user interfaces that foster trust and efficient interaction within AR environments.

6.1 Introduction

Visual search involves actively scanning through perceptual information to locate a specific target. For humans in complex visual environments, the search process will be effortful and serial [21–23]. Examples include a pilot scanning for potential hazards [78], a radiologist searching for a tumor or fracture [79], and a quality inspector searching a product for defects [80, 81]. In such real-world visual search settings, delays and errors in identifying targets can have profound implications. Support for visual search is possible through the use of Augmented Reality (AR) Head-Mounted Displays (HMDs) to cue individuals to targets identified through machine learning solutions. Users can see cues to target objects in relation to their real-world surroundings, which could be necessary for finding objects that might be outside their current field of view, hidden from a certain angle, or behind other things. Cues can provide essential contextual awareness, giving users a sense of where objects are located within their surroundings. In addition, these cues contribute to effective navigation and reduce the cognitive load associated by helping users perceive spatial relationships [82]. By enhancing immersion and increasing efficiency, visual cues serve as invaluable guides for visual search tasks in complex task environments, ensuring that users can quickly and accurately locate objects.

While benefits to both latency and accuracy from AR cues in visual search situations have been shown [83, 84], examining the impact of errors in the cues provided on performance is vital. Even where automated systems might outperform human operators in recognizing and recording the locations of targets, errors within such systems are still possible. In exploring the impact of imperfect cueing, Warden et al. found that when given correct arrow cues to the location of a target, as expected, participants were faster and more accurate in a visual search task [85]. However, incorrect arrow cues also produced higher error rates and longer search times than the uncued cases. This cost to accuracy suggests that participants were over-relying on the visual cues. Such work underlines the value of understanding how dependent users are on signals and how their performance is affected by imperfect signals. The choice of cue to deploy in a particular context ultimately might be determined not only by the magnitude of the benefits they provide when correct but also by the ability of humans to recognize when the cues are incorrect and overcome the tendency towards **automation bias** [86]. Automation Bias is a cognitive bias that people employ during decision-making when they rely on decisions and inferences of automation rather than seek information to make their own judgments and decisions [24].

Previous studies on visual cues, such as Warden et al. [85], used a Wizard-of-Oz (WoZ) approach to set up the detrimental effects of imperfect cues on user search performance. Such work offers valuable initial insight into the impact of incorrect cues on the accuracy of finding the target. However, human-specified errors in such systems introduce an essential limitation to the potential generalization of the findings. Errors can be determined either randomly – with incorrect cues occurring on a specified fraction of trials and directing people to a randomly selected incorrect distractor location; or systematically based on human input – with incorrect cues happening for ''hard'' or ''easy'' targets, and/or cues directing operators to distractors that are ''similar'' or "dissimilar" to the actual target. The issue this creates is that whichever path is chosen to create the imperfect cues, actual errors from machine learning could be very different. Humans are able to correlate different variables even when those variables aren't present in the dataset, while machines can't do that, leading to different errors.

Evidence that the types of representations developed through deep learning differ from human representations (see for example, [87]), that humans and machines make different types of errors [88, 89], and attempts to offer additional insights into the machine learning recommendations through aspects like explainable Artificial Intelligence (AI) to reduce automation bias (e.g., [90]), all point to an array of potential mismatches between humans and automation. Such differences create ways in which imperfect cues generated within a machine learning system could diverge from evidence produced through WoZ scenarios. In addition, a machine learning system working in real-time has the potential to change the cued location as it obtains more information and updates its predictions. The influence of varying locations being cued through a single search, which provides a further source of information about the potential unreliability of the cue being offered, adds an additional justification to explore actual machine learning-driven cues rather than just simulated versions of them.

In this paper, we implemented a real-time object recognition system using a machine learning (ML) algorithm. We analyzed whether the results from visual cues from such a system differ from prior WoZ approaches. For this, we used an arrow as a visual cue to point out the objects. This also helped us gain insight into automation bias in ML.

The expected consequences of automation bias are that people fail to respond to critical information or follow the aid (i.e., cues) when it is wrong. This bias is pervasive in situations where people trust technology to be more accurate and efficient than human judgment [24]. The findings are especially relevant in safety-critical applications where automation bias can lead to errors or overlooked information. For example, in autonomous vehicles, medical diagnosis, or aviation, understanding how users respond to automation cues is crucial for system design and training.

Our research aims to address the convergence of human performance and computer technology in cued search by (1) examining the effects of visual cues for searching in AR HMDs and (2) considering the imperfections of automation, such as vision-based AI, that may make the AI-based cueing imperfectly reliable. The visual cues are displayed on the near-field, e.g., within arm's reach. The far field is where the target objects are located (about 1 meter away from the user). The distinction between near- and far-field refers to the positioning in the real world, not the display on the video pass-through headset. Refer to Figure 6.1 for step-by-step visuals of a single trial.

The contributions of this research paper include:

1. Increasing the external validity given a more realistic scenario compared to a WoZ study for real-time AR HMDs cueing using ML during a visual search task prevalent in this type of research.

2. Demonstrated that the automation bias is high when using a visual cue. Due to the higher reliability of the imperfect visual cue, the magnitude of automation bias was greater using ML than in the WoZ study.

3. Demonstrated that the automation bias in WoZ studies of AR HMD cueing is replicated when using an AR HMD system using ML.

4. Released open-source code and ML training dataset (https://github.com/NuiLab/ML-Imperfect-Cueing), the dataset, and the ML methodology used to train blocks

using virtual objects as opposed to pictures. We hope this will help researchers replicate studies and conduct other studies related to visual search.

6.2 Related Work

The time required to perceive and interpret a particular cue is denoted by its Cue Effectiveness Value (CEV) [91–94]. The CEV considers the time needed to perceive and interpret a cue and orient attention from the cue to the cued target. This is an essential factor to consider when the aim is to reduce the search time using the cue [91–95]. CEV can be predicted and modeled by understanding the cognitive processes involved in visual search. Based on previous studies [66, 85], we have selected an arrow as the cue that benefits visual search performance the most. One crucial factor influencing these processes is whether the cue is exogenous or endogenous [93, 95]. An *endogenous cue* provides information about the target's location but does not appear at that location itself. For example, an image of the target cue that represents the target's appearance but does not directly indicate where attention should be directed is an endogenous cue. In contrast, an *exogenous cue* directs attention to the target's location, such as a flashing highlight near the object in the world (can be presented on an HMD). Exogenous cues have limitations of only being effective if the target is located in the field of view and may distract individuals when they are not actively searching for that specific target. We used an arrow cue to direct attention to the objects to overcome these limitations. We used an arrow because it can point to objects outside the field of view and is not as distracting as a flashing light. These types of cues can be referred to as world-referenced.

The design of AR HMDs can utilize world-referenced cueing and enables the creation of exogenous cues independent of the head's orientation. This cueing method involves placing a cue on an HMD to constantly overlay or point directly at the target, regardless of how the head is turned. This is achieved by presenting the cue on the display in world-referenced coordinates, which are continuously updated on the display as the head moves horizontally or vertically. In the present investigation, we used this type of cue in the form of an AR arrow that consistently indicates the direction of the target.

It is essential to consider the user's prior knowledge about the search target, as detailed information regarding the target may impact visual search tasks. In some cases, a user may only have general knowledge of the target, such as hazards [78] or physical injury [79, 96], which require the user to identify something that may warp or change on a case by case basis. In other cases, the user may have greater knowledge about a target or may have been prompted to search for a specific target. This may be a pedestrian [97], which may differ in some regards but have common, well-known attributes, office tools or supplies [98], or specific virtual geometries [99]; all of which vary to a lesser degree.

The cues used in previous studies, e.g., [66, 99], were always accurate, but in real-world situations, cues are generated through automation inferences based on what the system considers the target. Computer vision using ML algorithms can achieve this by comparing the features of each object in the search field, like humans, with a template of what the true target looks like. However, like many other functions of automation, this inference is not always accurate, as various factors such as lighting conditions, pattern complexity in the search scene, and the reliability of the machine-vision system can affect and degrade its performance [100–105].

The imperfect reliability of automation has significant implications for human performance. Imperfectly reliable automation can impact how much humans rely on the automation's decision of what should be attended to, and researchers have examined these factors in relation to human trust in imperfect automation [101, 102, 106, 107]. Several studies have specifically examined this issue in the context of visual search and target cueing, suggesting that cueing benefits are reduced as reliability is reduced [103–105,108,109]. Although Mifsud et al. [109] have conducted a study closely related to visual cues, only Yeh et al. [110], and Warden et al. [85], to our knowledge, have done so in the specific context of AR HMD target cueing (without the use of ML).

The above research has yielded two critical findings regarding human reliance on imperfect automation. First, a vast amount of research in human-automation interaction has examined the consequences of imperfectly reliable automation (see [100, 102, 111] for a summary). Such research has typically revealed that many participants follow automation's advice or attention guidance infrequently when automation is wrong, referred to as the automation bias [24]. Some research suggests that the automation bias is particularly prevalent because the guidance is offered in highly realistic AR format [109]. In the cueing work from Warden et al. [66], the search guidance cues were 100% reliable. A follow-up study examined the automation bias by conducting an experiment that used WoZ [85]. The study used WoZ methodology to examine the impacts of unreliable cues, which suggested that people exhibit an automation bias in visual search using AR HMDs. Three cues were provided: a mini-map, an icon image, and an arrow pointing to the object. The more realistic cue (i.e., the arrow cue) exhibited the highest automation bias. With this in mind, we designed the current study presented here to look at automation bias using real-time ML (i.e., no WoZ) to understand if there were differences, given the lack of control in the accuracy of the ML using the arrow cue.

Moreover, it has been found that the impact of cue imperfections can vary depending on the type of cue used. Specifically, the more reliable an AR cue is, the more it can assist users when correct. But it can also lead to more significant human errors caused by the automation bias when it is wrong. This prediction is somewhat supported by research on imperfect AR target cueing, which presented an 83% reliable cue pointing to a potential target object [85]. The study found that the AR HMD world-reference ego-centric (i.e., a cue that encodes the location of a target with respect to the viewer) arrow cue was more effective when the cue was correct, but more problematic when the cue was wrong, suggesting that the more immersive and world-referenced cue amplified the automation bias. This tendency is consistent with the notion that a more immersive cue leads to increased *attentional tunneling* [112, 113] ignoring the raw data in the real environment outside the HMD. Supporting this causal assumption, researchers have found that AR HMD cueing reduced the detection of other non-cued but high-priority threats in the scene [110, 112].

As stated above, our work differs from previous work, including recent findings from Warden et al. [66, 85], which used the WoZ method. In this paper, we used an ML algorithm for external validation to investigate the negative effects of imperfect visual cues. Additionally, our research addresses a longstanding inquiry concerning the WoZ method's efficacy in the context of visual cues. Additionally, we showed that ML algorithms produce a higher automation bias than the WoZ method.

6.3 Methodology

6.3.1 Hypotheses

We hypothesized that the interaction of the exogenous type of cue and the impact of cue imperfections would be significant, such that the benefits of cues would be greater for participants in the perfect automation condition compared to those in the imperfect automation condition. Overall, this design allowed us to test multiple hypotheses and explore the effects of automation quality and cues on human visual search.

- H_1 **The addition of cues (perfect/imperfect) produces higher accuracy performance relative to the control condition with no cues.** When a visual stimulus is available on the screen, the participants will pay attention to it, especially when it is generated by computers. This will increase participants' awareness, and they reach higher accuracy [114].

- H_2 **The addition of cues produces lower response times (i.e., users will find them faster) relative to the control condition with no cues.** The stimulus directs the attention of the users towards the target, reducing the search time.

H_3 **The benefits of imperfect cueing will be restricted to the frequent occasions when the cue is correct. In rare trials, when the cue is wrong, performance will be worse than when there is no cue at all.** Previous studies showed that users tended to follow the instructions and assistance provided by ML [115]. When there is an error in the ML algorithm, it will take time for users to realize that the system failed, and they must change their decisions. This process would decrease the overall search performance.

