DISSERTATION

INVESTIGATING NOTIFICATION PRESENTATION AND INTERACTION IN AUGMENTED REALITY

Submitted by

Lucas Plabst

Department of Computer Science Colorado State University

& Faculty of Mathematics and Computer Science University of Würzburg

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2025

And by Cotutelle agreement

For the Degree of Doctor rerum naturalium (Dr. rer. nat.)

University of Würzburg

Würzburg, Germany

Doctoral Committee:

Advisor: Francisco Ortega
Advisor: Florian Niebling

Mohammed Arefin
Hortensia Soto
Bianca Trinkenreich

ABSTRACT

INVESTIGATING NOTIFICATION PRESENTATION AND INTERACTION IN AUGMENTED REALITY

This thesis explores optimal notification display and interaction techniques in augmented reality (AR) environments. Through six experiments, we examined notification positioning, modality, interaction methods, real-world smartglasses usage, and task resumption strategies for interruptions. Our findings provide insights into designing effective AR notification systems while balancing usability and intrusiveness. First, we investigated notification placement in 3D AR environments using a memory task, finding that bottom-center positioning in the field-of-view (FOV) is ideal for general tasks, while world-anchored notifications suit stationary tasks. Next, we explored notification location and modality in a simulated cooking task, revealing that multimodal notifications placed on task-relevant objects improve user performance. We further analyzed notification interaction methods, determining that a body-attached notification list is the preferred option for users and that touch interaction outperforms gaze and voice input. In a follow-up study, we assessed multimodal interaction, noting its potential compared to unimodal interaction, but also highlighting reliability concerns due to modality fusion challenges. Beyond controlled environments, we conducted a five-day field study with smartglasses, demonstrating the benefits of hands-free, real-time notifications but also highlighting barriers such as social acceptability and attentional costs. Lastly, we designed an AR-based task resumption aid, showing that visual cues at the point of interruption significantly reduce resumption lag, potentially mitigating the disruptive effects of notifications. As AR devices become more pervasive, notification systems must be thoughtfully designed to balance intrusiveness with utility. This research informs future AR notification strategies, emphasizing adaptability to dynamic real-world contexts.

GERMAN ABSTRACT

BENACHRICHTIGUNGSDARSTELLUNG UND -INTERAKTION IN AUGMENTED REALITY

Diese Arbeit erforscht die optimale Anzeige von Benachrichtigungen und Interaktionstechniken in Augmented Reality (AR) Umgebungen. In sechs Experimenten untersuchten wir die Positionierung von Benachrichtigungen, die Modalität, Interaktionsmethoden, die Verwendung von Smartglasses im Alltag und Strategien zur Wiederaufnahme von Aufgaben nach Unterbrechungen. Unsere Ergebnisse geben Einblicke in die Gestaltung effektiver AR-Benachrichtigungssysteme, die ein Gleichgewicht zwischen Benutzerfreundlichkeit und Aufdringlichkeit herstellen.

Zunächst untersuchten wir die Platzierung von Benachrichtigungen in 3D-AR-Umgebungen anhand eines Gedächtnis-Kartenspiels und fanden heraus, dass die Positionierung unten in der Mitte im Blickfelds des Nutzers (FOV) ideal für allgemeine Aufgaben ist, während in der Welt verankerte Benachrichtigungen für stationäre Aufgaben geeignet sind. Als Nächstes untersuchten wir die Platzierung und Modalität von Benachrichtigungen in einer simulierten Küche und stellten fest, dass multimodale Benachrichtigungen, die auf aufgabenrelevanten Objekten platziert wurden, die Leistung der Nutzer verbessern. Wir analysierten außerdem die Interaktionsmethoden und Anzeigen für Benachrichtigungen und stellten fest, dass eine am Körper befestigte Benachrichtigungsliste von den Nutzern geschätzt wird und dass die Interaktion durch Berührung die Blick- und Spracheingabe übertrifft. In einer Folgestudie untersuchten wir die multimodale Interaktion und stellten fest, dass sie im Vergleich zur unimodalen Interaktion über ein größeres Potenzial verfügt, aber auch Bedenken hinsichtlich der Zuverlässigkeit aufgrund der Probleme bei der Modalitätenfusion aufwirft. Außerhalb kontrollierter Umgebungen führten wir eine fünftägige Feldstudie mit Smartglasses durch, in der wir die Vorteile freihändiger Echtzeit-Benachrichtigungen aufzeigten, aber auch Hindernisse wie soziale Akzeptanz und Aufmerksamkeitskosten hervorhoben. Schließlich

haben wir eine AR-basierte Hilfe zur Wiederaufnahme von Aufgaben entwickelt und gezeigt, dass visuelle Hinweise am Punkt der Unterbrechung die Verzögerung bei der Wiederaufnahme signifikant reduzieren und so die störenden Auswirkungen von Benachrichtigungen abmildern können.

Da AR-Geräte immer allgegenwärtiger werden, müssen Benachrichtigungssysteme sorgfältig entwickelt werden, um ein Gleichgewicht zwischen Aufdringlichkeit und Nutzen herzustellen. Diese Forschung dient als Grundlage für zukünftige AR-Benachrichtigungsstrategien, wobei die Anpassungsfähigkeit an dynamische reale Kontexte im Vordergrund steht.

ACKNOWLEDGEMENTS

I want to express my deepest thanks to all the people who made this journey possible and who helped me to complete it. First and foremost I would like to thank my wonderful family. I really could not have done this without you and your support. Thank you to my wonderful wife Lena, for being on this journey with me and always having my back. Thank you to my parents, for always supporting my endeavours. Thank you to my sister, for always being there for me.

This work would not have been possible without my incredible advisors Florian Niebling and Francisco R. Ortega. Both of you believed in me and encouraged me to pursue research further. I legitimately think that with other advisors, I would not have completed my PhD. So thank you for your time, knowledge, and patience.

I would also like to thank Sebastian Oberdörfer for mentoring me starting with my bachelor's thesis in 2017, all the way to now my dissertation. Thank you so much for your help with all of this. Thanks also go out to Marc Latoschik, who greatly fostered my interest in HCI and XR and who helped me find my internship abroad with Francisco, paving the way for this bi-national PhD.

Next, I would like to thank my friends for their patience and support all this time. A huge thank you goes out especially to *mindestmenschen*, you know who you are. There is a surprisingly large overlap between my friends and my colleagues, so I would like to next thank my amazing colleagues at both the HCI Group in Würzburg and the NUILab in Fort Collins. Being surrounded by such brilliant and helpful people has made my PhD so much easier, on a personal and academic level. Going to work and being happy is shaped tremendously by the people you surround yourself with there, and I can confidently say that seeing everyone was something I looked forward to in the mornings, no matter which continent. My thanks especially go to Brendan Kelley and Aditya Raikwar, for making us feel right at home after coming to Colorado and for many great experiences, whether it be hackathons or conferences.

Lastly, I want to thank my PhD committee, my reviewers, and my co-authors, thank you for your expertise and time.

DEDICATION

This work is lovingly dedicated to my family. To my parents Bettina and Rainer, without you literally none of this would have been possible. Thank you so much for your unwavering support and belief in me. So many opportunities in life I have only been able to take because of you. Know that I will be eternally grateful. To my sister Lucy, for being a beacon of positivity and for always being someone I knew I could depend on. To my childhood cat, Hobbes, who was a better person than most people and helped me through tough times. To my cats Clover and Franky, whose cuddliness and playfulness make the world right. Last but certainly not least, this work is dedicated to my wonderful, loving wife Lena. Starting a PhD in the middle of a global pandemic, and then moving to another continent for the second half, are immense challenges, but we braved these bitter storms together. Thank you for your encouragement, patience, and unconditional love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Figure 1.1: A cluttered AR view from the dystopian film “Hyper-reality” by Keiichi Matsuda, 2016. Permission for use was granted by original author.

With the rapid advancements in Augmented Reality (AR) technology, the prospect of humans primarily using AR for computing is becoming increasingly plausible. Daily computing often involves notifications for various events. These notifications can significantly impact users’ attention and responsiveness, as they compel users to react to new information, potentially disrupting their focus on a task. Previous studies, such as those by Czerwinski et al. [1], or the work of Bailey et al. [2], have demonstrated the disruptive effects of notifications and interruptions on ongoing computing tasks.

Mitigating these disruptions requires careful consideration of factors like timing, placement, and visualization methods of notifications. Users’ decisions on how to react are influenced by their perception of the importance and urgency of the notification. Research by McFarlane et al. [3] highlights the need to balance these factors to minimize interruptions.

While research on the presentation and interaction with notifications has been extensive on desktop and mobile systems, there has been considerably less focus on Virtual Reality (VR) and AR head-mounted displays (HMDs). Extended Reality (XR) is a growing market, and a future where a lot of daily computing is performed on a headset is plausible. However, there are many potential problems that these devices still face [4], so extensive research has and still needs to go into the usability and the user experience (UX) of these systems, including notifications that can appear anytime and potentially severely distract the user. XR designer and filmmaker Keiichi Matsuda [5] has illustrated a potential dystopian version of such a future in his film "Hyper-Reality" (see Figure 1.1).

1.1 Overview

This dissertation makes several key contributions to the study of notifications in AR:

- **Notification Position and Modality** – Investigating how the spatial placement and the modality of notifications affects user experience, attention management, and task performance.

- **Notification Interaction and Display Methods** – Analyzing the effectiveness of different unimodal and multimodal interaction techniques as well as display methods and their impact on user experience, task performance and perception in AR environments.

- **In-the-wild AR Notifications** — Exploring what the benefits and pitfalls of AR notifications are over the span of five days with real notifications outside a lab experiment.

- **Task Resumption Aid using AR** — Showing how AR can be used to assist users with task resumption after interruptions.

During this work, the term XR will be used as an umbrella term encompassing the different display types described in Milgram's & Kishino's work [6], which include AR, VR, Augmented

Virtuality (AV), and Mixed Reality (MR). For this thesis, we will refer to VR as "an approach that uses displays, tracking and other technologies to immerse the user in a virtual environment" and AR as an "approach that uses displays, tracking, and other technologies to enhance (augment) the user's view of a real-world environment with synthetic objects or information", and MR as "a set of approaches, including both VR and AR, in which real and virtual information is mixed in different combinations" according to the definitions by Bowman et al. [7]. This definition does not explicitly mention AV but includes it as a subset of MR.

Research is at its core a collaborative effort. While the research and writing of the text around the published reprinted papers were completed by the author of this dissertation, parts of the text will describe achievements and research efforts in the plural form “we”. Images used in this thesis are used with the creator's permission, which is found in the appendix or created by the author using assets in Microsoft PowerPoint.

The organization of the dissertation is described next. **Chapter 2** introduces the term "Notification" and discusses what their purpose and impact are. **Chapter 3** introduces work into Augmented Reality User Interfaces(UI). It discusses how AR UIs have been designed and how to interact with them. **Chapter 4** focuses on research efforts looking into notifications in XR. It shows different presentations of notifications, including modality and visuals, as well as interaction techniques. **Chapter 5** shows ways how notifications can be presented in AR. **Chapter 6** focuses on the interaction techniques with notifications. **Chapter 7** proposes research about long-term AR notifications outside a laboratory setting. **Chapter 8** describes techniques that can be used to reduce task resumption lag after interruptions, using AR cues. **Chapter 9** discusses the findings of the previous chapters and possible limitations. Finally, **Chapter 10** concludes this thesis with a summary of the contents.

Chapter 2

Notifications

A notification has been defined as a visual cue, auditory signal, or haptic alert generated by an application or service that relays information to a user outside the current focus of attention [8]. Notifications are also very closely related to cues. While there is considerable overlap, they are not the same thing. Cues (especially visual) are typically used to guide attention within a certain context, designed to integrate into a certain application unobtrusively, but rarely guide attention to something outside the current user's goals. Cues also usually spatially relate to whatever they are supposed to guide. Notifications, however, are a more explicit way of information delivery, as they are frequently interruptive and draw the user's focus away from their current task. Notifications also usually contain the information in itself and do not guide towards the information like cues. Therefore, while cues and notifications may use the same channels (e.g., haptic, visual, audio), their intent differs; therefore, the design philosophy around them might differ. We will explain this distinction, which we use for this thesis, with an example: Imagine wearing an AR HMD at home. You are wearing the headset and are watching a movie on it, not hearing the doorbell ring. Limiting this example to visual channels, a cue could be an arrow pointing towards the door to guide your attention, whereas a notification could be a text pop-up that tells you that someone just rang the doorbell. The cue unobtrusively guides your attention to the door, whereas the notification tells you the information directly. One place where notifications are deployed heavily is personal computing. Smartphones have quickly become the primary device that people use to access the internet and perform computing tasks, and are potentially one of if not the quickest spreading technology of all time [9]. Part of this smartphone use is notifications, ranging from phone calls and alarms to social media applications and ads. While it is difficult to know an exact number, some studies have found that their participants receive around 80 notifications per day, with some receiving up to 200 [10], [11]. Looking at only college students, Lee et al. [12] found some receiving over 400 daily notifications, mostly relating to messaging. They also found that receiving the notifications led to prolonged use of the apps, causing the corresponding notifications, showing that notifications may lead to excessive smartphone use. Measuring with an even younger population, 12 to 17-year-olds, Radesky et al. [13] found a median of 237 daily notifications, with maximums of over 4500 per day.

2.1 Categorizing notifications

To gain a more in-depth understanding of notifications, we first need to understand the key characteristics of notifications better, as when going by the definition, notifications are simply any kind of alert. According to current user experience design guidelines [14], [15] notifications may be categorized in one of two categories: **actionable notifications**, where the notification is followed by a user action, or **informational notifications**, whose aim is to pass information to the user. Actionable notifications may include but are not limited to phone calls, timers, or text-messaging notifications containing information that requires a response. Informational notifications may include but are not limited to system status updates or notifications about world news or stock prices. According to Jakob Nielsen's usability heuristics [16], "Visibility of System Status" is a highly important aspect of a system's usability, as users should be kept up to date about what the system is doing. Notifications can be an essential tool in informing the user of such changes. When designing notifications, the goal of the notifications heavily influences its presentation. McCrickard et al. [17] introduce the IRC Characterization Framework, a design model of user end goals for notifications based on three critical parameters:

- **Interruption:** This parameter is defined as the event prompting a user to switch their attention from their current task to the notification. It considers the timing, frequency, and context in which notifications are delivered. The goal is to minimize unnecessary interruptions that can negatively impact task performance and user experience, and to select the suitability of the interruption depending on context.
- **Reaction:** This parameter describes how promptly and accurately users can respond to notifications. This involves evaluating the user's ability to notice and act upon the notification.

- **Comprehension:** Comprehension measures the user's understanding of the notification content. High comprehension is crucial for users to make informed decisions based on the notification.

In this model, they assign values of high (1) or low (0) to each of the three parameters, to then select the appropriate notification type. They propose the model as a unified view of notification systems, that allows an improved usability evaluation process to emerge and that can help researchers and designers when designing such systems.

However, notifications are often considered secondary to an ongoing task and are therefore interruptive. Because of this they are frequently deferred to a later time [18], requiring a certain kind of triage and prioritization. According to Mehrotra et al. [19] based on the work by Clark et al. [20], there are four ways a user can choose to attend to a notification or interruption:

1. Handle it immediately.
2. Acknowledge it and agree to handle it later.
3. Decline to handle it (explicitly refusing to handle it).
4. Withdraw it (implicitly refusing to handle it).

Based on these ways of handling notifications, it is essential to understand which kind of notifications should be handled in specific ways. Pielot et al. [18] conducted an extensive study of real-world notifications by collecting smartphone notification data from 278 people over 10 days, resulting in a dataset of 794,525 notifications. These notifications were then analyzed and categorized into five groups: **Messaging**, **Group Messages**, **Email**, **Social**, and **Non-social**. To study how users attend to these notifications, they define five stages for notifications: **Shown**, **Seen**, **Checked**, **Consumed**, and **Removed**. In their analysis [18], they found that all notification types are seen in the same time frame. Still, notifications in the messaging or group-message category are consistently attended the most and also seen faster than other categories. Within those categories, they found that direct messages were being attended even faster than group messages.

They also had the highest conversion rate, meaning the highest rate of the corresponding app being opened after sending a notification. While these two categories were seen faster and had the largest portion of attendances, the time until it was attended, did not differ between the categories, except for emails, which were attended slower than the others. Lastly, non-social notifications were also attended to rather quickly; however, they were mostly removed without opening the app or clicking on the notification, highlighting their primarily informational purpose. When looking at the frequency of notifications, they found that social notifications are by far the most frequent notifications received on average in a day. This study leads to some interesting conclusions: First, messaging notifications seem to be the most important for users, as they were checked the most and comprised the most significant part of notifications received. Second, as the time to attend did not differ between categories, we can conclude that the notification category was not the deciding factor when deciding to participate in a notification, but rather external factors.

2.2 The impact of notifications

In a survey conducted in 2015 in the US, half of the surveyed adults reported that they were annoyed by their phone in the last week before the survey [21]. Given the number of notifications received daily, notifications might play a part in this annoyance, as it has been shown that users react negatively to notifications in many scenarios [22], [23]. Of those notifications, it is a reasonable assumption that not all notifications arrive at an appropriate time for the user, when they have time to react to them or want to react to them. However, even if a person receiving a notification does not engage with it at all, receiving a notification can significantly decrease the performance of an attention-demanding task, as has been shown by Stothart et al. [24].In their experiment, they recruited 212 undergraduates who were then asked to perform a sustained attention task [25]. During the second block of the task, half of the participants received either phone calls or text messages to the phone number they provided at the beginning. Participants who had received notifications were likelier to make an error in the second part of the experiment, even if they did not interact with their phones at all.

To get a more profound understanding of not just the short-term effects of notifications, but also of their long-term impact, and to understand notifications in a non-laboratory study better, Kushlev et al. [26] recruited 221 participants and assigned them to one of two groups: One group that had to maximize interruptions and notifications on their phone for a week and then minimize them the following week. The other group did the same but in an inverse order. Minimizing was done by turning off any notification indicators like haptics, sound, and visual notifications, as well as keeping their phone out of sight for most of the day. On the other hand, maximizing was achieved by enabling those indicators as well as keeping their phones within reach at all times. They found that people who reported more interruptions exhibited significantly higher levels of inattention and hyperactivity, traits usually found in Attention-Deficit-Hyperactivity-Disorder (ADHD), which are often associated with poor work/school performance and issues with social skills. They also reported lower productivity and psychological well-being. Warnock et al. [27] specifically looked at the modalities of notifications, to understand the different aspects of notification disruptiveness. In their experiment, they had participants play the game "Concentration" (also known as Memory) on a computer, during which they would receive notifications on the screen. These notifications either prompted the participants to press a button on the desk (target notifications) or to ignore the notification (distractor notifications). The modalities for the notifications used were: Text, Icons, Abstract visuals, Voice, Earcon, Auditory Icon, Tactile, and Smell. They found that the task performance suffered, no matter if the notifications were target or distractor notifications. However, if they did not receive any target notifications but only distractor notifications, the error rate suffered even more. This highlights that even notifications that are not relevant and can be ignored cause a significant increase in error rate while performing a task. As for the modality, they found that tactile and smell notifications caused a significantly longer pause than the other modalities when receiving a notification in that modality.

All the work presented showed a negative impact of notifications on task performance. The simple solution seems to be to simply turn off notifications or at least their indicators, such as sound or haptics, to help concentration. People simply wouldn't know if they had received a notification, and would just see it the next time they use their phone anyway. It would be easy to assume that this is the case since if something isn't doing anything it can't be distracting you, right? Unfortunately, not receiving notifications at all could also lower concentration and performance. Pielot et al. [28] asked 30 participants to turn off all notifications on all devices they used in their day-to-day lives, like smartphones or computers, for 24 hours. Participants reacted very differently to the experiment, with some reporting that they felt more productive at work or more relaxed at home. Other participants reported that they felt anxious that they were potentially missing important things but not receiving notifications for them, leading to more phone use as they constantly kept checking their apps. They also reported feeling disconnected from their social group, which is supported by the fact that most notifications on their phones are messaging-related [18]. None of the participants reported considering keeping notifications turned off, but 22 reported that they wished to manage notifications more consciously after the experiment, by either disabling some sources of notifications or using the do-not-disturb mode more often. Contacting the participants 2 years after the study, 13 of 22 reported that they were still managing their notifications after the experiment. Similar results were seen in an experiment by Iqbal et al. [29], where they asked participants to turn off email notifications at work for one week, after monitoring their usage for a week before. They found that 17 out of 18 participants wanted to turn notifications back on, even though they felt they were more productive without them. However, participants stated that they instead spent more time checking for emails or needing to catch up on numerous missed emails. Based on these findings, we can see that notifications have value for people, since almost all participants in these studies chose to rather receive notifications and be distracted sometimes, than not have them at all.

2.3 Managing notification delivery

To mitigate the attentional cost and other issues that notifications bring with them, without affecting their usefulness, researchers have studied several ways notifications can be delivered more optimally. One of these approaches is called *bounded deferral* [30], where notifications are delayed until an opportune moment or are delivered at specific times throughout the day. Fischer et al. [31] found in their work that notifications are attended to and dealt with significantly faster if they received them after finishing an episode of mobile interaction compared to random times. Supporting this finding, Pielot et al. [32] developed a machine-learning model to display notifications at such opportune moments and found that predicting engagement in users led to a significantly higher conversion rate from notifications to app content. Another approach that has been proposed is batching notifications together and delivering those batches of notifications several times throughout the day. Fitz et al. [33] conducted an experiment, in which participants would receive their notifications either as usual (control), in three batches throughout the day, or not at all. They found that participants with the batched notifications felt more attentive, productive, and in a better mood compared to the control group. As was established in other studies, participants who did not receive notifications at all felt feelings of missing out, did not report higher levels of concentration, and even reported higher levels of anxiety than control.

Based on bounded deferral, there has also been work done in developing context-aware notification delivery systems, like the work by Park et al. [34]. They developed a social context-aware notification system, that delivered notifications at breakpoints in social settings. Testing this system with ten friend groups of three people, they found that delivering notifications at breakpoints in the conversation reduced interruptions in an almost unnoticeable manner. This can be an important use case for notification deferral, as phones have been an issue in social settings for some time. The phenomenon of *phubbing* has been described as snubbing someone in a social setting by concentrating on one's mobile phone. Chotpitayasunondh & Douglas [35] investigated the perception of phubbing in social settings by showing different videos of dialogues to participants, where one conversation partner either phubbed never, sometimes, or most of the time. Asking the participants to put themselves in the role of the other person, they found that phubbing negatively impacted perceived communication quality and relationship satisfaction.

This leads us to the use of AR head-mounted displays for notification delivery as a means of improving notifications. In a future where these headsets might become pervasive, notifications will play a key role in the overall user experience and these headsets might improve interruptions of notifications. Orlosky et al. [36] found that the use of an HMD led to a higher contextual awareness compared to using a smartphone, which could be beneficial in many situations, for example pedestrian use, where it has been found that using a smartphone can lead to higher accident risk when crossing the street [37]. However, as notifications move beyond handheld devices into AR, new challenges may emerge. Unlike smartphones, which require explicit interaction and are often used in discrete sessions, AR may integrate notifications directly into the user's field of view, enabling more seamless but potentially more intrusive forms of delivery. The shift from screen-based to immersive, always-available notifications raises questions about how attention management, cognitive load, and user experience translate across these mediums. Effects found in desktop and smartphone research about interruption through notifications may be even stronger in AR, as they become harder to ignore. On the other hand, AR also provides new opportunities to deliver notifications, such as placing them in optimal positions to be informative but not intrusive. Since a large amount of AR headsets feature cameras and possess some understanding of the environment, they may be useful in predicting opportune moments for notification delivery. One way to potentially reduce the impact of notification delivery in AR will also be presented in chapter 8.

Chapter 3

Augmented Reality User Interfaces

While the emergence and popularity of AR devices are fairly recent, the technology behind it isn't, predating even the mobile phone, which was first released in 1973 by Motorola [38]. The first AR-HMD however was built in 1968, before the term AR was even coined. To build a prototype of his "Ultimate Display" [39], a system that could show a computer-generated environment that could be indistinguishable from reality, Sutherland et al. built the "head-mounted three-dimensional display", more colloquially called the "Sword of Damocles" [40]. The name comes from the fact that because this device was too heavy to wear comfortably, it was suspended from the ceiling, constantly hanging over the user's head, akin to the sword from Greek legend. It used a mechanical tracking system to adjust visuals based on the user's movements and incorporated a see-through optical display, which could display simple wire-frame 3D graphics. Despite its limitations, it laid the foundation for modern VR and AR technologies. Caudell et al. [41] were the first ones to use the term AR in their proposal about equipping Boeing factory workers with see-through HMDs that they could then use to aid the workers in the airplane assembly.