6.3.2 Participants

A total of 53 participants attended our studies. Each participant was assigned either of 2 conditions (perfect or imperfect cueing). There were 25 participants (17 male, 7 female, and 1 non-binary) with imperfect cueing and 28 (16 male, 12 female) with perfect cueing. Participants consisted of students and staff from Colorado State University, as well as people who were not affiliated with the university. There were 47 participants still studying at a university. The participants' ages ranged between 18 and 55 (M = 25.77, STD = 7.63). Among the 53 participants, 78.15% of the people reported that they didn't play any games regularly, and 84.62% of participants had either used AR or VR in the past. Participants received compensation in the form of $20 or class credits. None of the participants had participated in a prior study of cueing in our lab. All participants had normal or corrected-to-normal vision. This study was approved by the university IRB.

6.3.3 Materials

Participants completed the experiment using the Varjo XR-3 [116] video pass-through AR HMD. The Varjo XR-3 headset can display 3D content over a 90Hz video feed of the real-world environment taken by the headset's cameras. The field of view (FOV) of the device is 115° by 90°. The Varjo XR-3 was selected due to its FOV. To track the objects, an OAK-1 [117] camera was mounted to the front of the Varjo XR-3 (see Figure 6.2). The OAK-1, developed by OpenCV [118], has dedicated hardware for running neural net-

Figure 6.2: OAK-1 camera mounted on XR-3.

works on the device and has a field of view of 68.8° by 42.75°. The camera weighs 53.1g with dimensions of 36 by 54.5 by 27.8 mm and did not block the Varjo sensors. The neural network used was YOLOv5-Nano, which was trained from scratch on a synthetic dataset of 40,000 images. The synthetic dataset was created in Unity 2020.3.27 using 3D models of the 40 target objects + 3 practice objects. The study software was developed on Intel core i9-9900K 3.60GHz, 64GB RAM, and RTX 2080Ti graphics card.

6.3.4 Machine Learning for Imperfect Cueing

A convolutional neural network called YOLOv5 [119] was trained on a fully synthetic dataset with 43 classes. These 43 classes were the target objects (3 practice objects and 40 trial objects). The model never saw a real picture of the targets it searched for until the classification phase. The system used for real-time object detection is YOLOv5-Nano due to its speed and ease of deployment on the OAK camera [117]. The external camera is used to capture the scene, perform neural network inference on the device, and then finally send the coordinates of the objects to the headset. The frames captured by the camera are first down-scaled to 576 x 374 before being sent to the model for inference.

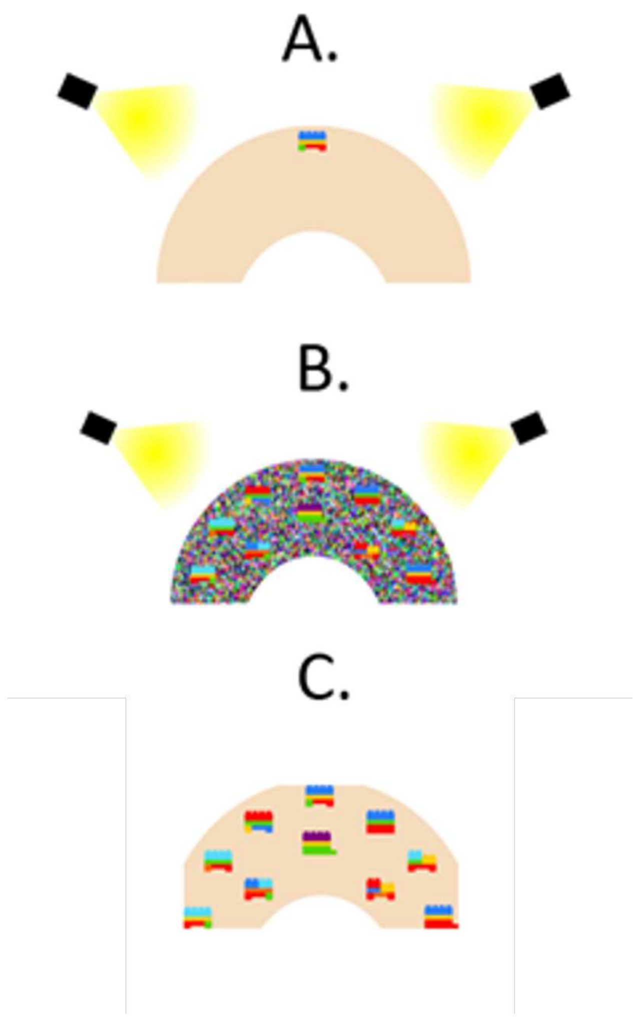

Figure 6.3: Tweaking different parameters A. Not being exposed to all blocks at once, B. confusing background images in the training set, C. not enough lighting in the training set.

40,000 synthetic images were generated using a custom Unity program. The program took 3D models of target objects as input. It created a dataset consisting of images of these models in different orientations and with different backgrounds. The ability to tweak specific parameters allowed us to test subjects' responses to different errors created by AI such as not being exposed to all blocks at once (see Figure 6.3A), confusing background images in the training set (see Figure 6.3B), or not enough lighting in the training set (see Figure 6.3C). The objective was to have a classification accuracy of at least 83% with the understanding that in a live system, this would never be constant. During the study, the average recognition across all the subjects was 88.9%. The recognition rate reported is the rate observed during actual trials, not during the testing phase.

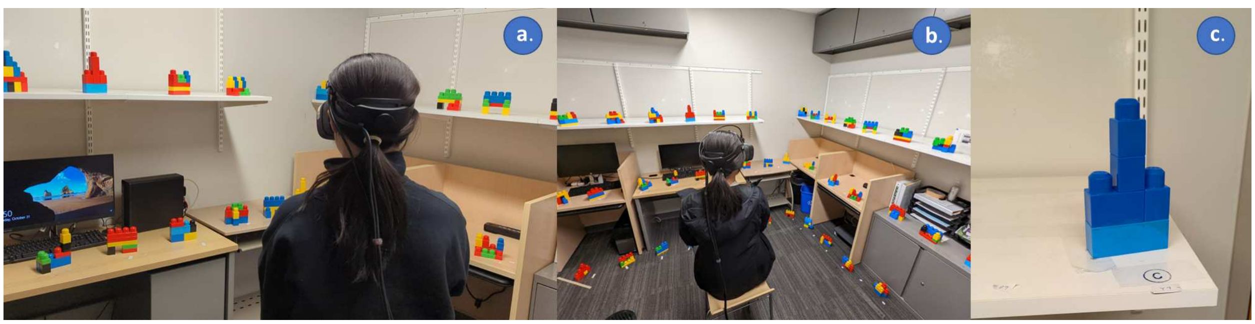

Figure 6.4: a. Head-level view from behind a participant; b. Top-level view from behind a participant; c. Example object used in the study.

The model took in 30,000 training images. The remaining 10,000 images were used as validation images. The model was trained for 3 hours over 300 epochs. Neural network inference on the OAK camera took 49.5ms on average (3.4ms Std Dev). Because of this, we limited the speed of the camera to 20fps. The data is sent over sockets from the OAK camera to Unity, where it is rendered for the headset. This process takes around 10ms, so there is very little delay on the system. The decision for YOLOv5 was to have a real-time recognition system that would be able to run in the computer described in Subsection 6.3.3.

During the actual experiment, the model continuously ran the detection algorithm. This meant that if initially, the model predicted the wrong location for the target object, it would correct itself if the user looked in the correct direction (i.e., the object is in the FOV of the camera). Pilot tests were conducted to check the accuracy of the model.

The coordinates of each object in the 3D environment were stored in the system at the start of the experiment. When the ML algorithm detected the target object in the frame, the location was projected out of the frame in the forward vector direction onto the 3D environment. This gave an approximate position of the target image. Since the coordinates of each object are stored in the system, the closest coordinate was selected as the target object. A threshold of 13 cm in the 3D environment was selected based on pilot study data.

6.3.5 Task

Participants completed a visual search task in 3D space in which they were asked to locate real-world objects (Mega Bloks [120], as shown in Figure 6.4b.), with and without the aid of a target cue presented via the Varjo XR-3 AR HMD video-pass through. Participants were told in the instructions that the system is not 100% reliable and may fail and that they should confirm the suggestion presented by the system. Each trial consisted of three steps (see Figure 6.1):

1. Target image is shown - an image of the target object is shown to the participants, and they press a button to continue when they feel comfortable recalling the object.

2. Searching - participants search the 3D environment for the designated object; there may be a cue or no cue depending on the trial condition.

3. In case of cue condition, selecting or rejecting the suggested object by the cue and searching for the target on their own. The target cue condition consisted of a 3D arrow pointing to the target object's location in 3D space.

The forward vector on the headset is the cursor. When no cues were presented, the center of the screen had a red dot representing the cursor. The red dot is replaced by cues in the cue conditions. When the participants looked at an object, the cursor changed to a circle enveloping the object. The participants held a Vive controller whose trigger button was used for selecting the target. The circle indicated that pressing the trigger would select the object. The red circle represents the system's predicted target. An ML selection was considered as rejected when the participants did not select the suggested object.

This was a mixed-subject design where the groups were split based on different levels of cue reliability (between subjects), and each group had two conditions, cue versus no cue (within-subjects, repeated measures). The two levels of cue reliability were: 1) the locations of all objects were known, and the cue always pointed at the correct object, and 2) the objects were located using ML and had an average accuracy of 88.9%. Participants did not receive a cue to help find the target object for the no-cue condition (i.e., the control condition). Thus, participants in the two groups had identical experiences in the no-cue condition. The search field of the room incorporated 180° surrounding the participant when looking forward from the chair in which they were seated. A total of 40 potential target objects + 3 practice objects were uniformly distributed across the 180° search field in the horizontal direction and approximately 15° in the vertical direction relative to the chair. Objects were placed at three different height levels as shown in Figure 6.4: ground, table, and shelf level (separation between levels was 28in).

6.3.6 Design and Procedures

First, participants signed consent forms and verified their vision. Before starting the experiment, the participants were asked to complete a pre-experiment survey. Next, participants were seated in the middle of the room and assigned either cue with ML or perfect cueing. Every participant was assigned the condition of no cues. The participants completed three practice trials. Participants were presented with either cue or no cue in blocks of 40 trials. The order of these two blocks was counterbalanced across participants. The 40 trials within each block presented 40 different target images in random order. Once the practice trials were done, participants had to find 40 target objects. After attempting to find all 40 objects, the next condition was presented (if a cue was presented first, then no cue for the second condition and vice versa.) In the two blocks, the 40 target objects were the same (40 objects with no cue condition and the same 40 objects with cue condition). In the end, the participants were asked to complete a post-experiment survey. The entire experiment consisted of 86 trials and lasted approximately 35 minutes.

6.4 Results

This study employed a 2 (cue reliability level) x 2 (cue condition) mixed analysis of variance (MANOVA) design to explore the effects of automation reliability on performance and response time. Specifically, the design included two within-subject conditions of cued and not cued and two between-subject conditions of perfect and imperfect automation cueing. The within-subject factor of cued or not cued refers to whether or not participants were given a cue to locate the objects. The cue was intended to guide participants toward more accurate performance. Automation reliability for imperfect conditions was significantly lower than perfect conditions ($t(51) = 8.03$, $p < .001$; mean perfect automation = 100%, mean for imperfect automation = 88.9%).

6.4.1 Accuracy

The effects of cue reliability and cue condition on the accuracy of finding the objects are shown in Figure 6.5. The data was normalized from 0 to 1 (e.g., 0.9 represents 36 correct selections out of 40) on the y-axis. The x-axis has 4 conditions grouped into 2 groups: ML cues and perfect cues. Each of these groups has 2 conditions, with cue and without cue.

A mixed model ANOVA on these data revealed that there was a significant benefit of cueing (Wilks' lambda = .67, $F(1,51) = 24.83$, $p < .001$), (No Cue: M = 89%; Cue: M = 95%). This effect supports H_1, "The addition of cues (perfect/imperfect) produces higher

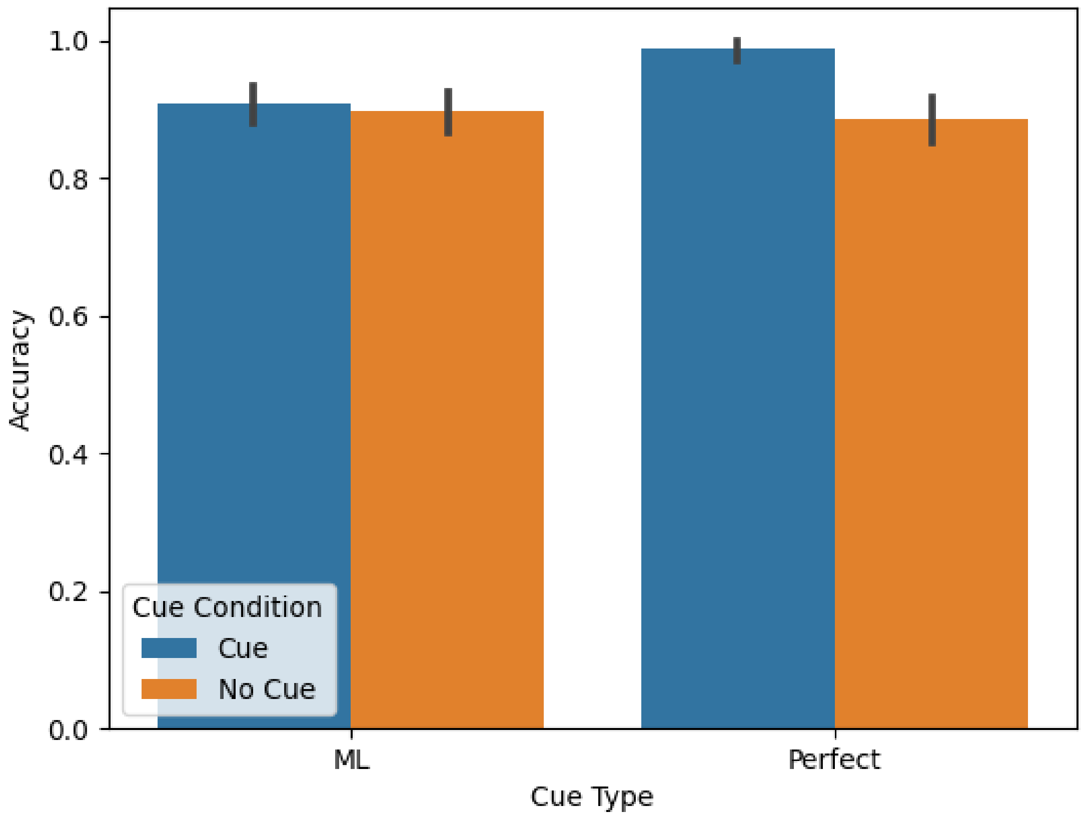

Figure 6.5: Percent accuracy of finding the target when the cues are presented (Blue) vs. no cues (Orange); Different cue conditions (ML vs. Perfect Cue); The bars are standard error bars.

accuracy performance relative to the control condition with no cues." The ANOVA also revealed a significant effect of Groups ($F(1,51) = 4.07$, $p < .05$).This effect can be further understood in the context of the highly significant interaction between cueing and groups (Wilks' lambda = .76, $F(1,51) = 16.15$, $p < .001$). The interaction revealed no difference in accuracy between the two groups in no cue search but a large cueing benefit for the perfect (M = 98.7%) over the imperfect (M = 90.7%) in cued search ($t(51) = 4.98$, $p < .001$). This result supports H_3, "Imperfect cueing automation negatively impacts the overall search performance due to high erroneous searches by humans on those imperfect trials."

When comparing accuracy with imperfect cues (M = 90.7%, SD = 2.85) vs. no cues (M = 89.6%, SD = 3.16), there was no significant difference ($t(24) = 1.05$, $p = .31$). On the other hand, for the perfect cueing group, accuracy with the cue (M = 98.75%), was significantly higher than with no cue (M = 88.48%; t(27) = 5.3, $p < .001$). Thus, the cueing benefit to overall accuracy was only observed when the cue was perfect.

Although there was a relatively large number of unique items to search for, one question is whether performance changed as a function of encountering search for the same objects twice across the experiment. Here we looked at changes in performance from Block 1 to Block 2, which includes both potential changes from a second exposure to items in the search task as well as any changes in general task learning across the session. When comparing the accuracy of finding objects, aided by imperfect cues, in Block 1 (M = 90.4%, SD = 2.82) compared to Block 2 (M = 89.9%, SD = 3.19), there was no significant difference (t(24) = 0.23, p = .82). When comparing the accuracy of finding objects, aided by perfect cues, in Block 1 (M = 94.28%, SD = 4.25) compared to the Block 2 (M = 92.95%, SD = 2.61), there was no significant difference (t(27) = 0.49, p = .62). Given that there was no evidence in this analysis of changes across the experiment, this is congruent with viewing the specific Lego Block shapes twice did not have any significant influence on performance.

6.4.2 Response Time

Response time for the four conditions is plotted in Figure 6.6. A corresponding MANOVA to that carried out with accuracy revealed a main effect of faster responses for cued trials (M = 5.9s) than without (9.4s) (Wilks' lambda = .43, F(1,51) = 66.67, $p < .001$). This effect supports H_2, "The addition of cues produces lower response times relative to the control condition with no cues." There was also a significant difference between the two groups (F(1,51) = 9.87, $p < .005$, partial eta squared = .16); however, as with accuracy, this difference can best be interpreted in the context of the significant interaction between the two variables (Wilks' lambda = .68, F(1,51) = 23.81, $p < .001$). Here, pairwise comparisons between response time in the no-cue conditions (the two orange bars in Figure 6.6) revealed no difference, but between the cued trials, more rapid responses were seen for those who had a perfect cue (M = 3.74s) compared to those for whom the cue was imperfectly reliable (M = 9.2s) ($t(51) = 4.90$, $p < .001$).

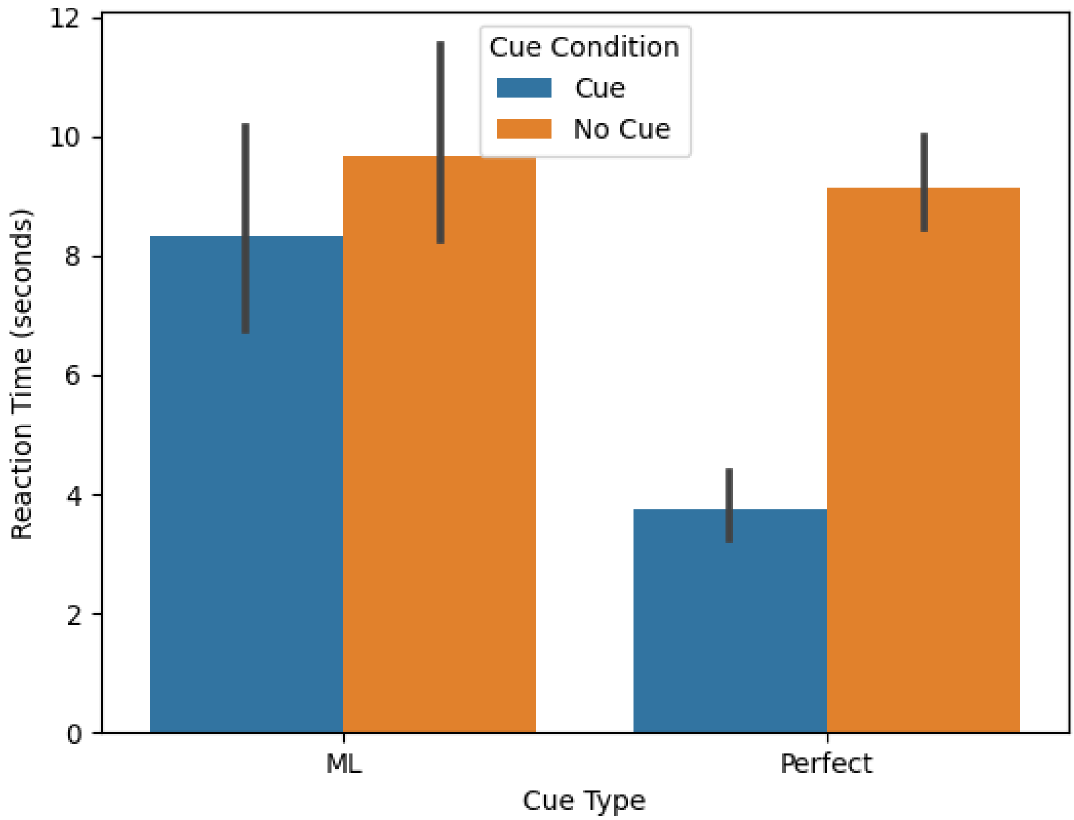

Figure 6.6: Response times of finding the target when the cues are presented (Blue) vs. no cues (Orange); Different cue conditions (ML vs. Perfect Cue); the bars are standard error bars.

When comparing response time when imperfect cues were presented (M = 8.31s, SD = 3.98) vs. no cues (M = 9.67s, SD = 4.57), there was a small but significant benefit of cueing ($t(24) = -2.12$, $p < .05$). In contrast, when presented with a perfect cue (M = 3.74s, SD = 1.49) vs. no cue (M = 9.16s, SD = 2.18), there was a large and significant benefit of cueing ($t(27) = -10.13$, $p < .001$). Thus, the cueing benefit to overall RT was observed for both perfect and imperfect cued conditions.

Although there was a relatively large number of unique items to search for, one question is whether performance changed as a function of encountering search for the same objects twice across the experiment. Here, we looked at changes in performance from Block 1 to Block 2, which includes both potential changes from a second exposure to items in the search task as well as any changes in general task learning across the session. When comparing the response time of finding objects, aided by imperfect cues, in Block 1 (M = 9.07 sec, SD = 5.06) compared to Block 2 (M = 8.91 sec, SD = 3.47), there was no significant difference ($t(24) = 0.23$, $p = .81$). When comparing the response time of finding objects, aided by perfect cues, in Block 1 (M = 6.02 sec, SD = 3.33) compared to Block 2 (M = 6.88 sec, SD = 3.18), there was no significant difference ($t(27) = 0.74$, $p = .47$). Given that there was no evidence in this analysis of changes across the experiment, this is congruent with viewing the specific Lego Block shapes twice did not have any significant influence on performance.