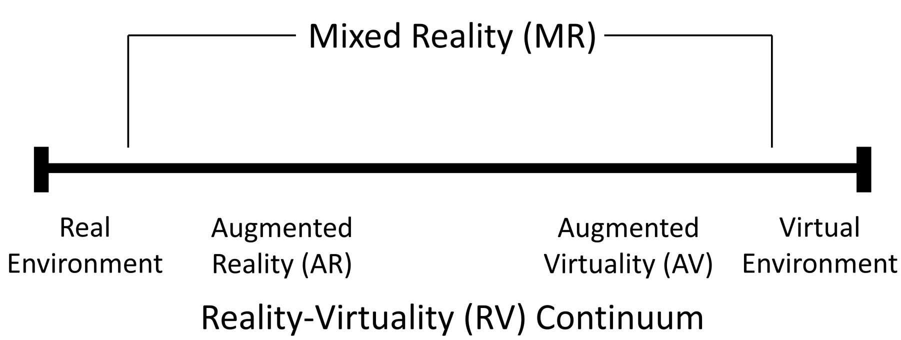


Figure 3.1: Reality–virtuality continuum.

The term AR was then firmly established by Milgram & Kishino [6] in their attempt to create a taxonomy for all systems that merge real and virtual environments. They proposed the "Reality-

Virtuality (RV) Continuum" (see Figure 3.1), which conceptualizes the spectrum between fully real and fully virtual environments (VE), providing a framework for understanding XR experiences. The VE represents the conceptual extreme of full virtuality, where VR is just a technology for experiencing this virtual environment. Between these extremes lie intermediate states such as AR—where digital elements overlay the real world—and AV—where real-world elements are incorporated into a predominantly virtual space. AR has then been further described by Azuma et al. [42], defining AR as a system that fulfills three key characteristics:

- **Combines real and virtual elements:** AR integrates computer-generated objects or information with the real-world environment, seamlessly blending the two.
- **Interactive in real time:** The system responds immediately to user input or changes in the environment, allowing for dynamic interaction.
- **Registered in 3D:** Virtual objects are accurately aligned with the real-world environment in three-dimensional space, maintaining spatial coherence.

However, some systems may be considered AR that do not technically fulfill the requirements of this definition, nor the classifications of the RV-continuum. Devices like Google Glass [43] can display information relating to the real world as a heads-up display, which is not interactive in real-time, or registered in 3D. The definitions also do not include technologies like spatial audio, as they focus on visual display technologies. Therefore, we have chosen to use the looser definition of AR set forth by Bowman et al. [7], which assumes that AR is an "approach that uses displays, tracking, and other technologies to enhance (augment) the user's view of a real-world environment with synthetic objects or information".

Looking at ways how this augmentation can be implemented, Billinghurst et al.[44] present three ways visual information can be displayed in AR environments, more specifically on AR headsets (see Figure 3.2):

- **Head-stabilized:** Information is fixed to the user's viewpoint.
- **Body-stabilized:** Information is fixed to the user's body.
- **World-stabilized:** Information is fixed to real-world locations.

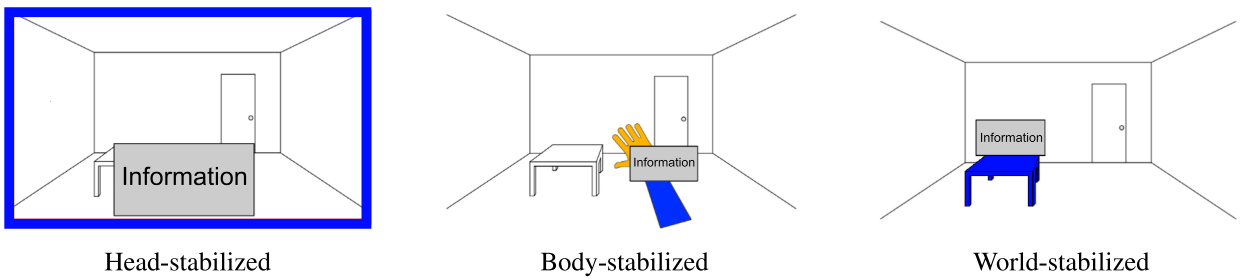


Figure 3.2: AR content placement options.

Feiner et al. [45] have also proposed a similar taxonomy in their work, where they rendered desktop windows on an AR HMD in each one of these placement methods. For accessing brief and glanceable information on an AR headset, Lu et al. [46] devised a system enabling users to summon information like weather updates or sports scores through short glances at the periphery. They propose three different interfaces to achieve this:

- An eye-glance interface, where content is placed display-stabilized in the four cardinal directions of the periphery and can be accessed by simply looking at it.
- A head-glance interface, where content is placed body-stabilized outside the four cardinal directions of the periphery and can be accessed by moving the head in one of the four directions.
- A gaze-summon interface, where content is placed display-stabilized outside the four cardinal directions of the periphery and can be accessed by gazing at the peripheral direction for 0.5s.

They evaluated this system in an experiment, where participants were asked to walk along a path during which they would need to answer questions related to the information in the glanceable interface. They were either asked direct questions, or instructed to notify the experimenter immediately if some information changed. They found that walking task performance was not affected by the interfaces, that question response time was slower with the gaze-summon interface, and that eye-gaze performed best when monitoring information since it was always in view. Based on this, they recommend the eye-gaze interface when information needs to be monitored or frequently accessed in a short time.

As described previously, interaction with a smartphone has been shown to decrease the quality of social interactions. To potentially address this issue, Davari et al. [47] built a socially context-aware AR interface, based on the previously mentioned Glanceable AR system [46]. This system adapts between two common contexts: a static scenario where the user is alone, and a dynamic social context. In the solo context, the system recognizes the current object of focus in the real world and places virtual content at the eye level above the object of focus with high transparency, therefore not occluding it. To highlight virtual content, it can simply be changed to opaque. As this would not be appropriate in a social setting, the system adapts to the topic of the conversation and the number of conversation partners, by highlighting relevant apps based on speech recognition and by not placing content over the partner's faces engaged in conversation with the user. This system was evaluated against an identical but context-unaware system and a mobile phone in solo and group settings. Their experiment shows the significant advantages of the context-aware system over a context-unaware system but especially the mobile phone in UX and preference, information access speed, and intrusiveness in both solo and group scenarios. Adapting content by changing the opacity has been shown by this experiment to be beneficial for guiding attention, but special care has to be taken. Unlike traditional displays, the background of digital content on optical-see-through (OST) AR headsets cannot be completely freely chosen, as it largely depends on the background of the surroundings. Especially text legibility decreases or increases greatly depending on the contrast ratio [48] and with OST AR headsets, the contrast ratio cannot always be kept constant. Because of this changing background problem, Debernadis et al. [49] evaluated the presentation of text on different AR headsets. They found that presenting text on a dark blue billboard with white font seems to be a good combination for indoor AR applications, regardless of device or background. Their findings also include that OST headsets generally produce higher text legibility compared to video-see-through headsets, where the text may be blurred because of the digitized background. This work is consistent with the findings from Gabbard et al. [50], who found a billboard style with a semi-transparent background to be the most well-suited text display style across several outdoor background textures. Their findings show that task performance in AR is significantly affected by background texture and text style, highlighting the importance of careful text design. Looking into reading on AR HMDs, Rzayev et al. [51] found that the position of text in AR also matters. In their experiment, presenting text in the top-right position (display-stabilized) led to significantly lower text comprehension and higher workload than bottom-center and center. Reading while walking also led to a slower reading speed than when sitting.

In addition to poor contrast, the increased display of virtual content in AR can negatively impact task performance due to clutter [52]; therefore, it should be ensured not to overload the user with UI elements, especially notifications. This is also supported by Miller's law, which strongly suggests that only seven pieces of information (±2) can be held in working memory at the same time [53].

3.1 Interaction

Looking closer at interaction techniques, the use of hands, eyes, and voice are all extensively studied techniques for XR devices [54]. The Microsoft Hololens line of devices is the most used OST AR headset in the experiments using such a class of device described in this research exam. The main way of interacting with these systems is through the use of hand gestures. While these gestures largely only use one hand, Chaconas et al. [55] evaluated the use of two-handed gestures for the Hololens. A significant limitation of the device in question is the limited hand-tracking area. In their experiment, they compared a one-handed scaling and rotation gesture, provided by the device's system, to four two-handed manipulation gestures. They found no differences in performance, however, their novel bi-manual technique was preferred by users, showing the feasibility and benefit of two-handed interfaces on these devices.

In their study on general interactions in AR, Wang et al. [56] explored the fusion of various input modalities. They discovered that a combination of Gaze, Gesture, and Speech yielded the most efficient results, although users preferred the combination of Gesture and Speech. Conversely, unimodal interactions using solely Gaze or Gestures generally underperformed compared to the multimodal options. Lee et al. [57] similarly investigated multimodal versus unimodal inputs, noting that while participants favored multimodal input over sole Speech or Gaze, they did not demonstrate objective performance improvements with it. Parisay et al. [58] proposed a system that integrates gaze for target selection with a mouth click to confirm manipulation. Their evaluation revealed that their system achieved faster task completion times, quicker movement times, and lower workload compared to sole voice recognition, but that it was surpassed by a gaze and dwell approach.

Pfeuffer et al. have also investigated interaction using gaze, but using a pinch gesture instead of a dwell time. Targets are indicated [7] using Eye-Gaze and are confirmed using a Pinch gesture. In an informal evaluation, they saw that while participants liked the concept, reliability issues with the eye tracker and the gesture recognition were prevalent. They suggest using a very explicit gesture operation, like the one used by the Hololens. In follow-up work, they suggest various considerations when using this technique [59], such as using eyes only for indication and the pinch gesture only for confirmation. They also note that this interaction technique requires users to "unlearn" previous behavioral patterns, as they might be used to moving their hand for interaction but can look away from potential interaction targets.

Chapter 4

Notifications in XR

After having presented several XR UI considerations and systems, focusing mostly on AR, it is now time to get into the core topic: **How can notifications work in XR?** Ghosh et al. [60] explored interruptions and notifications in VR, employing various modalities such as visuals, haptics, and audio. In their work, they first conducted a co-design experiment with XR developers and designers to come up with different notification ideas in VR.

For audio, they propose *Spatial Sound* (positioning sound in 3D space corresponding to the notification's location), *Skeuomorphism* (using metaphorical designs, like a door opening sound to signify someone entering a space), and *Gradual Intensification* (increasing sound gradually to indicate rising urgency). For the haptic channel, they propose *Skeuomorphism* (employing metaphorical vibrations, such as feet vibrating to symbolize footsteps), *Directionality* (directing vibrations toward the notification's direction), and *Gradual Intensification* (gradually increasing vibration to signify increasing urgency). Regarding the visual channel, their suggestions include *3D Representation* (displaying notifications in 3D space instead of attaching them to the display), *Skeuomorphism* (using metaphorical representations, like a clock face for alarms), *3D Pop-ups* (placing notification blocks in 3D space, on walls or surfaces), and *Flash and Pause* (temporarily halting immersion in critical situations to let users reorient in real-world space).

Collecting themes from this experiment, they designed five notification types (see Figure 4.1) that they then evaluated in a VR experiment. Each notification type had a visual, auditory, and haptic channel. Participants in their experiment had to continuously locate a ball in the environment, navigate to it, and bring it to a specified location while receiving notifications. As a secondary task, participants were asked to acknowledge notifications by pressing a button on the controller. Participants went through each notification type and each combination of modality components, resulting in 30 conditions. Besides objective measures like reaction time, they also asked participants four 7-point Likert scale questions about notifications (see Table 4.1).

Figure 4.1: Five notification types with three channels each. Permission for use was granted by the original author.

Attribute	**Explanation**
Noticeability	How easy or difficult is it to notice the notification?
Understandability	Once you notice the notification, how easy or difficult is it to understand what it stands for?
Perceived Urgency	What level of urgency does the notification convey?
Perceived Intrusiveness	How much of a hindrance was the notification to the overall VR experience?

Table 4.1: Definitions of Notification Attributes

These questions have since been used by other researchers studying XR notifications to quantify perceptual questions about notifications. Based on their findings, several design recommendations have been made for notifications:

1) Visual notifications should be supplemented with audio directing attention to the notification, and stand out from the environment to increase noticeability. 2) Controllers are often moved outside of view, so visual notifications on controllers should be enhanced by haptics to guide attention. 3) Visual search should be reduced by binaural audio or a dedicated place for notifications. 4) Sudden appearances should be reduced. 5) Familiar metaphors should be used. 6) Notifications should be interactable and able to be switched to being the primary task.

To enhance this first design recommendation [60], making notifications stand out from the environment, Zenner et al. [61] built a framework for notifications in VR. They integrate notifications into the environment to not break immersion and presence, two concepts in immersive virtual environments that describe the feeling of actually being in the virtual world [62]. In their work, they proposed sending notifications from outside of VR to the immersed user in VR by using a smartphone app to send messages with a specified priority. The VR environment could respond differently depending on the priority by drawing more attention to the notification depending on priority, integrating into the environment instead of appearing as an abstract pop-up (see Figure 4.2). While this work is demonstrated in VR, it is thinkable to use this for notifications in AR, placing notifications in locations where they make sense for the current task and environment to make them seem more natural and real.