6.4.3 The Speed-Accuracy Trade-off Function

The combined effects of imperfect cueing on accuracy and response time – that is, on overall performance as stated in H_3 - are represented in the speed-accuracy trade-off space in Figure 6.7. Within the figure, high performance (high accuracy, fast responses) can be seen in the upper left region, while poorer performance is in the lower right corner. Figure 6.7 first illustrates, with high prominence, the general equivalence in both measures of the no cue condition, the two circled points. Thus, any concerns about fundamental differences in search performance abilities between the two populations are eliminated. The differential impact of cueing on the two groups is illustrated by the two arrows. When the cues are perfect, their benefit to both dimensions of performance is pronounced, as illustrated by the long dashed orange arrow pointing to the upper left in Figure 6.7. But for the imperfect cueing group, with this relatively small drop in cueing reliability, from 100% to 88.9%, the cues' benefit to accuracy is eliminated, and its benefit to processing

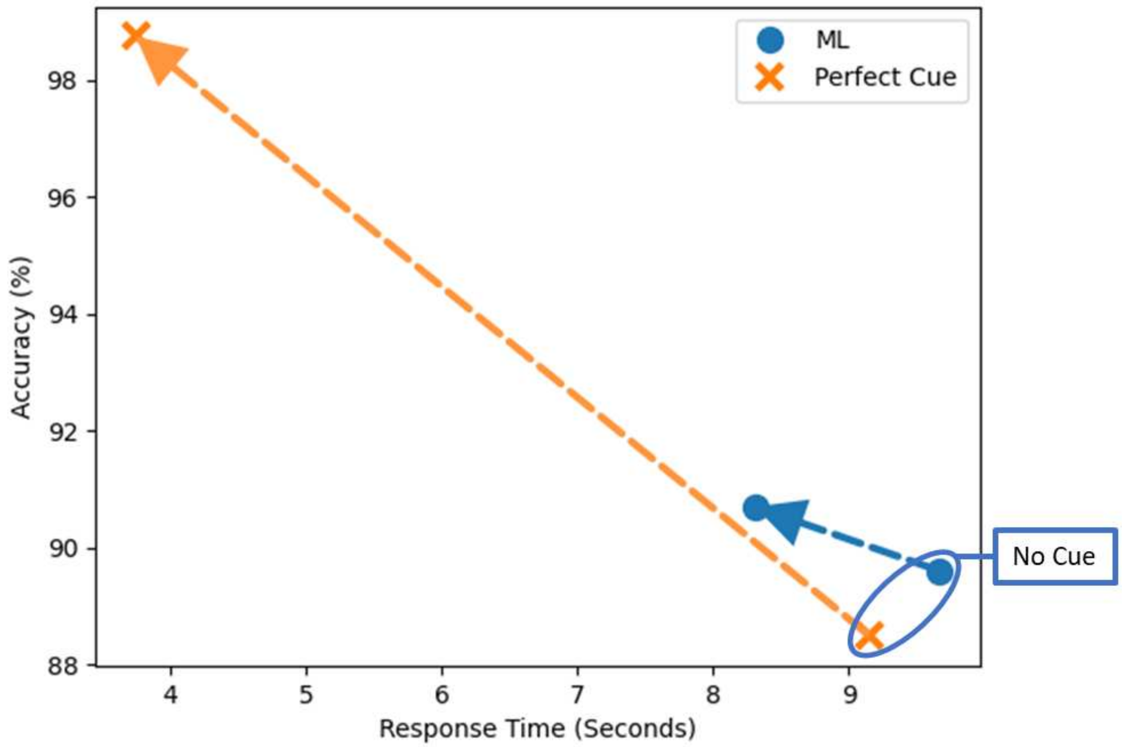

Figure 6.7: In the speed-accuracy trade-off space, the combined impacts of imprecise cueing.

speed, while not eliminated, is greatly reduced, as illustrated by the short dashed blue arrow.

6.4.4 Automation Bias

Section6.1 described automation bias as the tendency to automatically follow the automation's recommendation (here, regarding where the target is). For statistical reasons, this tendency is revealed most clearly by examining participant responses when automation is wrong.

In Figure 6.8, we have used the same speed-accuracy space as in Figure 6.7, to depict the automation bias observed in the current data for the imperfect automation group. In the upper left is depicted the excellent performance when the cue was correct. In the lower right is the performance when the cue was wrong. The figure reveals the large and statistically significant loss in accuracy from 88.3% to 33.1% ($t(23) = 8.15$, $p < .001$) as well as the large increase in response time, from 7.3s to 16.32s ($t(25.7) = 5.04$; $p < .01$).

Collectively, both of these effects further confirm H_3 "Imperfect cueing automation negatively impacts the overall search performance due to high erroneous searches by humans on those imperfect trials." Not only does performance suffer on all search trials with imperfect automation cueing, as shown in Figure 6.7, but time and accuracy performance is particularly reduced when automation errors and the wrong object are cued.

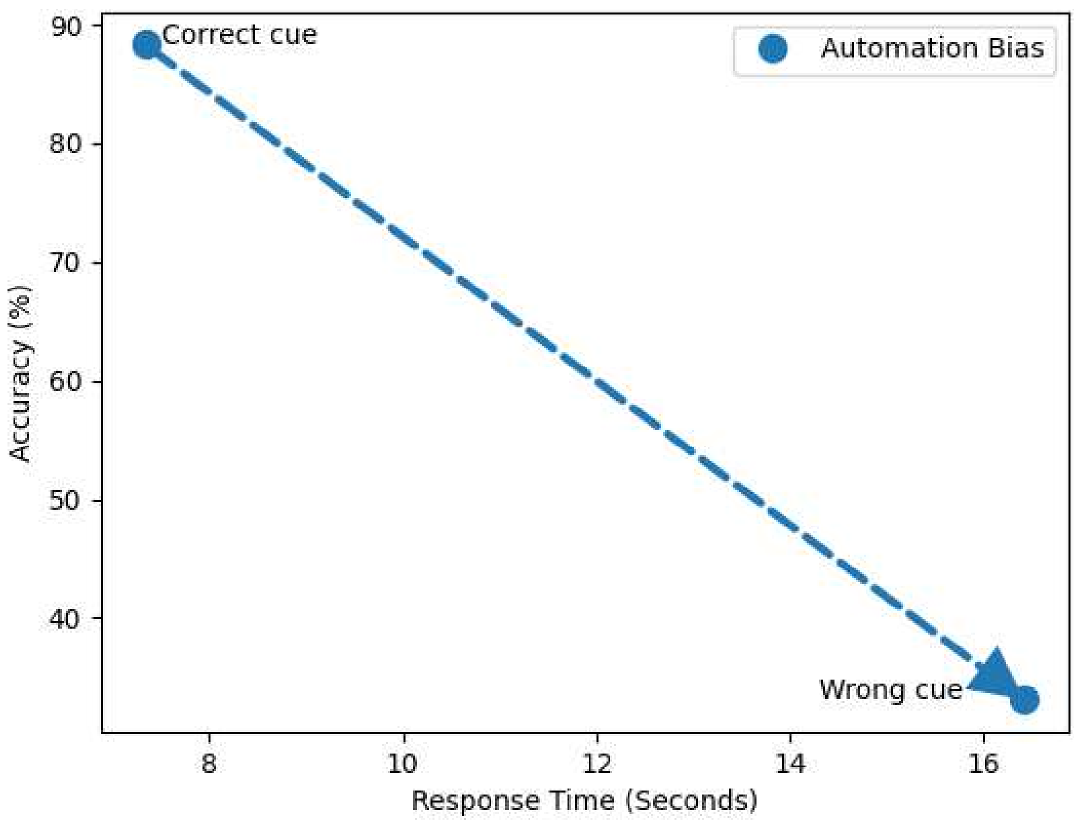

Figure 6.8: In the speed-accuracy trade-off space, automation bias.

The finding of the response time delay when automation is wrong illuminates the cognitive strategies employed by the participants. This delay indicates participants did not always "blindly follow" the automation recommendation. Rather, when they saw a cue they thought was wrong, they tried to figure out the correct target (based on their imperfect memory of the image viewed at the beginning of the trial) but failed. Following that, they followed the automation recommendation nevertheless. This whole cycle took them an additional average of 11.4s.

The two strategies to be followed when the cue is wrong would seemingly produce a different pattern of effects of response time and accuracy. "Blindly following" (the automation bias) would keep response time at its very short value typical of the control condition (8.3s) or the cue-correct condition (7.3s), and produce a near 100% error rate. In contrast, noticing the error and triggering a "re-search" for the correct target would produce this prolonged delay. Accuracy would be improved but would be far from perfect because, by this time, the participant's memory of what the true target looked like would be degraded. We cannot estimate the proportion of trials on which each of these two strategies was deployed, nor the extent to which different participants or different trials for the same participant contributed to this difference in strategy choice, as this study was not intended for such analysis.

6.5 Discussion

The present study investigated the impact of cue reliability on the accuracy and response time during a visual search task presented in an AR environment. It was hypothesized that the addition of cues would produce higher accuracy performance (H_1) and lower response times (H_2) compared to the control condition with no cues consistent with the prior findings of [24,109]. The study results provided partial support for the two previously presented hypotheses. Specifically, when the cue was consistently accurate, it significantly improved both speed and accuracy. However, in the case of imperfect cues (as observed in the ML condition), there was a significant benefit only to speed (reduced RT) but not to accuracy. Furthermore, the study also revealed that search performance is highly impacted by imperfect cueing automation. This was shown both by the lower performance and particularly longer time with imperfect-than-perfect cueing and by highly erroneous searches by participants on the trials when automation was wrong.

A task that required the integration of near domain (display cue) information with far domain (search field) information supported a 5.95% increase in the performance (accuracy) of searching for the target objects. The study's first hypothesis (H_1 - "The addition of cues (perfect/imperfect) produces higher accuracy performance relative to the control condition with no cues") was supported, as the addition of cues produced higher accuracy performance relative to the control condition with no cues. The use of cues can guide participants toward more accurate performance and lead to faster response times. This finding is consistent with previous research that has demonstrated the efficacy of AR cues in improving visual search performance (e.g., [109, 110]).

In comparing the different groups, there was only a 1.1% increase in accuracy performance when the cueing was not reliable. The observation that imperfect cues did not enhance accuracy could be explained by participants harboring doubts about the reliability of these cues. In such instances, they opted to rely on their own flawed perceptual judgment instead. In the case of perfect cue vs. no cue condition, we saw an increase of 10.27% in accuracy performance with the perfect cue. Looking at the mean accuracy performance of the group with perfect cues (M = 98.75%), the results suggest that when cues are perfect, the users tend to follow the cue. The participants who did not perform perfectly were probably influenced by the statement in the instructions about the system not being completely reliable. So, when they may have encountered objects they were unsure of, their trust in the system wavered.

The study's second hypothesis (H_2 - "the addition of cues produces lower response times relative to the control condition with no cues") was fully supported, as the addition of cues produced lower response times relative to the control condition with no cues. Participants who received cues responded (found the objects) significantly faster than those who did not receive cues. The use of cues can guide participants to find objects more rapidly. This finding is also consistent with previous research that has shown the efficacy of AR cues in reducing response times in visual search tasks (e.g., [66]).

Regarding H_3 – "The benefits of imperfect cueing will be restricted to the frequent occasions when the cue is correct. In rare trials when the cue is wrong, performance will be worse than when there is no cue at all." – there was an increase of 5.42s in response time when the cueing was not reliable. Such an increase observed is attributable to two factors: (1) the large increase in 11% of the trials when the automation was wrong (see Figure 6.8), associated with double checking the ML-inferred target, and (2) the general slowing on all trials, associated with the appropriately increased caution by participants who realized the imperfection of the cueing automation. Hence, our hypothesis H_3 was supported.

The use of ML can simplify a lot of tasks and may be essential when the precise nature of the target cannot be specified in advance, but the reliability of such ML is not perfect. While ML can be a powerful tool for automating the detection of target objects, it is important to consider the potential for error. In this study, the average error of the ML model was 11.1%. Looking at just the accuracy of searching for the target, we may be able to interpret that the participants were blindly following the cues presented by the system in most cases (88 out of 112), even when we told each participant that the system may fail regardless of the condition (i.e., perfect vs. imperfect). This led to similar results between the imperfect cue and no cue conditions. But by exploring the results from their response times, we can paint a different picture. There was an increase in response times when the cue was wrong. The participants were able to figure out that the cue provided was false and started looking for the correct target. Some participants found the correct target after that, but most failed. This additional search took, on average, an additional 5 seconds.

Another interesting analysis is how our study compared the recent findings of incorrect cueing using WoZ by Warden et al. [85] to a new system with real-time ML cueing. Our study used a Varjo XR-3 AR video pass-through, which has a much bigger FOV compared to Hololens 2 used by Warden et al. [85]. Another difference is that we used a higher ML accuracy rate of 89% versus a lower accuracy (yet constant by using WoZ) of 83% accuracy. The procedures and participant population demographics were similar between the two studies. Human performance accuracy with the imperfect cues there was 92%. Using our approach with real-time ML was 90.7%. In both studies, the mean response time in the imperfect cueing condition was approximately 8 seconds.

Most critically, the drop in accuracy when the cue was wrong, reflecting the automation bias, was large and highly significant in both studies: a drop to 15% was reported by Warden et al. [85] compared to 31% of our study. This underlines that a real system not only reproduces the automation bias from a WoZ study but, in this case, is worse. Another analysis was the response time on those automation error trials; both studies also revealed a lengthening, indicating that participants did not just blindly follow the incorrect automation guidance.

Furthermore, these results support the WoZ methodology as a reliable experimental approach for studying user interaction. It suggests that the controlled environment of WoZ experiments can accurately simulate real-world interactions, indicating the robustness of this method. We remain highly confident that additional research with ML cueing will continue to produce a pattern of results that are essentially equivalent to those where the cueing error is experimentally imposed (i.e., using WoZ) when the accuracy is around 83% [85, 110].

6.5.1 Limitations

Although the study used different shapes and color combinations of the Mega Bloks™, making them easier to distinguish from the natural environment, a study with actual objects that belong to the environment contextually may provide further insight (e.g., a pen in a home office). Another limitation concerns the ecological validity of the study. Many cases of cued search in the real world do not require comparing objects with an image to be detected. For instance, airport agents scanning luggage for a weapon are not comparing items with an image to be detected; rather, they have a long-term memory representation of what a typical weapon may be. Yet, the objective of our study was to search items to which participants have had no previous exposure, and the current population had not participated in the prior study of Lego block search [85]. Also, we did not use experts in visual searches, such as radiologists personnel. Future work may examine the effects of AR cueing with visual search experts. Finally, while the study may improve external validity by using a real-time ML system for visual search, it limits its internal validity due to the classification accuracy of the ML model presented to each participant varied by a small number. In a WoZ study, the classification accuracy is kept constant. However, as described in the Result and Discussion Sections, our study validates the control WoZ studies done by other researchers (e.g., [85, 110]), which also allows for rapid prototyping and iterative design, potentially saving time and resources during the early stages of research and application development.

6.5.2 Design Implications

The study has several implications for the design of cueing systems in visual search tasks. First, the study highlights the potential benefits of using cues to improve search performance and reduce response times, which can inform the design of user experiences. Our results can be used in real-world applications, and developers can create more intuitive and user-friendly interfaces.

However, the study also highlights the potential risks associated with imperfect automation, which can negatively impact overall search performance. To mitigate these risks, designers of cueing systems should consider the reliability of the automation and provide clear feedback to users on the reliability of the cues.

Additionally, designers should consider the potential for automation bias and design cues that encourage users to actively engage in the search task rather than relying solely on automation.

Finally, the study has implications for the use of ML in the design of cueing systems. While ML can provide a powerful tool for automating the detection of target objects, it is important to consider the potential for error and design cues that take into account the limitations of ML systems.

6.6 Conclusion and Future Work

The presented research study provides insightful information on the potential benefits and risks of using cues to improve visual search efficiency. This is in the context of imperfect automation. The study emphasizes the necessity for cueing system designers to carefully evaluate the reliability of automation. Taking this into account, they can develop cues that motivate users to actively participate in the search task. The study emphasizes the potential for ML as an effective tool for automating the identification of target objects. However, it also emphasizes the necessity of considering limitations and the potential for errors when in the design of cueing systems.

In future work, the study may be expanded to look at different objects (real-world objects that contextually belong to the environment) and add walking to the search. In addition, introducing multiple target objects and multiple cues may present interesting findings. For example, if you have two cues, is it possible to decrease automation bias? We believe that this may be possible through a combination of cues that further increase the trust and reliability of the systems being used and, therefore, increase user performance. It will also be interesting to investigate how predictive cue information influences user trust when guiding through a sequence of tasks. We also plan to analyze the effect of cueing error sources with a within-subjects experiment in the future.

Chapter 7

Conclusion

This dissertation investigated how to optimize user experience in Augmented Reality (AR) environments by examining information delivery and visual search tasks.

Chapters 4 and 5 explored notification design, finding that notifications positioned directly on relevant objects with synchronized audio cues were most effective. This informs the development of AR interfaces that minimize disruption while maximizing information delivery.

Chapter 6 focused on visual search tasks within AR, highlighting the impact of imperfect cues generated by machine learning systems. The research demonstrated a high level of automation bias when users rely on visual cues. These findings emphasize the importance of reliable cueing systems and user training to mitigate reliance on imperfect cues.

Future research directions include exploring alternative notification designs, investigating user experience with mixed physical and virtual environments, and delving deeper into visual search tasks with real-world objects and increased complexity. By understanding user experience in AR environments, we can develop AR interfaces that are informative, user-friendly, and effective in supporting a variety of tasks.

7.1 Contributions

This dissertation advances the understanding of HCI and AR through the following key contributions, each elaborated upon for a deeper understanding:

1. Improved Notification Design:

This research analyzed user performance and preferences for visual and auditory notification modalities within AR HMDs. The study provides clear and actionable guidelines for developers by identifying the most effective placement (directly on relevant objects) and modality (combined visual and audio cues). These guidelines promote the design of user-centered AR interfaces that deliver information efficiently while minimizing user distraction.

2. Mitigating Disruption and Enhancing User Experience:

The research demonstrates how strategically designed AR notifications can minimize disruption and even assist users in completing tasks. This knowledge is crucial for developers who aim to create user-centered AR applications that prioritize a seamless and efficient user experience.

3. Understanding User Behavior in AR:

The study provides valuable insights into how users respond to different notifications within an AR environment. This knowledge can be leveraged to develop future AR interface design principles that cater to user needs and optimize interaction patterns within AR environments.

4. Impact of Imperfect Cues in AR Visual Search:

This research investigates a critical factor in AR visual search tasks: the impact of errors in visual cues. The study explores how these errors can affect user performance and potentially lead to automation bias, where users over-rely on cues even when they are incorrect. These findings can inform the design of reliable AR guidance systems that minimize cue errors.

5. Quantifying Automation Bias in AR Visual Search with Machine Learning:

The study demonstrates a high level of automation bias when users rely on visual cues generated by a machine learning system during a visual search task. Furthermore, the research reveals that the magnitude of automation bias was even greater using the ML system compared to a WoZ study due to the better performance of "automation".

7.2 Design Implications

This dissertation offers valuable insights for developers and designers of AR interfaces, particularly focusing on user interaction through notification design and visual search tasks with cues.

The research continuously emphasizes the advantages of integrating auditory and visual modalities as notifications in AR apps. The total task performance, noticeability, and user engagement are all improved by this integrated approach. The placement of visual notifications is important. Their capacity to locate and engage with items in the AR environment is enhanced when they are positioned over relevant objects. The research proposes fixed visual alerts inside the user's field of view as a workable substitute for voice notifications in situations when they are unsuitable. This method keeps the visual signals noticeable and performs similarly without excessively interfering with the primary activity.

The study also draws attention to the possible dangers of imperfect automation since search performance may suffer from imperfect cues. In order to solve this, designers must consider the automation system's reliability and give consumers explicit feedback on how reliable the cues are. Although ML is a strong tool for automating target object recognition in augmented reality experiences, designers need to be aware of possible inaccuracies. Cues ought to be created with the limits of ML systems in mind.

Developers and designers may improve user experience and enable successful task completion in a range of AR applications by integrating these design implications into their work to produce more efficient and user-centered AR interfaces.

7.3 Limitations

Although every chapter in this dissertation provides insightful information about the user experience with AR interfaces, it's crucial to recognize certain limitations in all of the presented research. These drawbacks point to areas that need further research and improvement as we work to create AR systems that are centered around users:

Ecological Validity: One issue that keeps coming up in every chapter is how controlled the research environments used in this study are. Although these controlled environments provided a relatively realistic environment, allowed for the manipulation of experimental conditions and allowed for the isolation of variables, they might not have adequately captured the complexity of real-world AR use cases. A user's experience using AR interfaces can be influenced by a wide range of elements found in real-world surroundings, including background clutter, ambient noise, and different task difficulties.

Task Complexity: The studies' challenges had varying degrees of complexity. As mentioned in Chapters 4 and 5, simpler tasks might have made some participants believe that notifications are not as important as they had thought, which could have an effect on how noticeable and dependent they are. Future research ought to examine the relationship between task complexity, notification design, and user behavior in augmented reality interfaces.

In addition to the general limitations outlined above, some limitations are specific to each chapter:

Notification Design: Although ARtisan Bistro was deemed a useful platform for researching notification design in Chapter 3, it is vital to take into account the limits of Microsoft HoloLens 2's hand-tracking technology and its FoV. These restrictions could have affected how users engaged with notifications and virtual objects in the AR environment.

Visual Search Activities: Chapter 6 examined how users interacted with AR cues while completing visual search activities. The usage of Mega Bloks with distinct forms and colors may not translate to real-world search circumstances where things naturally blend into the surroundings. Furthermore, the study concentrated on item search with little prior exposure, which would not accurately represent search behavior in the actual world that depends on long-term memory representations. Future studies might involve people with search expertise and examine how well AR cues perform in more ecologically realistic search tasks.

Despite these limitations, the study described in this dissertation significantly adds to our knowledge of how users interact with augmented reality interfaces, especially regarding notification design and visual search tasks with imperfect cues. By recognizing these limitations and suggesting directions for future study, this dissertation lays the way for the creation of more intuitive, user-friendly, and successful AR interfaces for a wider variety of applications.

7.4 Future Work

This dissertation has laid a strong foundation for understanding user experience in AR environments. However, the field of AR is rapidly evolving, presenting numerous exciting avenues for future research. Here, we explore potential directions to build upon the findings of this dissertation:

1. Expanding Notification Design Research:

Advanced Personalization: Notifications can be further personalized to user preferences and task context. Research could explore user-configurable notification settings, allowing users to adjust modality (visual, auditory), placement, and level of detail based on their needs. Additionally, the system could learn individual user preferences over time, dynamically adapting notification design for optimal effectiveness.

Transitioning to Physical and Virtual Blends: This research has focused on purely virtual environments. Future studies can investigate user experience in mixed reality (MR) settings, where virtual objects and notifications interact with the physical world. This requires exploring how AR notifications adapt to real-world contexts, potentially blending with or dynamically responding to physical objects and user interactions with the physical environment.

Haptic Integration: Haptic feedback can be a powerful tool in MR. Future research can explore how haptic notifications can be integrated with AR interfaces, providing users with additional information channels and potentially enhancing task completion efficiency.

2. Advanced Visual Search in AR:

Real-World Object Search: This research focused on virtual objects within a controlled environment. Future studies can investigate how AR visual search functions in real-world settings with a wider variety of objects that belong in the context. This includes exploring the impact of object occlusion, lighting variations, and cluttered environments on visual search performance and cue reliability.

Multiple Target Objects and Cues: The research examined single target object searches. Future studies can explore how visual search functions with multiple target objects and investigate the effectiveness of combining different cue modalities (e.g., visual, audio, haptic) or just different cue types to guide users toward multiple targets. For example, an image companies the arrow so that users can reaffirm their choice.

Attention Management and Cognitive Load: AR visual search tasks can introduce cognitive load. Future research can explore strategies to manage user attention within AR environments, potentially including dynamic cue adjustment based on task complexity or visual clutter.

Bibliography

[1] Aditya Raikwar, Domenick Mifsud, Christopher D. Wickens, Brendan Kelly, Amelia C. Warden, Benjamin A. Clegg, and Francisco R. Ortega. Beyond the wizard of oz: Using imperfect machine learning to examine the impact of reliability of augmented reality cues on visual search performance. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 437–438, 2023.

[2] Iring Koch, Edita Poljac, Hermann Müller, and Andrea Kiesel. Cognitive structure, flexibility, and plasticity in human multitasking—an integrative review of dual-task and task-switching research. *Psychological bulletin*, 144(6):557, 2018.

[3] Sandy Gould, Duncan Brumby, Anna Cox, Víctor González, Dario Salvucci, and Niels Taatgen. Multitasking and interruptions: a sig on bridging the gap between research on the micro and macro worlds. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 1189–1192. Association for Computing Machinery, 2012.

[4] Tilo Westermann, Sebastian Möller, and Ina Wechsung. Assessing the relationship between technical affinity, stress and notifications on smartphones. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '15, page 652–659, New York, NY, USA, 2015. Association for Computing Machinery.

[5] Martin Pielot and Luz Rello. The do not disturb challenge: A day without notifications. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, page 1761–1766, New York, NY, USA, 2015. Association for Computing Machinery.

[6] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. Large-scale assessment of mobile notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 3055–3064, New York, NY, USA, 2014. Association for Computing Machinery.