Low priority: Notification is put on unobtrusive display in a virtual supermarket

Medium priority: Lights are dimmed and notification is highlighted

High priority: User is blocked from the environment to stop and notice the notification

Figure 4.2: Three immersive notifications based on priority, Zenner et al. [61]. Permission for use was granted by the original author.

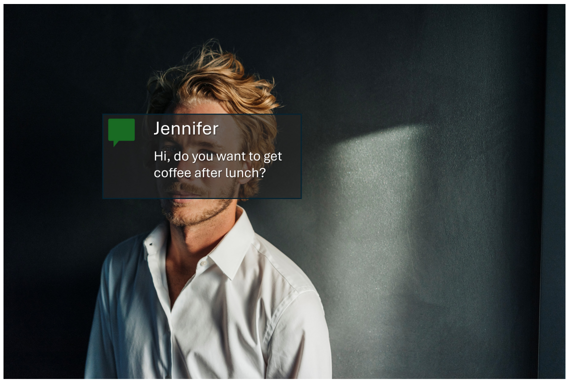
4.1 Position

Speaking of placing notifications in locations where they make sense: where **do** they make sense? One of the most important questions about how to design XR notifications is where to even put them. Whereas smartphone or desktop notifications are traditionally displayed on a 2D display, moving into a 3D environment brings with it ample ways of displaying UI elements, including notifications. So moving into notification position in XR, Rzayev et al. [63] studied the positioning of notifications in VR. In their experiment, they looked at three tasks in three different

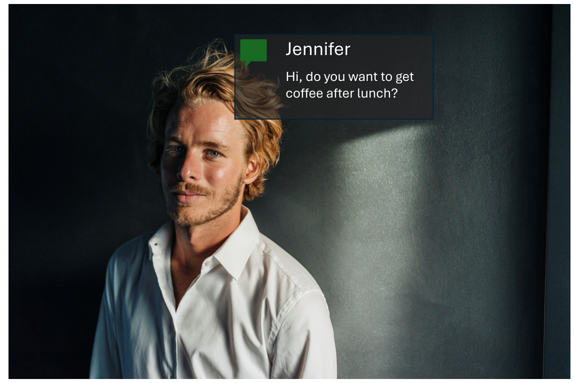
VR environments and displayed notifications in one of four positions: 1) Attached to the headset (*Heads-up-display*); 2) Attached to the controller (*On-Body)*; 3) *Floating* in front of the user and 4) Attached to a wall in the environment (*In-situ*). They found that *Heads-up-display* notifications resulted in the lowest missed notification count and the shortest response time. However, while having the highest noticeability, participants preferred this location the least, while also rating it the highest on intrusiveness. Placing them *in-situ* on the other hand, led to the highest count of missed notifications, the highest reaction time, and the lowest noticeability and urgency. Users preferred *floating* and *on-body* over the others, and thought they were the most understandable. From this, they recommend using a *heads-up-display* for urgent notifications only, *in-situ* for unimportant notifications only, and *floating* or *on-body* for general notifications. Since there was not one clear best option, they show that it is important to give the user control over the priority of notifications and change the presentation style accordingly.

Since it has been shown that phone notifications can disrupt conversational settings [35], Rzayev et al. continued their research into notification placement by shifting to a dialogue setting and using AR instead of VR [64]. They invited participants and paired them up in teams, that would then engage in 7-minute long discussions, which the researchers provided topics for. In each team, one participant (*receiver*) would be given a Hololens 2 AR OST HMD [65], which would start displaying notifications after two minutes of the conversation. The notifications were displayed in four different positions (see Figure 4.3): a) Center position on the *observer* (dialogue partner); b) Top-right position on the *observer*; c) Center position locked to display and; d) Top-right position locked to display.

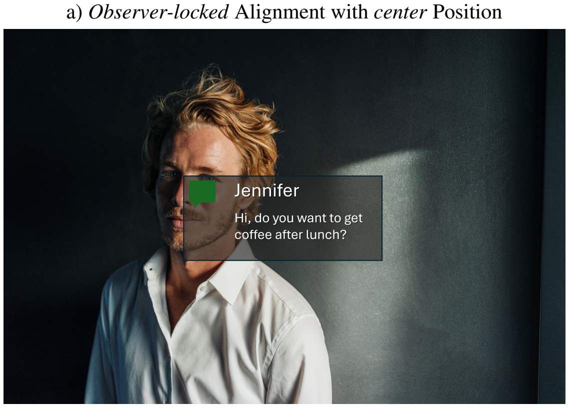
Their findings suggest that general notifications should be placed on the *observer*, as that placement was found to be the least intrusive while also having the lowest task load for reading the notification while still participating in the conversation, especially when placed in the top-right. For more important notifications, placement in the center and attached to the *receiver* showed to be the most urgent, but also the most intrusive, however, participants stated that they would accept



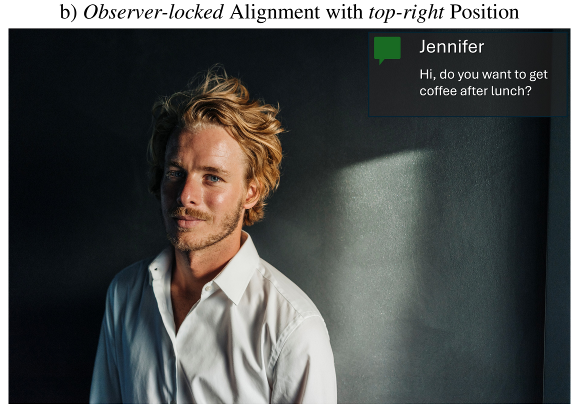
a) *Observer-locked* Alignment with *center* Position



b) *Observer-locked* Alignment with *top-right* Position



c) *Receiver-locked* Alignment with *center* Position



d) *Receiver-locked* Alignment with *top-right* Position

Figure 4.3: Four positions used in the experiment [64]

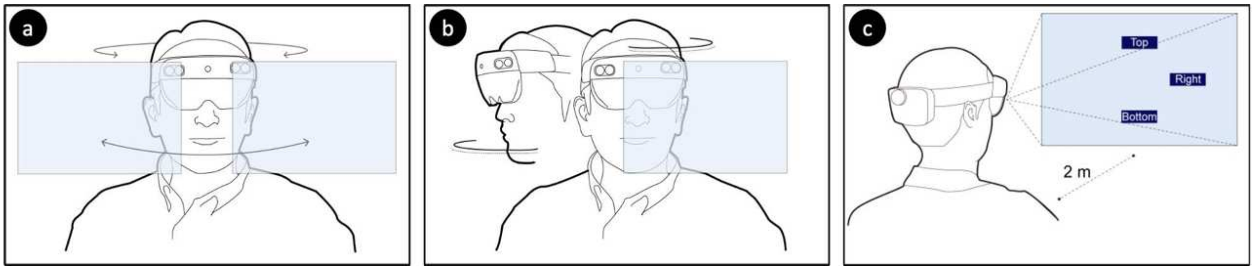


Figure 4.4: Attachment methods a) Head-stabilized b) body-stabilized; c) Notification positions used [66]. Permission for use was granted by the original author.

the negative impact to ensure receiving important notifications. The majority of both *observers* and *receivers* expressed that they would prefer notifications on an AR headset, rather than a phone.

Lee et al. [66] studied the position of AR notifications during a walking task. For their experiment, they chose body-stabilized and head-stabilized methods, with three different positions in each method (top, right, bottom, see Figure 4.4). The notifications they used contained instructions to perform specific hand gestures when read, as this would reflect not only if a notification was seen, but also if the content was understood. Furthermore, notifications were split into high priority (which had to be immediately attended) and low priority (which could be handled later). Participants were tasked with following a specified path in the room while receiving notifications on a Hololens 2. They found that a bottom position in the display-fixed method resulted in a significantly higher noticeability and comprehension for notifications compared to the right and top positions, while also showing lower task load. Comparing the two attachment methods, they found that display-stabilized notifications are perceived as having higher urgency and are reacted to faster, and therefore recommend using this for high-priority notifications. In contrast, body-stabilization was perceived as less urgent, but also affected the walking speed significantly less when receiving a notification, causing the authors to recommend using body-stabilized notifications for unimportant notifications. It is worth noting that all the positions used in the described studies about position fell within the classification by Billinghurst et al. [44], highlighting the strength of their taxonomy.

Figure 4.5: A "caught" butterfly showing a notification [67]

4.2 Presentation & Modality

The work shown so far has largely focused on the location of visual notifications based on text, so we will now report on more ways how notifications can be presented in XR. Taking the research out of a lab and "into the wild" [68], Lucero et al. [67] studied AR notifications and a dedicated attendance device in an urban navigation task, using the Epson Moverio BT-100 HMD [69]. They employed a minimal user interface and a discrete thumb touchpad device for controlling notifications, where display-stabilized butterflies would fly across the display indicating a new notification (see Figure 4.5). These could then be opened by swiping in the direction of the flight path on the thumb touchpad, upon which the notification text would be displayed. The notification would stay open as long as the touchpad is held after swiping (catching the butterfly) and would close as soon as the finger is removed from the device (releasing the butterfly), causing the butterfly to continue on its path (meaning it could be reopened). Their findings indicated that participants faced minimal difficulties in managing notifications while being exposed to potential hazards in an urban environment.

Looking further into output modalities of notifications, Lazaro et al. [70] conducted an experiment where they designed notification indicators that were either visual-only (in the form of a bell icon in the periphery), auditory-only (where an alert sound was played), or multimodal (combining auditory with visual). Participants were given a Hololens 2 HMD and asked to perform a visual search and matching task, during which they received notifications. The content of the notification was placed outside the field of view to the left of their main task in all conditions, but participants were instructed to read the notification and then confirm they had read it by either verbalizing it or pushing a virtual button. They found that visual notifications were missed more than 50% of the time, whereas auditory notifications were missed 5%, and multimodal notifications were not missed at all. This difference was also reflected in the preference ranking, where auditory and multimodal were ranked significantly higher. However, task completion time was not affected, revealing that even though auditory and multimodal notifications were attended to more, this did not affect main task performance. When looking at the frequency of the confirmation action, they saw that participants used verbal confirmation more often than gesture action, suggesting that voice input was more intuitive than hand gestures.

Investigating a hand-summonable menu in AR, Pfeuffer et al. [71] created a menu that can be summoned by the users by looking at their palm. This interface also features a notification delivery system, where notifications are displayed in a vertical list in the menu. They opted not to include buttons on the notifications to display more information and used a two-step selection process for notification attendance. Context actions of notifications appear when the notifications are selected using Gaze + Pinch, but this results in more steps to attend a notification.

Taking the context of the notification into consideration, Faulhaber et al. [72] proposed three different notification visualizations to reflect the priority of the notification. Participants would receive notifications from a public transport app with low priority (informing the user but needing no reaction), medium priority (informing and maybe needing reaction, but not (time) critical), and high priority (informing and needing immediate action). All notifications were presented in the form of icons in the peripheral vision and differed in animation, choice of icon, and color. Low-priority notifications were represented with a green icon that would slowly fade in to not attract too much attention, medium-priority notifications with a yellow icon that moved up and down slowly to attract some attention, and high-priority with a red icon that flashed in and out while changing color to attract immediate attention. Using the Nreal Light [73] HMD, participants were instructed to perform the n-back task [74], which is a visual attention and concentration task, during which they would receive notifications with different priority. They found that the low-priority visualization resulted in significantly higher task reaction time, while also ranked as lower noticeability. They were consequently also ranked as least distracting. The performance of the main task was not affected by the priorities.

Lastly, looking at iconography in AR notifications, Janaka et al. [75] conducted a series of experiments to understand the use of icons in notifications. In their first two experiments, they found that users preferred having icons instead of text; however, no objective measurements were significantly different. In their third experiment, they found that if the users were allowed to choose icons themselves, therefore increasing familiarity, and the density of information represented by the icon was high, icons significantly outperformed text. While the first three experiments were conducted in research labs, the last experiment was conducted on university grounds, where participants had to walk 1.2 km while wearing an HMD, to more accurately reflect real-world conditions. In this experiment, they found that the lab results could be applied to a realistic setting, however because of outdoor background and lighting changes affecting legibility, the longer text notification became more noticeable than the icons.