[7] Philipp Müller, Sander Staal, Mihai Bâce, and Andreas Bulling. Designing for noticeability: Understanding the impact of visual importance on desktop notifications. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[8] Alexandra Voit, Dominik Weber, and Niels Henze. Qualitative investigation of multi-device notifications. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, UbiComp '18, page 1263–1270, New York, NY, USA, 2018. Association for Computing Machinery.

[9] Seul-Kee Kim, So-Yeong Kim, and Hang-Bong Kang. An analysis of the effects of smartphone push notifications on task performance with regard to smartphone overuse using erp. *Intell. Neuroscience*, 2016:9, jun 2016.

[10] Joshua D Upshaw, Carl E Stevens Jr, Giorgio Ganis, and Darya L Zabelina. The hidden cost of a smartphone: The effects of smartphone notifications on cognitive control from a behavioral and electrophysiological perspective. *PloS One*, 17(11):e0277220, 2022.

[11] Sandra Ohly and Luca Bastin. Effects of task interruptions caused by notifications from communication applications on strain and performance. *Journal of Occupational Health*, 65(1):e12408, 06 2023.

[12] Kostadin Kushlev, Jason Proulx, and Elizabeth W. Dunn. "silence your phones": Smartphone notifications increase inattention and hyperactivity symptoms. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 1011–1020, New York, NY, USA, 2016. Association for Computing Machinery.

[13] Shamsi T Iqbal and Eric Horvitz. Notifications and awareness: a field study of alert usage and preferences. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 27–30, 2010.

[14] Ronald Azuma and Chris Furmanski. Evaluating label placement for augmented reality view management. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pages 66–75. IEEE, 2003.

[15] Mark Billinghurst, Adrian Clark, Gun Lee, et al. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, 8(2-3):73–272, 2015.

[16] Jarkko Polvi, Takafumi Taketomi, Atsunori Moteki, Toshiyuki Yoshitake, Toshiyuki Fukuoka, Goshiro Yamamoto, Christian Sandor, and Hirokazu Kato. Handheld guides in inspection tasks: Augmented reality versus picture. *IEEE transactions on visualization and computer graphics*, 24(7):2118–2128, 2017.

[17] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE transactions on visualization and computer graphics*, 23(6):1706–1724, 2016.

[18] Timothy Sohn, Kevin A Li, William G Griswold, and James D Hollan. A diary study of mobile information needs. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 433–442, 2008.

[19] JS Wolffsohn, NA McBrien, Graham K Edgar, and T Stout. The influence of cognition and age on accommodation, detection rate and response times when using a car head-up display (hud). *Ophthalmic and physiological optics*, 18(3):243–253, 1998.

[20] G. E. Burnett and R. A. Donkor. Evaluating the impact of head-up display complexity on peripheral detection performance: a driving simulator study. *Advances in transportation studies*, 28:5–16, 2012.

[21] Ulric Neisser. Visual search. *Scientific American*, 210(6):94–103, 1964.

[22] Anne M. Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980.

[23] Jeremy M Wolfe. Guided search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, 28(4):1060–1092, 2021.

[24] Raja Parasuraman and Dietrich H. Manzey. Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3):381–410, 2010. PMID: 21077562.

[25] Celeste Lyn Paul, Anita Komlodi, and Wayne Lutters. Interruptive notifications in support of task management. *International Journal of Human-Computer Studies*, 79:20–34, 2015.

[26] Hvuniin Lee and Woontack Woo. Investigating display position of a head-fixed augmented reality notification for dual-task. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 584–585. IEEE, 2022.

[27] Nicky Kern and Bernt Schiele. Context-aware notification for wearable computing. In *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.*, pages 223–223. Citeseer, 2003.

[28] Kim Salazar. Indicators, validations, and notifications: Pick the correct communication option. https://www.nngroup.com/articles/indicators-validations-notifications/. Published: 2015-07-25.

[29] Markus Wallmyr, Taufik Akbar Sitompul, Tobias Holstein, and Rikard Lindell. Evaluating mixed reality notifications to support excavator operator awareness. In *IFIP Conference on Human-Computer Interaction*, pages 743–762. Springer, 2019.

[30] Chris Harrison, Brian Y Lim, Aubrey Shick, and Scott E Hudson. Where to locate wearable displays? reaction time performance of visual alerts from tip to toe. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 941–944, 2009.

[31] Android and Google. Android notifications. https://developer.android.com/develop/ui/views/notifications.

[32] Apple. Apple notifications. https://support.apple.com/en-us/108781.

[33] Dominik Weber, Sven Mayer, Alexandra Voit, Rodrigo Ventura Fierro, and Niels Henze. Design guidelines for notifications on smart tvs. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, pages 13–24, 2016.

[34] John D Lee, Joshua D Hoffman, and Elizabeth Hayes. Collision warning design to mitigate driver distraction. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pages 65–72, 2004.

[35] Koen Crommentuijn and Fredrik Winberg. Designing auditory displays to facilitate object localization in virtual haptic 3d environments. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '06, page 255–256, New York, NY, USA, 2006. Association for Computing Machinery.

[36] André Zenner, Marco Speicher, Sören Klingner, Donald Degraen, Florian Daiber, and Antonio Krüger. Immersive notification framework: adaptive & plausible notifications in virtual reality. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2018.

[37] Sarthak Ghosh, Lauren Winston, Nishant Panchal, Philippe Kimura-Thollander, Jeff Hotnog, Douglas Cheong, Gabriel Reyes, and Gregory D Abowd. Notifivr: exploring interruptions and notifications in virtual reality. *IEEE transactions on visualization and computer graphics*, 24(4):1447–1456, 2018.

[38] Facebook. Facebook smart glasses in partnering with rayban. https://www.ray-ban.com/usa/discover-ray-ban-stories/clp. Published: 2021-09-09.

[39] Kate Kozuch. Apple glasses updates. https://www.tomsguide.com/news/apple-glasses. Published: 2021-09-30.

[40] Xiaomi Team. Xiaomi smart glasses. https://blog.mi.com/en/2021/09/14/xiaomi-unveils-xiaomi-smart-glasses/. Published: 2021-09-14.

[41] Andrés Lucero and Akos Vetek. Notifeye: using interactive glasses to deal with notifications while walking in public. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*, pages 1–10, 2014.

[42] Rufat Rzayev, Paweł W Woźniak, Tilman Dingler, and Niels Henze. Reading on smart glasses: The effect of text position, presentation type and walking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–9, 2018.

[43] Pin-Sung Ku, Yu-Chih Lin, Yi-Hao Peng, and Mike Y Chen. Peritext: Utilizing peripheral vision for reading text on augmented reality smart glasses. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 630–635. IEEE, 2019.

[44] WOLFGANG JASCHINSKI*, Herbert Heuer, and Hannegret Kylian. A procedure to determine the individually comfortable position of visual displays relative to the eyes. *Ergonomics*, 42(4):535–549, 1999.

[45] Nicholas F Polys, Seonho Kim, and Doug A Bowman. Effects of information layout, screen size, and field of view on user performance in information-rich virtual environments. *Computer Animation and Virtual Worlds*, 18(1):19–38, 2007.

[46] Varjo Team. Varjo xr-3. https://varjo.com/products/xr-3/.

[47] Bireswar Laha, Doug A Bowman, and James D Schiffbauer. Validation of the mr simulation approach for evaluating the effects of immersion on visual analysis of volume data. *IEEE transactions on visualization and computer graphics*, 19(4):529–538, 2013.

[48] Donghao Ren, Tibor Goldschwendt, YunSuk Chang, and Tobias Höllerer. Evaluating wide-field-of-view augmented reality with mixed reality simulation. In *2016 IEEE Virtual Reality (VR)*, pages 93–102. IEEE, 2016.

[49] Mathieu Rodrigue, Andrew Waranis, Tim Wood, and Tobias Höllerer. Mixed reality simulation with physical mobile display devices. In *2015 IEEE Virtual Reality (VR)*, pages 105–110. IEEE, 2015.

[50] Samat Imamov, Daniel Monzel, and Wallace S Lages. Where to display? how interface position affects comfort and task switching time on glanceable interfaces. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 851–858. IEEE, 2020.

[51] Lucas Plabst, Sebastian Oberdörfer, Francisco Raul Ortega, and Florian Niebling. Push the Red Button: Comparing Notification Placement with Augmented and Non-Augmented Tasks in AR. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction*, pages 1–11, Online CA USA, December 2022. ACM.

[52] Hyunjin Lee and Woontack Woo. Exploring the Effects of Augmented Reality Notification Type and Placement in AR HMD while Walking. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pages 519–529, March 2023.

[53] May Jorella Lazaro, Sungho Kim, Jaeyong Lee, Jaemin Chun, and Myung-Hwan Yun. Interaction Modalities for Notification Signals in Augmented Reality. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, pages 470–477, Montréal QC Canada, October 2021. ACM.

[54] Aditya Raikwar, Lucas Plabst, and Francisco R Ortega. Artisan bistro: A cooking task environment to conduct studies in augmented reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 909–910. IEEE, 2022.

[55] Fast food mega constructor. https://assetstore.unity.com/packages/3d/props/food/fast-food-mega-constructor-209796.

[56] Pizza oven. https://www.turbosquid.com/3d-models/3d-model-morello-forni-pizza-oven-1554795.

[57] Coffee maker. https://www.turbosquid.com/3d-models/coffee-maker-3d-model/896640.

[58] Blender Online Community. *Blender - a 3D modeling and rendering package*. Blender Foundation, 2024.

[59] Starbucks coffee cup. https://www.turbosquid.com/3d-models/starbucks-cup-3ds-free/650917.

[60] Tray. https://www.turbosquid.com/3d-models/3d-tray-06-model-1504272.

[61] Aditya Raikwar, Lucas Plabst, Anil Ufuk Batmaz, Florian Niebling, and Francisco R. Ortega. Ping! your food is ready: Comparing different notification techniques in 3d ar cooking environment, 2024.

[62] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. Windows on the world: 2d windows for 3d augmented reality. In *Proceedings of the 6th annual ACM symposium on User interface software and technology*, pages 145–155, 1993.

[63] Colorado State University NUILAB. Artisan bistro github repo. https://github.com/NuiLab/ARtisan-Bistro.

[64] Sandra G. Hart. Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006.

[65] 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, 2015.

[66] Amelia C. Warden, Christopher D. Wickens, Domenick Mifsud, Shannon Ourada, Benjamin A. Clegg, and Francisco R. Ortega. Visual search in augmented reality: Effect of target cue type and location. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 66(1):373–377, 2022.

[67] Christopher D Wickens, Jason S McCarley, and Robert S Gutzwiller. *Applied attention theory*. CRC press, 2022.

[68] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 143–146, New York, NY, USA, 2011. Association for Computing Machinery.

[69] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. An aligned rank transform procedure for multifactor contrast tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, page 754–768, New York, NY, USA, 2021. Association for Computing Machinery.

[70] Michael G Kenward and James H Roger. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, pages 983–997, 1997.

[71] James R. Lewis. Psychometric evaluation of the post-study system usability questionnaire: The pssuq. *Proceedings of the Human Factors Society Annual Meeting*, 36(16):1259–1260, 1992.

[72] Hyunjin Lee and Woontack Woo. Exploring the effects of augmented reality notification type and placement in ar hmd while walking. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pages 519–529, 2023.

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

[74] S. S. SHAPIRO and M. B. WILK. An analysis of variance test for normality (complete samples)†. *Biometrika*, 52(3-4):591–611, 12 1965.

[75] Morton B. Brown and Alan B. Forsythe. Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346):364–367, 1974.

[76] William H. Kruskal and W. Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621, 1952.

[77] Will T. Interpreting system usability scale (sus) score. https://uiuxtrend.com/measuring-system-usability-scale-sus/.

[78] Nicolas Barbotin, James Baumeister, Andrew Cunningham, Thierry Duval, Olivier Grisvard, and Bruce H. Thomas. Evaluating visual cues for future airborne surveillance using simulated augmented reality displays. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 213–221, 2022.

[79] Chia-Chien Wu and Jeremy M. Wolfe. Eye movements in medical image perception: A selective review of past, present and future. *Vision*, 3(2):32, Jun 2019.

[80] CG Drury. *Visual search in industrial inspection*. Taylor & Francis London, 1990.

[81] A.K. Gramopadhye, C.G. Drury, X. Jiang, and R. Sreenivasan. Visual search and visual lobe size: can training on one affect the other? *International Journal of Industrial Ergonomics*, 30(3):181–195, 2002.

[82] Marvin M. Chun. Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4(5):170–178, 2000.

[83] Frank Biocca, Arthur Tang, Charles Owen, and Fan Xiao. Attention funnel: Omnidirectional 3d cursor for mobile augmented reality platforms. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, page 1115–1122, New York, NY, USA, 2006. Association for Computing Machinery.

[84] Steven J Henderson and Steven Feiner. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 135–144. IEEE, 2009.

[85] Amelia C. Warden, Christopher D. Wickens, Daniel Rehberg, Francisco R. Ortega, and Benjamin A. Clegg. Fast, accurate, but sometimes too-compelling support: The impact of imperfectly automated cues in an augmented-reality head-mounted display on visual search performance. *IEEE Transactions on Human-Machine Systems*, pages 1–12, 2023.

[86] Kathleen L. Mosier and Linda J. Skitka. Automation use and automation bias. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 43(3):344–348, 1999.

[87] Samuel Dodge and Lina Karam. A study and comparison of human and deep learning recognition performance under visual distortions. In *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–7, 2017.

[88] Lucas Beyer, Olivier J. Hénaff, Alexander Kolesnikov, Xiaohua Zhai, and Aäron van den Oord. Are we done with imagenet? *CoRR*, abs/2006.07159, 2020.

[89] Dhruva K. Patil. *Something Is Fishy! - How Ambiguous Language Affects Generalization of Video Action Recognition Networks*. PhD thesis, Colorado State University, 2022. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2023-03-08.

[90] Max Schemmer, Niklas Kühl, Carina Benz, and Gerhard Satzger. On the influence of explainable ai on automation bias, 2022.

[91] Steven Yantis. Stimulus-driven attentional capture. *Current Directions in Psychological Science*, 2(5):156–161, 1993.

[92] Michael I. Posner. Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1):3–25, 1980.

[93] Michael I Posner, Charles R Snyder, and Brian J Davidson. Attention and the detection of signals. *Journal of experimental psychology: General*, 109(2):160, 1980.

[94] Michael I. Posner. Orienting of attention: Then and now. *Quarterly Journal of Experimental Psychology*, 69(10):1864–1875, 2016. PMID: 25176352.

[95] Christopher D Wickens, Jason S McCarley, and Robert S Gutzwiller. *Applied attention theory*. CRC press, 2023.

[96] Chia-Chien Wu and Jeremy M. Wolfe. Eye movements in medical image perception: A selective review of past, present and future. *Vision*, 3(2):32, Jun 2019.

[97] Minh Tien Phan, Indira Thouvenin, and Vincent Frémont. Enhancing the driver awareness of pedestrian using augmented reality cues. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1298–1304, 2016.

[98] Weiquan Lu, Henry Been-Lirn Duh, Steven Feiner, and Qi Zhao. Attributes of subtle cues for facilitating visual search in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):404–412, 2014.

[99] Radha Kumaran, You-Jin Kim, Anne E Milner, Tom Bullock, Barry Giesbrecht, and Tobias Höllerer. The impact of navigation aids on search performance and object recall in wide-area augmented reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[100] Ben Shneiderman. *Human-centered AI*. Oxford University Press, 2022.

[101] JOHN LEE and NEVILLE MORAY. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10):1243–1270, 1992. PMID: 1516577.

[102] Kevin Anthony Hoff and Masooda Bashir. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3):407–434, 2015. PMID: 25875432.

[103] Megan L. Bartlett and Jason S. McCarley. Benchmarking aided decision making in a signal detection task. *Human Factors*, 59(6):881–900, 2017. PMID: 28796974.

[104] Melanie M. Boskemper, Megan L. Bartlett, and Jason S. McCarley. Measuring the efficiency of automation-aided performance in a simulated baggage screening task. *Human Factors*, 64(6):945–961, 2022. PMID: 33508964.

[105] Juliana Goh, Douglas A. Wiegmann, and Poornima Madhavan. Effects of automation failure in a luggage screening task: A comparison between direct and indirect cueing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(3):492–496, 2005.

[106] John D. Lee and Katrina A. See. Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1):50–80, 2004. PMID: 15151155.

[107] Arne Seeliger, Raphael P. Weibel, and Stefan Feuerriegel. Context-adaptive visual cues for safe navigation in augmented reality using machine learning. *International Journal of Humanâ€"Computer Interaction*, 0(0):1–21, 2022.

[108] Michelle Yeh, Christopher D. Wickens, and F. Jacob Seagull. Target cuing in visual search: The effects of conformality and display location on the allocation of visual attention. *Human Factors*, 41(4):524–542, 1999. PMID: 10774124.

[109] Domenick Mifsud, Chris Wickens, Michael Maulbeck, Peter Crane, and Francisco R. Ortega. The effectiveness of gaze guidance lines in supporting jtac's attention allocation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 66(1):2198–2201, 2022.

[110] Michelle Yeh and Christopher D. Wickens. Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors*, 43(3):355–365, 2001. PMID: 11866192.

[111] Robert Sargent, Brett Walters, and Chris Wickens. Meta-analysis qualifying and quantifying the benefits of automation transparency to enhance models of human performance. In *Human-Computer Interaction. Theoretical Approaches and Design Methods: Thematic Area, HCI 2023, Held as Part of the 25th HCI International Conference, HCII 2023, Hybrid Event, July 23–July 28, 2023, Proceedings, Part I*. Springer, 2023. In print.

[112] Michelle Yeh, James L. Merlo, Christopher D. Wickens, and David L. Brandenburg. Head up versus head down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling. *Human Factors*, 45(3):390–407, 2003. PMID: 14702991.

[113] Brandon Victor Syiem, Ryan M. Kelly, Jorge Goncalves, Eduardo Velloso, and Tilman Dingler. Impact of task on attentional tunneling in handheld 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.

[114] Takuya Honda, Nobuhiro Hagura, Toshinori Yoshioka, and Hiroshi Imamizu. Imposed visual feedback delay of an action changes mass perception based on the sensory prediction error. *Frontiers in psychology*, 4:760, 2013.

[115] Mary Jo Bitner, Amy L Ostrom, and Matthew L Meuter. Implementing successful self-service technologies. *Academy of management perspectives*, 16(4):96–108, 2002.

[116] Varjo xr-3. https://varjo.com/products/xr-3/. Accessed: 2023-03-24.

[117] Oak-1. https://store.opencv.ai/products/oak-1. Accessed: 2023-03-24.

[118] Opencv. https://opencv.org/. Accessed: 2023-03-24.

[119] Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, NanoCode012, Yonghye Kwon, Kalen Michael, TaoXie, Jiacong Fang, imyhxy, Lorna, Zeng Yifu, Colin Wong, Abhiram V, Diego Montes, Zhiqiang Wang, Cristi Fati, Jebastin Nadar, Laughing, UnglvKitDe, Victor Sonck, tkianai, yxNONG, Piotr Skalski, Adam Hogan, Dhruv Nair, Max Strobel, and Mrinal Jain. ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation, nov 2022.

[120] Mega bloks. https://shop.mattel.com/collections/mega-bloks. Accessed: 2023-03-24.

Appendix A

Experiment Raw Results

A.1 Notifications Experiment 1 Post-Study System Usability Questionnaire

Participant Number	Overall, I am satisfied with how easy it is to use this system.	It was simple to use this system.	I could effectively complete the tasks and scenarios using this system.	I was able to complete the tasks and scenarios quickly using this system.	I felt comfortable using this system.	It was easy to learn to use this system.
1	0	0	2	2	0	0
2	2	2	1	4	3	0
3	1	2	1	3	2	2
4	1	1	3	3	3	0
5	4	2	2	2	2	3
6	2	2	4	4	1	0
7	0	0	0	0	0	0
8	2	2	2	3	2	0
9	1	1	2	1	2	1
10	1	2	0	1	1	0
11	1	1	1	1	0	0
12	0	0	4	4	0	0
13	2	1	1	1	1	0
14	1	0	1	0	0	0
15	0	0	0	0	0	1
16	2	1	3	3	1	0
17	1	0	2	2	2	0
18	1	0	0	0	0	0
19	2	1	0	1	1	0
20	3	2	3	4	3	0
21	0	1	1	1	0	1
22	1	1	0	1	2	0
23	2	3	2	3	0	0
24	0	0	0	0	0	0
25	0	0	3	3	0	0
26	1	1	1	1	1	0
Mean	**1.19**	**1.00**	**1.50**	**1.85**	**1.04**	**0.31**
STD	**1.00**	**0.88**	**1.25**	**1.38**	**1.06**	**0.72**
<1.75	**Yes**	**Yes**	**Yes**	**No**	**Yes**	**Yes**

Figure A.1: PSSUQ survey results - First Six Questions.

Participant Number	I believe I could become productive quickly using this system.	The system gave error messages that clearly told me how to fix problems.	Whenever I made a mistake using the system, I could recover easily and quickly.	The information (such as on-line help, on-screen messages, and other documentation) provided with this system was clear.	It was easy to find the information I needed.	The information provided for the system was easy to understand.
1	0	6	3	4	4	4
2	1	0	1	1	3	0
3	3	4	3	2	2	2
4	3	2	3	3	3	3
5	3	3	2	1	1	1
6	1	0	0	1	1	1
7	0	0	1	0	0	0
8	2	4	1	0	0	0
9	2	3	2	0	0	0
10	0	3	2	1	2	1
11	0	2	1	1	0	0
12	3	0	1	3	3	0
13	0	4	2	2	1	1
14	0	3	1	2	5	0
15	1	6	3	2	2	1
16	1	2	1	3	2	1
17	1	0	5	0	1	1
18	0	0	0	0	0	0
19	2	1	1	0	1	0
20	1	5	1	0	1	1
21	0	0	0	0	0	0
22	0	5	4	1	2	1
23	0	3	3	0	0	0
24	1	2	1	0	0	0
25	0	2	2	1	1	1
26	1	2	1	1	1	1
Mean	**1.00**	**2.38**	**1.73**	**1.12**	**1.38**	**0.77**
STD	**1.07**	**1.88**	**1.23**	**1.15**	**1.33**	**0.97**
<1.75	**Yes**	**No**	**Yes**	**Yes**	**Yes**	**Yes**

Figure A.2: PSSUQ survey results - Middle Six Questions.

Participant Number	The information was effective in helping me complete the tasks and scenarios.	The organization of information on the system screens was clear.	The interface of this system was pleasant.	I liked using the interface of this system.	This system has all the functions and capabilities I expect it to have.	Overall, I am satisfied with this system.
1	4	4	0	0	1	1
2	0	0	3	5	2	2
3	1	1	2	4	1	2
4	3	3	3	3	3	3
5	2	1	2	2	1	1
6	0	0	1	1	2	1
7	0	0	1	1	1	1
8	0	0	1	1	0	1
9	0	0	2	1	1	1
10	3	2	1	1	1	1
11	0	0	0	0	0	0
12	0	0	0	0	1	0
13	1	2	3	2	3	2
14	6	0	0	0	0	0
15	1	1	1	1	1	1
16	3	3	1	1	2	2
17	1	1	1	0	0	0
18	0	0	0	0	0	0
19	1	1	0	1	2	1
20	2	0	2	2	2	2
21	0	1	3	3	1	1
22	1	2	2	1	4	1
23	0	0	0	0	1	1
24	1	1	1	1	0	0
25	1	1	0	0	0	0
26	1	1	2	2	1	1
Mean	**1.23**	**0.96**	**1.23**	**1.27**	**1.19**	**1.00**
STD	**1.48**	**1.09**	**1.05**	**1.29**	**1.04**	**0.78**
<1.75	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

Figure A.3: PSSUQ survey results - Last Six Questions.

A.2 Notifications Experiment 1 System Usability Survey

Figure A.4: SUS survey results - NoO WS Condition - Six Questions.