To mitigate the attentional costs that come with receiving a notification, Janaka et al. [76] researched different animation techniques for the delivery of notifications in AR. They found that fading-in notifications for two seconds had the least impact on primary task performance, as users could prepare for the notification better than when they suddenly appeared or scrolled in from the side. However, the fade duration has to be considered cautiously, as a too-short duration can attract attention too quickly, and a too-long duration would distract and cause lower noticeability of the notification.

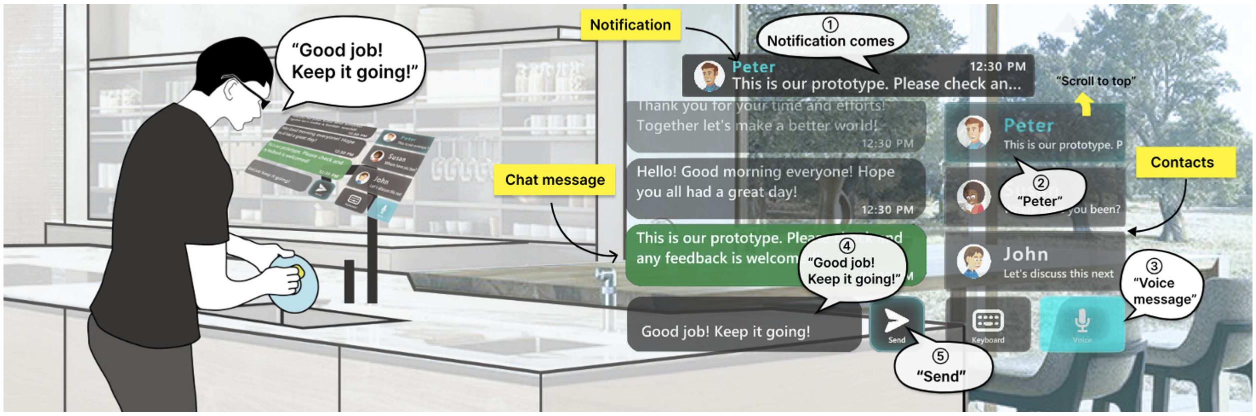


Figure 4.6: GlassMessaging: Messaging in AR [78].Permission for use was granted by the original author.

4.2.1 Attendance and Interaction

As we saw in the research about smartphone notifications, there are several ways that users can attend to notifications. We will now take a closer look at how this attendance can be designed in XR. The importance of this can be seen in some of the previously explained literature [60], [63], where even though the notifications in the experiments did not relate to the participants personally, they still expressed the wish to reply to the notifications or manually dismiss then when they feel they are distracting.

Kosch et al. [77] evaluated three AR notification selection techniques while riding a bicycle: gaze with a dwell time, gestures, and gaze for selection with a physical button press for confirmation. They found that gaze with dwell time resulted in the lowest error rates, but required more attention and focus, whereas the gaze with button approach was preferred by users and had the lowest task completion time. Since the predominant subject of notifications in smartphones is messages, Janaka et al. [78] have researched a messaging system for AR (see Figure 4.6). They first conducted an online survey to identify the most common two scenarios for messaging: On transportation and eating alone. Participants were then observed in these scenarios to identify messaging and multitasking behavior. Interactions with notifications in these scenarios were mostly viewing and replying to messages, and participants did mostly not open the messaging app, but rather looked at their notification interface to select which messages to respond to. They also identified common difficulties participants faced: 1) their hands were too busy with a different task to comfortably and reliably use their phones, and 2) there were frequent attention switches between their phone and their environment, which limited attention on both the primary task and their phone. Based on their observations, they iteratively designed and implemented a messaging application for Hololens 2, which they then evaluated against phone messaging in simulated day-to-day activities. They found that messaging using an AR-HMD showed better multitasking capabilities and faster response times than using a smartphone, and suggest the usage of HMDs as supplemental devices to the phone. They mostly attribute this to the hands-free operation modes, the wearable always accessible nature that showed notifications in any situation, and the multimodal input methods that could be adapted depending on the situation the user was currently in. However, participants encountered issues with the accuracy of the messages, since text entry was mostly done using voice input, which sometimes had issues understanding the correct dictation. Participants also mentioned that correcting misinputs was harder, as they would rather not use the midair keyboard. Participants also reported higher cognitive load, stating that it was challenging to get used to midair gestures and voice commands.

4.3 Contributions

The study of notifications in XR has explored various factors such as presentation, content, positioning, and interaction techniques. Prior research has established the importance of optimizing notifications to balance awareness and interruption, ensuring they remain informative without being overly disruptive. However, gaps remain in understanding notification positioning and interaction, particularly in generalized, foundational settings rather than highly specialized applications. AR encompasses a wide range of contexts, making it crucial to understand what works best for different scenarios. To investigate notification position and interaction in AR, we conducted our studies in a standing cooking environment and a sitting desk environment. While the tasks in these environments are specific, they serve as representative use cases for general AR applications where users must manage tasks while interacting with their surroundings. The settings involve multitasking, hand-occupied interactions, and dynamic environmental changes—challenges that are also present in industrial, medical, and everyday productivity applications. As such, our findings offer insights that extend beyond the tasks and apply to broader AR scenarios. Our work aims to understand how notifications should adapt to different scenarios rather than searching for a single universal solution.

One gap in the literature is the lack of a comprehensive investigation into the various positions of notifications with all ways of AR UI placement as outlined by Billinghurst et al. [44]. While some studies have examined specific positioning strategies, these have often been tailored to particular use cases, limiting their generalizability or not including all three ways of display. Our research addresses this gap by systematically evaluating notification positioning at a base level, providing a foundational understanding that can inform future studies and applications.

Similarly, interaction techniques in XR—such as touch, voice commands, gaze, or gestures—have been explored in prior work, but often in constrained or application-specific contexts or not using notifications. There remains a need to understand how different interaction modalities affect user experience in more varied settings. In our work, we investigate several commonly used uni-and multi-modal interaction techniques for use with notifications and present shortcomings and benefits of the techniques. These findings show what techniques could be used depending on the current context of the user and help in understanding when specific techniques should be deployed.

Lastly, to the best of our knowledge, there is no research on real notifications in XR, since notifications in the works presented are all artificially sent during the experiments. However, most notifications are messages directly to the user [18], which is not possible to truly replicate in a controlled experiment.

Chapter 5

Presentation of Notifications in AR

5.1 Position of Notifications

To address the problem of position for notifications, in our first experiment[1] we researched the impact of notification position on task performance and perception in an AR scenario [79] using the Hololens 2. We used a card game task that was either fully digital or with real playing cards and sent participants notifications throughout the game, which prompted users to press a specific button on a game controller. We assigned participants either the physical or digital cards and then had participants play the game four times for eight minutes each, once for each of the four notification placements we investigated. Those placements were:

- Top-right in the FOV (head-stabilized)
- Bottom-center in the FOV (head-stabilized)
- On the wrist (body-stabilized)
- On the table (world-stabilized)

[1]This work was published as Lucas Plabst, Sebastian Oberdörfer, Francisco R. 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*, 1–11. Online CA USA: ACM, 2022. HTTPS://DOI.ORG/10.1145/3565970.3567701.

Push the Red Button: Comparing Notification Placement with Augmented and Non-Augmented Tasks in AR

LUCAS PLABST, HCI Group, University of Würzburg, Germany
SEBASTIAN OBERDÖRFER, HCI Group, University of Würzburg, Germany
FRANCISCO ORTEGA, Computer Science & NUILAB, Colorado State University, USA
FLORIAN NIEBLING, University of Applied Sciences Fulda, Germany

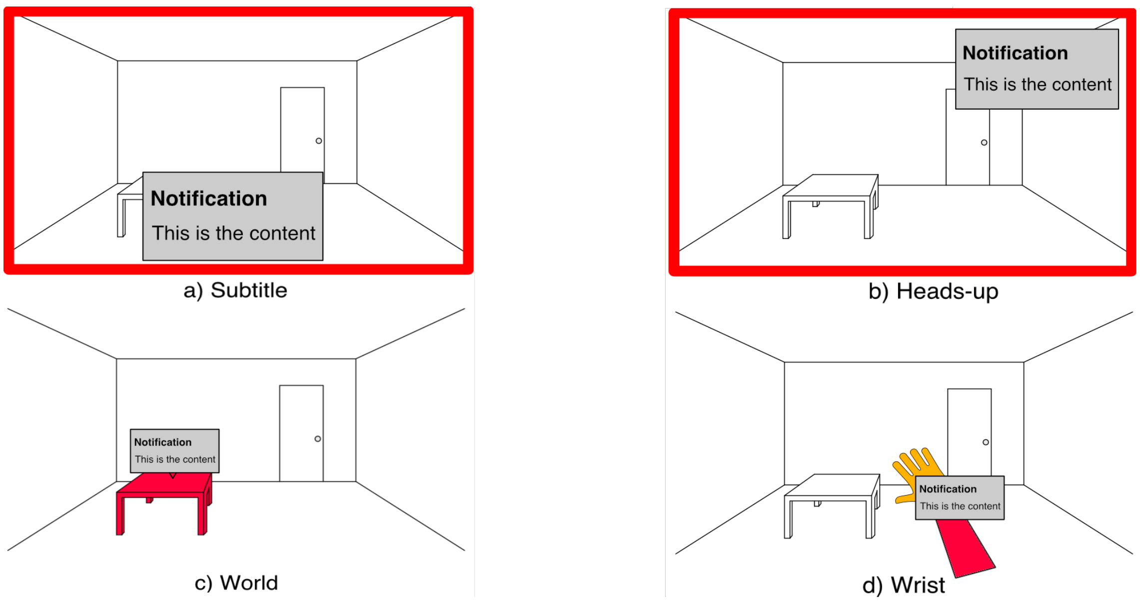


Fig. 1. Different notification placements (location is illustrated by the red color, a) and b) are presented in screen space) : Subtitle, Heads-up. World and Wrist

Visual notifications are omnipresent in applications ranging from smart phones to Virtual Reality (VR) and Augmented Reality (AR) systems. They are especially useful in applications where users performing a primary task have to be interrupted to react to external events. However, these notifications can cause disruptive effects on the performance of users concerning their currently executed primary task. Also, different notification placements have been shown to have an influence on response times, as well as e.g. on user perceived intrusiveness and disruptiveness.

We investigated the effects and impacts of four visual notification types in AR environments when the main task was performed (1) in AR and (2) the real world. We used subtitle, heads-up, world space, and user wrist as notification types. In a user study, we interrupted the execution of the main task with one of the AR notification types. When noticing a notification, users responded to it by completing a secondary task. We used a Memory card game as the main task and the pressing of a correctly colored button as the secondary task. Our findings suggest that notifications at a user's *wrist* are most suitable when other AR elements are present. Notifications displayed in the *World* are quick to notice and understand if the view direction of a user is known. *Heads-up* notifications in the corner of the field-of-view, as they are primarily used in smart glasses, performed significantly worse, especially compared to *Subtitle* placement. Hence, we recommend to use different notification types depending on the overall structure of an AR system.

CCS Concepts: • **Human-centered computing → User studies**; *Mixed / augmented reality*.

Additional Key Words and Phrases: augmented reality, notifications, attention

ACM Reference Format:
Lucas Plabst, Sebastian Oberdörfer, Francisco Ortega, and Florian Niebling. 2022. Push the Red Button: Comparing Notification Placement with Augmented and Non-Augmented Tasks in AR. In *Symposium on Spatial User Interaction (SUI '22), December 1–2, 2022, Online, CA, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3565970.3567701

1 INTRODUCTION

Visual notifications in AR can be employed to draw a user's attention away from their current main task, towards specific, potentially important events. This is especially significant in safety-critical environments such as control rooms, medical care, disaster response, or construction, where missed incidents can sometimes have fatal consequences. Nevertheless, interruptions have disruptive effects on the user's task performance and lead to a higher memory load of users at the time of interruption [3]. Also, different forms of presentations in mobile device applications have an influence on response time and the disruption perceived from a notification [27, 35]. Rzayev et al. show that placement has an impact on the perceived urgency and intrusiveness of visual notifications in AR [32]. In Virtual Reality, presentation and placement of notifications have also been shown to influence response time, noticeability, distraction, intrusiveness [33], as well as perceived disruptiveness [17]. Thus far, the effects of notifications and notification placement on a main task are well established for desktop, mobile applications and VR. In contrast, research mostly focused on the effects of AR-based notifications on the performance of real world tasks. This, however, leaves out situations in which AR also displays the main task in addition to the notifications. Here, the perception of notifications can be vastly different depending on the amount of virtual content displayed. An AR-based notification during a real world task can stand out more than during the purely virtual task, e.g., due to the vergence-accommodation conflict, latency, differing color appearance and depth perception, especially when using Optical See-Through devices.