Participant number	I think that I would like to use this system frequently.	I found the system unnecessarily complex.	I thought the system was easy to use.	I think that I would need the support of a technical person to be able to use this system.	I found the various functions in this system were well integrated.	I thought there was too much inconsistency in this system.
1	4	4	4	4	4	4
2	1	2	2	4	2	1
3	3	4	3	3	2	4
4	3	3	3	3	3	2
5	4	4	4	4	4	4
6	4	4	4	4	4	4
7	4	3	4	3	4	3
8	3	3	3	4	3	3
10	2	2	2	4	2	4
11	2	3	3	1	3	1
12	4	4	4	4	4	4
13	2	3	3	4	3	4
14	3	3	3	4	3	4
15	3	2	3	3	3	3
16	4	3	4	2	4	3
17	3	4	4	4	4	4
18	4	4	4	3	3	2
19	2	2	3	4	2	3
20	4	0	3	4	3	4
22	1	3	2	3	3	3
23	2	2	2	2	3	3
24	3	4	4	3	3	4
25	4	4	4	4	4	4
26	1	2	2	4	1	1
AVG	**2.92**	**3.00**	**3.21**	**3.42**	**3.08**	**3.17**
STD	**1.04**	**1.00**	**0.76**	**0.81**	**0.81**	**1.03**
SE	**0.21**	**0.20**	**0.16**	**0.17**	**0.17**	**0.21**

Participant number	I would imagine that most people would learn to use this system very quickly.	I found the system very cumbersome to use.	I felt very confident using the system.	I needed to learn a lot of things before I could get going with this system.	**Sum**	**Score**
1	4	4	4	4	**40**	**100**
2	1	3	1	3	**20**	**50**
3	4	4	3	3	**33**	**82.5**
4	3	3	3	1	**27**	**67.5**
5	4	4	4	4	**40**	**100**
6	4	4	4	4	**40**	**100**
7	4	3	4	2	**34**	**85**
8	3	4	4	3	**33**	**82.5**
10	3	3	4	4	**30**	**75**
11	4	4	3	3	**27**	**67.5**
12	4	4	4	4	**40**	**100**
13	4	4	4	4	**35**	**87.5**
14	3	4	3	3	**33**	**82.5**
15	3	3	3	3	**29**	**72.5**
16	4	3	4	2	**33**	**82.5**
17	4	4	4	4	**39**	**97.5**
18	4	2	4	4	**34**	**85**
19	4	3	2	4	**29**	**72.5**
20	3	1	3	1	**26**	**65**
22	3	3	3	2	**26**	**65**
23	3	1	1	2	**21**	**52.5**
24	4	4	3	4	**36**	**90**
25	4	4	4	4	**40**	**100**
26	3	2	3	4	**23**	**57.5**
AVG	**3.50**	**3.25**	**3.29**	**3.17**		
STD	**0.71**	**0.92**	**0.89**	**0.99**		
SE	**0.14**	**0.19**	**0.18**	**0.20**		

Figure A.5: SUS survey results - NoO WS Condition - Four Questions and Totals.

Figure A.6: SUS survey results - NoV Condition - Six Questions.

Participant number	I think that I would like to use this system frequently.	I found the system unnecessarily complex.	I thought the system was easy to use.	I think that I would need the support of a technical person to be able to use this system.	I found the various functions in this system were well integrated.
1	4	4	4	4	4
2	3	3	4	4	3
3	4	4	3	3	3
4	3	4	3	3	2
5	1	3	3	4	1
6	4	4	4	4	4
7	4	3	4	3	4
8	2	3	3	3	4
10	1	2	2	4	2
11	2	3	3	1	3
12	2	3	3	4	3
13	2	4	4	4	4
14	3	4	4	4	4
15	3	2	3	3	3
16	4	4	4	2	4
17	1	2	1	4	4
18	4	3	1	4	4
19	3	3	4	4	3
20	0	2	2	0	3
22	3	3	3	3	3
23	2	1	2	3	3
24	1	1	2	4	2
25	4	4	4	4	4
26	3	3	4	4	3
AVG	**2.63**	**3.00**	**3.08**	**3.33**	**3.21**
STD	**1.18**	**0.91**	**0.95**	**1.03**	**0.82**
SE	**0.24**	**0.19**	**0.19**	**0.21**	**0.17**

Figure A.7: SUS survey results - NoV Condition - Four Questions and Totals.

Participant number	I thought there was too much inconsistency in this system.	I would imagine that most people would learn to use this system very quickly.	I found the system very cumbersome to use.	I felt very confident using the system.	I needed to learn a lot of things before I could get going with this system.	**Sum**	**Score**
1	4	4	4	4	4	**40**	**100**
2	3	1	2	4	4	**31**	**77.5**
3	3	4	4	3	3	**34**	**85**
4	3	2	1	2	1	**24**	**60**
5	1	3	2	3	4	**25**	**62.5**
6	4	4	4	4	4	**40**	**100**
7	3	3	3	4	3	**34**	**85**
8	3	2	4	4	4	**32**	**80**
10	4	3	2	2	4	**26**	**65**
11	3	4	4	3	3	**29**	**72.5**
12	3	4	2	2	4	**30**	**75**
13	4	4	4	4	4	**38**	**95**
14	4	4	3	4	4	**38**	**95**
15	3	3	2	3	3	**28**	**70**
16	4	4	4	4	1	**35**	**87.5**
17	4	1	3	2	4	**26**	**65**
18	4	4	2	4	2	**32**	**80**
19	3	4	3	3	4	**34**	**85**
20	4	1	2	4	0	**18**	**45**
22	3	3	3	3	3	**30**	**75**
23	3	3	2	2	2	**23**	**57.5**
24	4	4	1	2	4	**25**	**62.5**
25	4	4	4	4	4	**40**	**100**
26	2	4	3	4	2	**32**	**80**
AVG	**3.33**	**3.21**	**2.83**	**3.25**	**3.13**		
STD	**0.75**	**1.04**	**0.99**	**0.83**	**1.17**		
SE	**0.15**	**0.21**	**0.20**	**0.17**	**0.24**		

Figure A.8: SUS survey results - NoO WA Condition - Six Questions.

Participant number	I think that I would like to use this system frequently.	I found the system unnecessarily complex.	I thought the system was easy to use.	I think that I would need the support of a technical person to be able to use this system.	I found the various functions in this system were well integrated.	I thought there was too much inconsistency in this system.
1	2	3	4	4	4	4
2	1	3	2	4	1	1
3	3	3	3	3	2	4
4	3	3	3	3	2	3
5	2	2	3	4	3	3
6	3	4	4	4	3	4
7	4	3	4	3	4	2
8	3	1	3	2	3	3
10	0	1	1	4	2	2
11	3	3	3	1	3	1
12	0	1	1	4	1	2
13	2	4	3	4	3	2
14	3	4	4	4	2	1
15	2	1	3	4	3	3
16	4	3	4	1	4	1
17	1	2	1	4	4	4
18	2	3	4	4	2	1
19	1	3	2	4	3	3
20	0	0	0	0	0	4
22	1	1	2	4	2	3
23	3	3	3	3	4	4
24	4	4	4	4	4	4
25	4	4	4	4	4	4
26	1	0	1	3	3	2
AVG	**2.17**	**2.46**	**2.75**	**3.29**	**2.75**	**2.71**
STD	**1.28**	**1.26**	**1.20**	**1.14**	**1.09**	**1.14**
SE	**0.26**	**0.26**	**0.24**	**0.23**	**0.22**	**0.23**

Figure A.9: SUS survey results - NoO WA Condition - Four Questions and Totals.

Participant number	I would imagine that most people would learn to use this system very quickly.	I found the system very cumbersome to use.	I felt very confident using the system.	I needed to learn a lot of things before I could get going with this system.	**Sum**	**Score**
1	3	4	4	3	**35**	**87.5**
2	2	0	2	3	**19**	**47.5**
3	1	1	1	3	**24**	**60**
4	3	3	3	1	**27**	**67.5**
5	3	2	3	4	**29**	**72.5**
6	4	4	4	4	**38**	**95**
7	3	1	4	1	**29**	**72.5**
8	3	3	3	2	**26**	**65**
10	2	1	1	4	**18**	**45**
11	4	4	3	0	**25**	**62.5**
12	3	0	1	4	**17**	**42.5**
13	4	4	4	4	**34**	**85**
14	3	4	3	4	**32**	**80**
15	3	1	3	3	**26**	**65**
16	4	4	4	2	**31**	**77.5**
17	2	2	3	4	**27**	**67.5**
18	2	2	1	1	**22**	**55**
19	4	2	2	3	**27**	**67.5**
20	0	0	0	4	**8**	**20**
22	3	1	3	3	**23**	**57.5**
23	3	2	3	3	**31**	**77.5**
24	4	4	4	4	**40**	**100**
25	4	0	4	3	**35**	**87.5**
26	3	0	2	2	**17**	**42.5**
AVG	**2.92**	**2.04**	**2.71**	**2.88**		
STD	**1.00**	**1.51**	**1.17**	**1.17**		
SE	**0.20**	**0.31**	**0.24**	**0.24**		

Figure A.10: SUS survey results - Control Condition - Six Questions.

Participant number	I think that I would like to use this system frequently.	I found the system unnecessarily complex.	I thought the system was easy to use.	I think that I would need the support of a technical person to be able to use this system.	I found the various functions in this system were well integrated.	I thought there was too much inconsistency in this system.
1	3	3	4	4	4	4
2	2	3	3	4	4	2
3	3	3	1	1	1	1
4	3	3	3	3	3	3
5	1	4	3	4	2	4
6	4	4	4	4	4	4
7	0	3	1	2	2	2
8	3	1	3	3	4	3
10	0	4	3	4	2	4
11	2	3	3	1	3	1
12	3	4	4	4	3	4
13	2	3	3	4	3	4
14	1	4	4	4	4	3
15	3	3	2	4	1	3
16	4	4	3	1	4	4
17	4	4	4	4	4	4
18	2	4	2	3	4	4
19	2	4	4	3	3	4
20	1	2	3	1	2	4
22	3	3	3	3	3	3
23	2	1	2	2	1	3
24	3	3	2	4	4	4
25	4	4	4	4	4	4
26	2	2	3	3	3	3
AVG	**2.38**	**3.17**	**2.96**	**3.08**	**3.00**	**3.29**
STD	**1.15**	**0.90**	**0.89**	**1.11**	**1.04**	**0.93**
SE	**0.23**	**0.18**	**0.18**	**0.23**	**0.21**	**0.19**

Participant number	I would imagine that most people would learn to use this system very quickly.	I found the system very cumbersome to use.	I felt very confident using the system.	I needed to learn a lot of things before I could get going with this system.	**Sum**	**Score**
1	4	4	4	4	**38**	**95**
2	3	2	3	1	**27**	**67.5**
3	3	1	1	1	**16**	**40**
4	3	4	3	1	**29**	**72.5**
5	4	3	3	4	**32**	**80**
6	4	4	4	4	**40**	**100**
7	1	4	1	3	**19**	**47.5**
8	3	4	3	2	**29**	**72.5**
10	3	1	1	4	**26**	**65**
11	4	4	3	2	**26**	**65**
12	4	4	4	4	**38**	**95**
13	4	3	3	4	**33**	**82.5**
14	3	4	3	4	**34**	**85**
15	3	3	3	4	**29**	**72.5**
16	4	4	4	2	**34**	**85**
17	4	4	4	4	**40**	**100**
18	3	2	4	0	**28**	**70**
19	4	3	3	2	**32**	**80**
20	0	4	2	1	**20**	**50**
22	3	3	3	3	**30**	**75**
23	1	2	2	1	**17**	**42.5**
24	3	4	3	4	**34**	**85**
25	4	4	4	4	**40**	**100**
26	4	2	3	4	**29**	**72.5**
AVG	**3.17**	**3.21**	**2.96**	**2.79**		
STD	**1.07**	**1.00**	**0.93**	**1.35**		
SE	**0.22**	**0.20**	**0.19**	**0.28**		

Figure A.11: SUS survey results - Control Condition - Four Questions and Totals.