In this work, we focus on the perception and notability of four different AR-based notifications displayed either during a real world only or an AR-based task: subtitle, heads-up, world space, and user wrist. In particular, we investigate which notification type is more suited when (1) the only virtual information are notifications and (2) the virtual information is used to display the task and the notifications. The main task consists of a card game known as *Concentration* or *Memory* [37], where users have to find matching pairs of cards that are initially laid out face down on a surface. In the real world condition, no virtual content besides the visual notifications is presented to the user. In the AR condition, the card game itself is performed in AR. In both conditions, users have to interrupt their main task to perform an additional activity, i.e., pressing the button mentioned in the notification. We report on task performance between the conditions concerning *e.g.* reaction time, missed notifications, and error rate in the interrupting activity.

Our main contributions are foundations for AR notification display location, including (1) understanding the ideal position for AR notifications, which includes the effect the position of a notification in AR has on the main task performance and the effect the position has on the perception of the notification; (2) how these effects change depending on whether the main task is of a physical or fully virtual nature.

2 RELATED WORK

2.1 Notifications

A notification has been defined as a visual cue, auditory signal, or haptic alert generated by an application or service that relays information to a user outside of the current focus of attention [19]. Notifications have become an essential part of our interaction with technology on a day-to-day basis, especially with the unprecedented rise of the smartphone [7]. While difficult to accurately pin down, some studies have found their participants to receive around 80 notifications per day, with some receiving up to 200 [1]. Given the amount, it is reasonable to assume that not all notifications always arrive at an appropriate time for the user, which is problematic as Stothart et al. [36] has shown, that receiving a notification can significantly decrease the performance of an attention-demanding task. Several researchers have tried to manage the attentional cost of notifications by approaches such as grouping many notifications together in small batches delivered multiple times throughout the day [9] or by developing context aware delivery systems [30]. However, not receiving notifications can lead to increased frustration and actually lower productivity [22]. Also, not every notification can even be delayed until a later time, for example phone call notifications or time-critical alerts like in safety-critical-systems need to be delivered regardless of opportune timing or context. Orlosky et al. [29] have shown that the use of a head-mounted display for notification delivery can lead to increased spatial awareness with minimal performance impact over the use of a smartphone.

2.2 Information Acquisition in 3D

When it comes to placing content in AR environments, there are three possibilities according to the classification of Billinghurst et al.[4]:

- Head-stabilized: Information is fixed to the user’s viewpoint.
- Body-stabilized: Information is fixed to the user’s body.
- World-stabilized: Information is fixed to real-world locations.

Rzayev et. al. [33] experimented with different notification positions in VR. They concluded that there was not a preferred notification placement for all contexts, as each position was perceived differently from the others, but rather, that position should depend on the context of the notification and the current task the user is performing. Also researching notifications in VR, Ghosh et al. [12] explored interruptions and notifications in VR with several modalities like haptic and audio and derived design guidelines based on their findings. To evaluate the perception of notifications they created several questions, which will also be used during the course of this work. Lu et al. [25] developed an interface for quickly accessing short information at the periphery of vision using different glancing methods, which could be employed for notifications as well. Chua et al. [5] investigated the display-position of a monocular head-mounted-display and how it affected the performance and usability in a dual-task scenario. They found that middle-right, top-center, and top-right are most suited when the center of vision is needed for the main task and when the secondary stimulus is not urgent. Middle-center and bottom-center positions were preferred when the secondary stimulus required high noticeability. Rzayev et al. [32] also looked at notification position in AR during social interactions and found that displaying notifications in the user’s field-of-view (FOV) was seen as favorable in social interactions. Participants could not agree on whether they preferred a center or top right position. Also based on a more casual day-to-day activity like talking to another person, Lucero et al. [26] developed and researched notifications on an AR-headset while walking and performing a pedestrian navigation task in a busy city center. They used a minimal UI and a discrete thumb touch-pad device to control notifications and found that participants had little issue with dealing with the notifications while being exposed to potential hazards in an urban environment. This might change with the increase in AR-content displayed, as it was shown that more virtual objects in an AR scene decreased task performance due to clutter [10].

2.3 Text in AR

Unlike traditional displays, the background of digital content on AR headsets can not be completely freely chosen, as it largely depends on the background of the surrounding. Especially text legibility decreases or increases greatly depending on the contrast ratio [24] and with AR headsets, the contrast ratio cannot always be kept constant. Debernadis et. al. [6] evaluated the presentation of text on different AR-headsets. They found that presenting text on a dark blue billboard with white text seems to be a good combination for indoor AR-applications, regardless of device or background. If the notification should also convey information through color, using the color as a background with white text is preferable. This was also corroborated by Gabbard et al. [11], who found a billboard-style with a semi-transparent background to be the most well-suited text display style across several outdoor background textures. Rzayev et al. [34] presented a study to evaluate reading text on an AR headset, looking especially at the positioning of text and presentation method (Rapid Serial Visual Presentation (RSVP) and line-by-line scrolling). The positions researched were center, upper right, and bottom-center. When text was displayed in the lower-center or center position, comprehension increased while perceived workload decreased, with participants preferring bottom-center for reading. Text in the top right was least favorable but might be suited for quick alarms or notifications, as longer reading caused eyestrain and reduced text comprehension.

3 EXPERIMENT

We conducted an experiment to determine if the position of a notification could affect primary and secondary task performance and to also examine if the perception of the notification changes depending on its location. In this study, participants were instructed to play a card game (primary task), during which they received notifications on an optical see-through AR headset, the Microsoft Hololens 2, to which they had to respond to (secondary task). The headset features a resolution of 1440x936 pixels per eye with a FOV of 43 degree horizontal, 29 vertical and 52 diagonal.

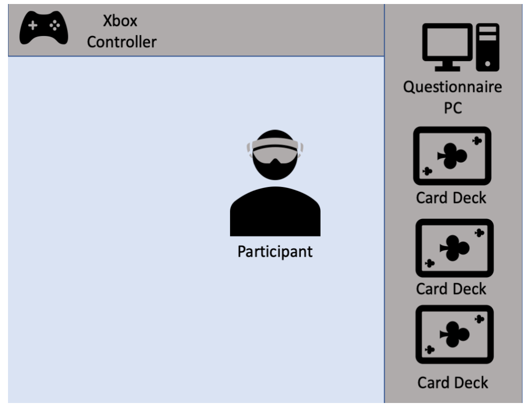
3.1 Design

In the experiment, we deployed four different notification Placements (see Figure 1) in an AR-environment with two different Task scenarios. This resulted in a mixed group design with two independent variables. The notification Placements consisted of a notification in the (1) top right *(heads-up-display)*, (2) bottom middle *(subtitle)* portion of the AR headset display, (3) projected on the *wrist*, and (4) situated above the Task in the *world*. The participants were exposed to all notification types during the experiment, i.e., type of notification was a within-subjects condition.

Cards were either all physical (real card task) or exclusively virtual AR playing cards (AR card task), depending on the Task scenario. This between-subject variable allowed for a comparison of whether more virtual content influences notification perception. Also, this approach enabled an investigation on the influence of the FOV of the AR device. The real card condition was not limited to the boundaries of the AR device's FOV and resembled a typical task that could be enhanced with AR information like assisting in safety-critical medical procedures [31]. In contrast, the digital card condition explored the effectiveness of notifications in a more AR-focused situation.

The Task for the experiment was a memory card-playing game, as a sustained attention task was needed for the experiment. Because this game requires a lot of recall ability, intrusive interruptions should have a large impact on the

(a) Card setup for the real-card condition.



(b) **Setup of the experiment room (approx. 3m wide x 3m long).**

(c) **Real Cards with notification. (Image has lower contrast due to capturing technique used by Hololens.**

Fig. 2. **Experiment setup**

performance, which makes this task suited for investigating the effects of the notifications. The game rules are also very simple so preexisting knowledge about the game should not be an issue.

Participants were given three playing-card decks (cut down to 15 pairs from 28) spread out face-down in a five by six grid each (see Figure 2a). The decks were kept small to allow displaying all cards within the FOV of the AR headset used. Both card types measured 64mm X 89mm. Two cards had to be flipped and discarded if their color (red or black) and value matched. If they did not match, they had to be returned face-down. This process was repeated until all cards of a deck had been discarded, and then the participant had to move on to the next deck. Users playing with the digital cards could use their right index finger to tap on a card, which would flip it. If the cards matched, the cards would automatically disappear, or in the case of a mismatch, be flipped face-down again.

While playing the memory game, notifications with different PLACEMENTS were shown to the participants in the AR-headset. Each experiment run lasted eight minutes and the notifications appeared every 50 seconds resulting in a total of nine notifications per run. Timings were kept constant in an attempt to reduce potential confounds. Timing could influence the perception and we want to research this in the future. Ending each experiment run after eight minutes ensured every participant got shown exactly the same amount of notifications. The amount of total cards was chosen to make sure no participant could finish within the given time, which would cause them to sit idle, most likely influencing the results.

Every notification contained an instruction, which the participant needed to perform. Each instruction was the pushing of a specific button on an Xbox game-controller located on the far end of the room. The controller was placed away from the participant as it would cause a more severe distraction from the main task because it required the participants to completely stop the main task in order to attend to the instruction, resembling for example the dismissal of a patient monitoring system alarm in emergency health care. The four face-buttons on the controller are colored (yellow, red, blue, green), so each instruction told the participant to press a specific colored button, which was chosen at random. Notifications were not able to be dismissed by the user but disappeared after five seconds on their own, no matter if a correct button was pressed. The Xbox controller was paired with the Hololens 2 using a Bluetooth connection and plugged into a USB charger to avoid power loss during the experiment. Because of the power cable, and also to discourage participants from taking the controller to their seat, the controller was taped down. The study had a duration of approximately 65 minutes per participant.

3.2 Notification Design

All notifications had a rectangular form, mimicking the alerts most commonly seen on mobile and desktop operating systems. They featured a bold title, which read "Notice" for all notifications and a text-body, containing the instruction to be carried out. To ensure high legibility of the text, a dark gray background was chosen, along with white text color, keeping in line with design recommendations by Microsoft [28] and Jankowski et al. [20]. The font size was set to 20pt, putting it within range of Microsoft's guidelines for text legibility in AR applications. All notifications automatically aligned to face the user, with the exception of the z-axis(roll), therefore ignoring head tilting. It was found that as focal switching distance increased in AR, eye fatigue increased while performance decreased [2], so the notifications are displayed at about the same distance as the cards.

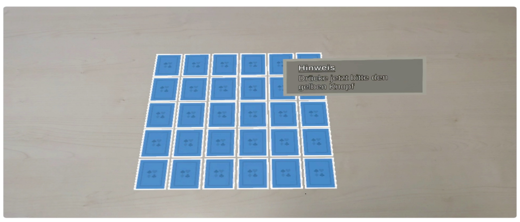
3.2.1 Heads-up. The notification types *Heads-up* and *Subtitle* are fixed to a specific position in the display of the headset and are therefore head-stabilized according to the categorization by Billinghurst and Kato [4]. As such they are visible regardless of position and orientation of the user and resemble traditional 3D user interfaces found in for example video games. Aside from the location in the display, both notifications are identical. The heads-up notification (see Figure 3a) is placed at the top-right border of the FOV, closely mimicking headsets such as Google Glass. Participants in Rzayev et al.'s [32] study disliked this position for longer reading but expressed that it would be well suited for short texts.

3.2.2 Subtitle. The subtitle notification (see Figure 3b) is placed at the bottom center border of the FOV, as suggested by Chua et al. [5] for dual-task scenarios that require high noticeability on the secondary stimuli. The bottom-center was chosen rather than middle-center, because we wanted to have the least visibility-impact on the main task, while also having good noticeability of the notification. Work by Rzayev et al. [34] also showed that participants preferred the bottom-center over the middle-center notification position when it came to reading text in AR. System notifications that are displayed by the Windows OS running on the Hololens 2 are displayed using the subtitle notification placement. With content placed close to the user, special attention needs to be paid to the vergence-accommodation-conflict [16]. The Hololens 2 display is fixed at an optical distance of approximately two meters away from the user, so Microsoft recommends placing content close to this point for extended interactions, with the optimal zone being one to five meters.

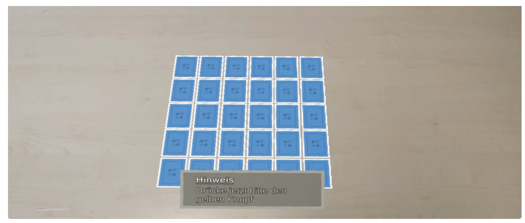
Both *Subtitle* and *Heads-up* notifications are displayed at a distance of one meter away from the user, in accordance with the comfort guidelines by Microsoft. The notifications do not move in depth and the Hololens 2 automatically calibrates the interpupillary distance (starting with OS Version 20H2 released Nov. 2020) , which both lessens the potential discomfort caused by the vergence-accommodation-conflict.

3.2.3 Wrist. The body-stabilized notifications (see Figure 3c) were positioned at the user’s right wrist, inspired by notifications that a user wearing a smartwatch might receive. Unlike a smartwatch, rotating the wrist did not change the position of the notification as it always remained centered above the wrist, independent of rotation.

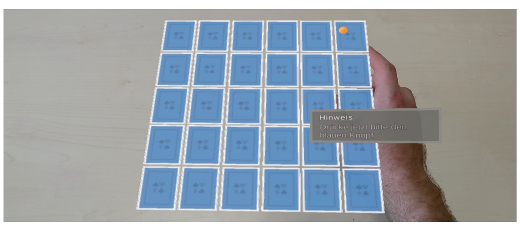
3.2.4 World. The world-stabilized notifications (see Figure 3d) were placed on the center top edge of the card deck, which the user is currently closest to, hovering slightly over the table to not cause any alignment issues with the real world or block sight of the cards. It was found that world-stabilized interfaces centrally and closely located to the task improved task completion time [18].



(a) **Heads-up**: Notification in top-right.



(b) **Subtitle**: Notification in bottom-center.



(c) **Wrist**: Notification on the users right wrist.

(d) **World**: Notification over the current task.

Fig. 3. **Different notification placements used in the experiment, here with the AR-Task.**

3.3 Implementation

To run the experiment, a prototype was developed which could periodically send different notifications and allowed the user to play a memory-card-game. The application was implemented using Unity Engine 2020.3.16f1 with the help of Microsoft’s Mixed Reality Toolkit (MRTK) v2.4.0. Aligning the digital playing cards with the table, as well as anchoring the world-notifications above the playing cards, was realized using Vuforia Image target recognition. Three tracking markers were generated using the ARMaker Tool [23] and printed on white paper with a size of 21cm x 21cm. Each deck had its own marker for tracking, as this reduced any shift that might have occurred by small registration errors. Every event (notification sent, button press, card flips and card matches) was logged by the application running on the Hololens 2 and saved to a text-file for analysis. Digital cards could be turned over by touching them with a small orange orb attached to the tip of the index finger on the right hand for additional visual feedback.

3.4 Measurements

We used the following measures to evaluate the notifications.

Performance. When a notification was sent and the participant subsequently reacted with pressing the corresponding button on the controller, it was measured as a *correct button press*. If a notification was sent and no button was pressed, the notification was counted as missed after 20 seconds of popping up. *Missed notification count* was measured as a dependent variable. In the event the button pressed did not match the instruction, it was counted as a *wrong button press*. Other measurements done during the experiment were the time after a notification was sent until a button was pressed (*reaction time*) and the number of *correct card matches*. As each experiment run is capped at eight minutes, the correct card matches are used as a measure of main task performance, instead of time to completion.

Usability and Task Load. After each round of the TASK, participants had to complete a set of questionnaires. To assess the overall usability of the notifications, the System Usability Scale (SUS) [13] was used, along with a NASA Task-Load-Index questionnaire (NASA-TLX) [15] to assess task load. Subscales of the NASA-TLX questionnaire were not weighted (Raw-TLX) as doing so does not seem to impact the results [14].

Perception. To gain a further understanding of the perception of notifications we used the questions from the work of Ghosh et al. [12] about noticeability, understandability, urgency and intrusiveness of notifications. In addition to the questionnaires, the participants were asked to rank each notification type by personal preference and give a short explanation of their ranking.

3.5 Procedure

Task and order of notification Placements were assigned to each participant at random, so everyone would play with either real or digital cards with every notification type. Also, lighting conditions were kept the same across all conditions and participants. After welcoming the participants, we asked them to read and sign experimental consent forms and fill out demographic information questionnaires. Following that, we explained the study procedure and explained how to correctly put on the Hololens 2. Participants had the rules of the game explained to them and were instructed to read the notifications and act upon the instructions contained in them. They were then given the headset and when the participant correctly put it on, we played a demo scene to show what kind of notification types the participant could expect and to ensure that each notification type was legible to the participant. The Hololens 2 automatically calibrates the display to the wearer's interpupillary distance after wearing it for about 30 seconds, so no separate calibration was done, as the process would automatically complete during the demo, ensuring optimal clarity for every participant. As soon as the demo finished, a blue text box was displayed, telling the participants that the experiment would start as soon as the text box disappeared. Also, over each tracking marker a purple cube was displayed to indicate if the marker was being correctly tracked. To set the different experiment conditions before each experiment run, the buttons on the Xbox Controller were used. When the participant confirmed that all markers were being tracked, the experiment was started by experiment leader by also pressing a button on the controller. The participant would then start playing the card game for eight minutes, with a pop-up text box notifying the participant of the end of the experiment when that time had passed. After each game had finished, the participant was instructed to take off the headset and complete a set of questionnaires. This procedure was then repeated (with the omission of the demo), so each participant would perform the Task a total of four times, once for every notification type. Participants had to sit down during the experiment and only got up from their chair to press the button on the controller.

The institutional review board of Human-Computer-Media Würzburg approved our ethics proposal for this study.

3.6 Participants

Participants were recruited from a pool of university students studying Human-Computer-Systems or Media-Communication. They are required to gather a certain amount of experiment participation hours for their coursework and were rewarded with 1.25 hours participation time for the experiment. In total, 40 participants were recruited (12 Male and 28 female). Age ranged from 19 to 30 years (M = 22.15, SD = 2.3). All either had normal or corrected-to-normal vision. Of those participants, all 40 stated that they used smartphones and the internet daily and 30 had either never used AR before or only in experiments (16 in experiments, 14 never). Additionally, 11 participants stated that they played video games somewhat regularly and 36 were right-handed. The data of one participant could not be evaluated at all. Two participants were missing the log files from the playing card game, but their questionnaires were still evaluated. This brings the total to 39 participants for the questionnaires and 37 for the task. All participants were fully vaccinated against COVID-19 and were required to show a negative test that was taken at most 24 hours before their participation. During the experiment participants wore medical gloves which were disposed of after the experiment.

4 RESULTS

We analyzed our data with RStudio in version 1.4.1106. To compare the means of the conditions for the measured factors, because our data did not meet the assumptions for ANOVA, we transformed it using the Aligned Rank Transform[38], before computing a non-parametric ANOVA on the transformed data. Pairwise tests were done using the ART-C Procedure [8].

Most notable, we found that *Wrist* notifications performed substantially different depending on *Task*, that *Heads-up*-performed worse overall than *Subtitle*- notifications and that *World* notifications performed the best overall in most measurements. *World* notifications were ranked as the most preferable, while *Heads-up* notifications were ranked the lowest. In the following we will go through the results of the experiment in detail, starting with the quantitative performance measurements like reaction time, then moving on to results gathered from the questionnaires and qualitative results. Table 1 provides a summary of the descriptive statistics.

4.1 Performance

Correctly Pressed Buttons. We found a significant two-way interaction between Placement and Task explaining the number of correctly pressed buttons, $F(3, 105) = 5.57$, $p = 0.001$. We found a significant main effect for the Placement $F(3, 105) = 3.88$, $p = 0.011$. However we found no significant effect for the Task overall $F(1, 35) = 1.7849$,$p = 0.190$. Contrast tests showed a significant difference between *Subtitle* and *World* $t(105) = 3.12, p = 0.012$, and *Subtitle* and *Wrist* $t(105) = 2.62, p = 0.049$, regardless of Task. Given the AR card condition, we found significant differences between *World* (6.47) and *Wrist* (8.37), $t(105) = 3.71, p = 0.008$ with respect to the number of correctly pressed buttons, (see Figure 4a). We also found a significant difference for *Wrist* depending on Task $t(129.8) = 3.44, p = 0.017$. When conducting difference of difference testing, we found that the difference between *Subtitle* and *World* $t(105) = 3.12, p = 0.012$, as well as *Subtitle* and *Wrist* $t(105) = 2.61, p = 0.049$, changed significantly depending on the Task.

Wrongly Pressed Buttons. We found no significant effect between Placement and Task explaining the number of wrongly pressed buttons, $F(3, 105) = 0.311$, $p = 0.817$. There was also no main effect regarding Placement or Task.

Table 1. Descriptive statistics; $N = 37$, real cards $n = 18$, AR cards $n = 19$. Values are $M(SD)$.

Scale	Overall	Heads-up	Subtitle	World	Wrist
Correctly Pressed Buttons (0-9)					
Real Cards	6.82 (2.8)	6.89 (2.54)	7.89 (2.27)	7.11 (2.03)	5.39 (3.70)
AR Cards	7.34 (2.36)	6.68 (2.79)	7.84 (2.01)	6.47 (2.44)	8.37 (1.67)
Wrongly Pressed Buttons (0-9)					
Real Cards	0.111 (0.358)	0.167 (0.383)	0.167 (0.514)	0.111 (0.323)	0.0 (0.0)
AR Cards	0.132 (0.411)	0.263 (0.562)	0.158 (0.501)	0.053 (0.229)	0.053 (0.229)
Missed Notifications (0-9)					
Real Cards	2.07 (2.84)	1.94 (2.58)	0.944 (2.29)	1.78 (2.02)	3.61 (3.70)
AR Cards	1.53 (2.28)	2.05 (2.72)	1.0 (1.92)	2.47 (2.29)	0.579 (1.68)
Reaction Time (in seconds)					
Real Cards	6.83 (1.58)	6.54 (0.92)	6.13 (0.84)	5.95 (0.56)	8.69 (1.88)
AR Cards	7 (1.68)	7.3 (1.72)	6.89 (1.65)	6.01 (1.08)	7.8 (1.75)
Correct Card Matches (0-45)					
Real Cards	23.4 (6.97)	23.4 (8.07)	24.7 (7.24)	24.3 (5.89)	21.3 (6.52)
AR Cards	24.6 (6.78)	21.4 (6.51)	24.7 (7.88)	28.5 (4.95)	23.8 (5.93)
SUS (0 - 100)					
Real Cards	71.1 (14.4)	67.2 (16.0)	75.1 (13.0)	71.7 (14.2)	70.3 (14.3)
AR Cards	67.9 (20.4)	62.5 (23.9)	70.6 (18.2)	70.1 (18.4)	68.4 (21.2)
NASA TLX (0-100)					
Real Cards	35.4 (15.1)	34.7 (15.5)	34.3 (14.4)	33.6 (14.7)	38.9 (16.4)
AR Cards	42.1 (17.7)	46.9 (19.7)	40.0 (18.7)	38.1 (16.8)	43.2 (15.2)
Noticeability (1-7)					
Real Cards	4.87 (1.84)	4.32 (1.89)	5.68 (1.42)	5.26 (1.79)	4.21 (1.9)
AR Cards	4.75 (1.90)	4.45 (2.21)	5.45 (1.73)	4.35 (1.93)	4.75 (1.62)
Understandability (1-7)					
Real Cards	5.71 (1.84)	5.21 (2.07)	6.42 (0.97)	6.26 (1.05)	4.95 (2.46)
AR Cards	5.74 (1.69)	4.9 (2.1)	5.8 (1.8)	6.2 (0.95)	6.05 (1.5)
Urgency (1-7)					
Real Cards	4.63 (1.30)	4.84 (1.12)	4.74 (1.37)	4.74 (1.2)	4.21 (1.51)
AR Cards	4.2 (1.72)	4.15 (1.93)	4.3 (1.62)	4.05 (1.76)	4.3 (1.69)
Intrusiveness (1-7)					
Real Cards	4.46 (1.12)	4.32 (0.95)	4.79 (0.92)	4.53 (0.96)	4.21 (1.55)
AR Cards	4.28 (1.37)	4.15 (1.5)	4.65 (0.99)	3.7 (1.49)	4.6 (1.31)

Missed Notifications. We found a significant two-way interaction between Placement and Task explaining the number of completely missed notifications, $F(3, 105) = 5.09$, $p = 0.002$. We also found a significant main effect for the Placement , $F(3, 105) = 4.82$, $p = 0.003$.

Contrast tests showed a significant difference between *Subtitle*(0.97) and *World*(2.13), $t(105) = 3.57, p = 0.003$, and *Subtitle*(0.97) and *Wrist*(2.05) $t(105) = 2.91, p = 0.022$, regardless of Task. Of note, though not statistically significant, the remaining *Subtitle* comparison to *Heads-up* had p = 0.0751. These results show that there may be a significant difference between subtitle and other type of tested notification regardless of Task. In the real card condition, we found significant differences between *Subtitle*(0.94) and *Wrist*(3.61) $t(105) = 3.42, p = 0.019$, with respect to the number of missed notifications regardless of cards. In the AR card condition, we found significant differences between *World*(2.47) and *Wrist*(0.579) $t(105) = 3.91, p = 0.004$ and between *Subtitle*(1) and *World*(2.47) $t(105) = 3.11, p = 0.04$, with respect to the number of missed notifications. When conducting difference of difference testing, we found that the difference between *HUD* and *Wrist* $t(84) = 2.29, p = 0.024$, changed significantly depending on the Task.

Reaction Time. Two participants didn't respond to any notifications so these participants are excluded from this calculation for reaction time. We only used notifications that produced a response to calculate reactions. We found a significant main effect for Placement, $F(3, 84) = 27.66$, $p < 0.001$. However we found no significant effect for the Task $F(1, 28) = 0.16$,$p = 0.069$ or for an interaction between the Placement and Task explaining the reaction time $F(3, 84) = 2.27$, $p = 0.08$. Contrast tests revealed significant differences between every Placement $p < 0.001$ except *Heads-up* and *Subtitle* $p = 0.17$, disregarding the Task. Given the real card condition, we found significant differences between *Heads-up*(6.54) and *Wrist*(8.69) $p < 0.001$, between *Subtitle*(6.13) and *Wrist*(8.69), $p < 0.001$, and between *World*(5.95) and *Wrist*(8.69) $p < 0.001$ with respect to the reaction time (see Figure 4c). Given the AR card condition, we found significant differences between *Heads-up*(7.3) and *World*(6.01) $p < 0.001$ and between *Wrist*(7.8) and *World*(6.01) $p < 0.001$ with respect to the reaction time (see Figure 4c). When conducting difference of difference testing, we found a significant difference between *Heads-up* and *Wrist* $t(105) = 2.29, p = 0.024$,*Subtitle* and *Wrist* $t(105) = 3.03, p = 0.03$ and *World* and *Wrist* $t(105) = 3.47, p < 0.001$ depending on the Task.

Correct Card Matches. We found a significant two-way interaction between Placement and Task explaining the number of correctly matched cards, $F(3, 108) = 2.75$, $p = 0.046$. We found a significant main effect for Placement, $F(3, 108) = 5.87$, $p < 0.001$. However, the Task caused no significant main effect, $F(1, 36) = 0.71$, $p = 0.40$. Pairwise contrast tests revealed significant differences between *Heads-up*(22.4) and *World*(26.4) , $t(108) = 3.36, p = 0.006$ and between *Wrist*(22.55) and *World*(26.4) , $t(108) = 3.64, p = 0.002$, with respect to the correct card matches regardless of Task. In the AR condition we found a significant difference between *Heads-up* and *World* $t(108) = 4.18, p = 0.001$. When conducting difference of difference testing, we found that the difference between *Subtitle* and *World* $t(108) = 2.72, p = 0.007$, changed significantly depending on the Task.

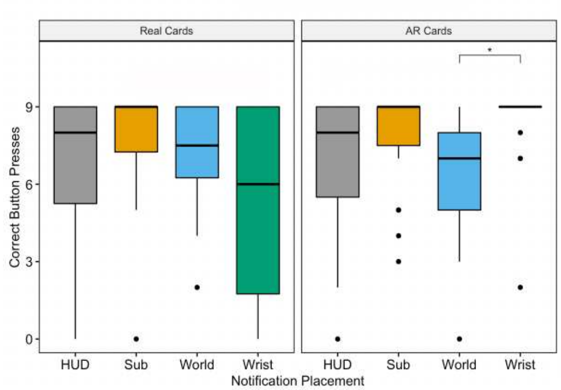
4.2 Questionnaires

Usability. We found a significant main effect for Placement in explaining the Usability score, $F(3, 111) = 3.43$, $p = 0.019$. However, the Task caused no significant main effect, $F(1, 37) = 0.09$, $p = 0.76$. Pairwise contrast tests revealed significant differences between *Heads-up*(64.85) and *Subtitle*(72.85) with respect to the SUS score regardless of Task, $p = 0.013$.

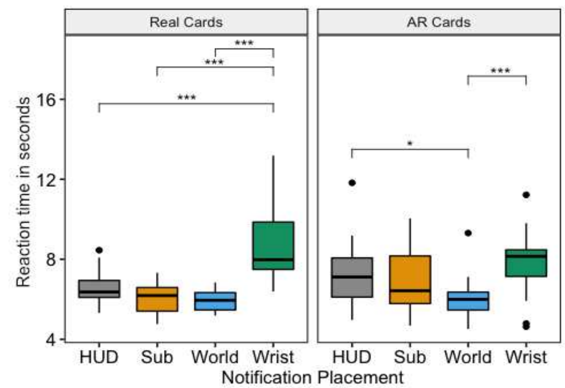
Task Load. We found a significant main effect for Placement in explaining the task load, $F(3, 111) = 4.29$, $p = 0.006$. However, the Task caused no main effect, $F(1, 37) = 0.93$, $p = 0.33$. Pairwise contrast tests revealed significant differences between *Heads-up*(40.8) and *World*(35.85) , $t(111) = 2.63, p = 0.047$, and *Wrist*(41.05) and *World*(35.85) , $t(111) = 3.04, p = 0.015$, with respect to the task load regardless of Task.

4.2.1 Perception.

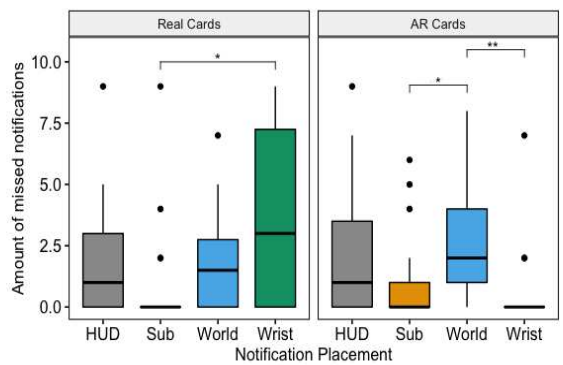
Noticeability. We found a significant main effect for the Placement in assessing the noticeability, $F(3, 111) = 3.58$, $p = 0.016$. However, the Task caused no main effect, $F(1, 37) = 0.27, p = 0.604$. Pairwise contrast tests revealed significant



(a) **Correct button presses**: Depending on notification and card condition. $^{*}p < 0.05$



(b) **Reaction time** : Depending on notification and card condition. $^{*}p < 0.05$, $^{***}\ p < 0.001$



(c) **Missed notifications**: Depending on notification and card condition. $^{*}p < 0.05$,$^{**}\ p < 0.01$

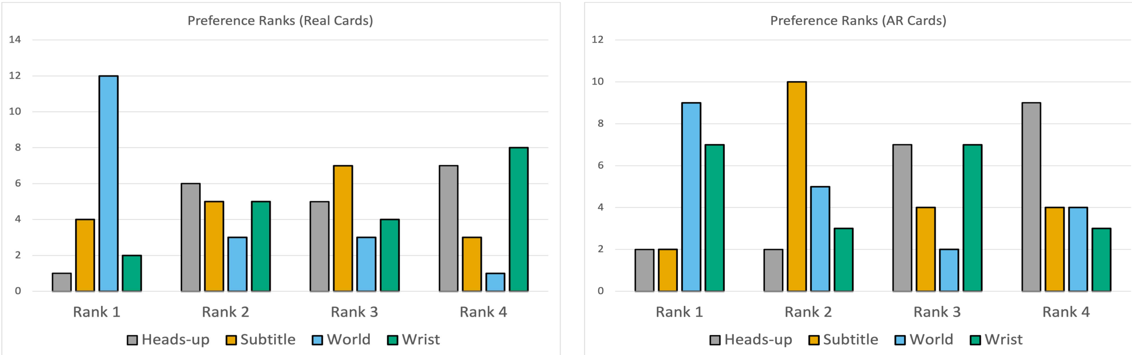
Fig. 4. **Pairwise tests with Notification condition regarding correct button presses, reaction time and missed notifications.**

differences between *Heads-up*(4.38) and *Subtitle*(5.56) , $t(111) = 2.79, p = 0.03$, and *Wrist*(4.48) and *Subtitle*(5.56), $t(111) = 2.88.p = 0.024$, with respect to the perceived noticeability regardless of Task.

Understandability. We found a significant main effect for Placement in assessing the understandability, $F(3, 111) = 3.67$, $p = 0.014$. However, the Task caused no main effect, $F(1, 37) = 0.46$, $p = 0.49$. Pairwise contrast tests revealed significant differences between *Heads-up*(5.05) and *Subtitle*(6.11) $t(111) = 2.74, p = 0.035$, and *Heads-up*(5.05) and *World*(6.23) , $t(111) = 2.88, p = 0.025$, with respect to the perceived understandability regardless of Task.

Urgency. We did not find any significant effects for the Task $F(1, 37) = 0.147$, $p = 0.70$, Placement $F(3, 111) = 0.317$, $p = 0.812$ or the interaction between the two $F(3, 111) = 0.591$, $p = 0.621$ regarding the perceived urgency.

Intrusiveness. We found a significant main effect for the Placement in assessing the intrusiveness, $F(3, 111) = 2.77$, $p = 0.044$. We did not find any significant effects for the Task $F(1, 37) = 0.99$, $p = 0.32$ or the interaction between the two $F(3, 111) = 2.33$, $p = 0.078$ regarding the perceived intrusiveness. Pairwise contrast tests revealed significant



(a) Preference Ranking for the real card Task.

(b) Preference Ranking for the AR-card Task.

Fig. 5. **Preference rankings for each notification type depending on the Task.**

differences between *World*(4.12) and *Subtitle*(4.72) $t(111) = 2.74, p = 0.035$ with respect to the perceived intrusiveness regardless of Task.

4.2.2 Subjective measures. At the end of the experiment, participants were asked to rank the Placement by preference. Ranks were weighted, so a Rank 1 preference would be worth 1 point and Rank 4 is 4 points. This number was then divided by the number of responses to reach an average ranking, with 1 being the most favorable ranking and 4 the least, which can be seen in Table 2 and Figure 5. Across both Task conditions, *World* notifications were ranked statistically significantly higher than the others (p < 0.05).

Table 2. Descriptive statistics; $N = 39$, real cards $n = 19$, AR cards $n = 20$. Values are average rank from 1 to 4 (1 being most favorable).

Scale	Real Cards	AR Cards	Total
Heads-up	2.95	3.32	3.05
Subtitle	2.47	2.63	2.49
World	1.63	2.16	1.85
Wrist	2.95	2.42	2.62

5 DISCUSSION

This study revealed significant effects of Placement and interaction effects between Placement and Task.

Wrist produced the best and also the worst score regarding the correctly pressed buttons, depending on Task. With the AR cards, *Wrist* produced more correct button presses than *Heads-up* or *World* and less missed notifications than *World*. The picture changes quite a bit when looking at the real card Task. Correct button matches, missed notifications and noticeability were all worse than *Subtitle* and reaction time was worse than all other Placements. When looking at the ranking responses, it can also be seen that the preference of *Wrist* notifications depends largely on the card condition, going from tied to last with real cards to second place with AR cards. This can be explained by the fact that with the real cards it was possible to perform the Task without looking through the FOV of the AR headset. Participants often glanced under the center of the display, which caused them to miss notifications on the *Wrist*, with two participants completely missing all *Wrist* notifications. When playing with AR cards, the user was forced to look through the display to see the cards, which ensured that they would also have their *Wrist* in the FOV when tapping on the cards. The other notification placements were largely unaffected by this behavior as they were rendered anyway. This is also supported by research done by Kruijff et al. [21], where they found lower FOV to negatively impact discovery rates of target objects, as they enter the FOV less often. Some participants with real cards explained their ranking with statements like "I did not pay attention to my wrist at all" which explains why they missed some of the notifications. *Wrist* notifications were also rated as the hardest to notice in the questionnaire.

World was better in reaction time in the AR Task than any other and better in the real cards as *Wrist*. They were also better in correct card matches and TLX score than *Heads-up* or *Wrist* across and showed better understandability than *Heads-up* across both Tasks. *World* was the most preferred position for notification Placement, independent of card. This was explained with statements like "over the playing field the notifications were the clearest and quickest to read". The decreased reaction time can be explained because having the notification positioned at their current focus point might lead to a quicker registration than when the user first needs to look somewhere else. Because the participants were quicker to respond, they also had more time to play the card game, leading to a higher match count. Beside the quicker reaction time in AR, *World* always performed the same as *Subtitle* with no significant differences between the two in any other measurement except rank, where *World* placed first overall, and *Subtitle* second.

Subtitle notifications had higher correct button presses and reaction time than *Wrist* with real cards and scored higher in the SUS than *Heads-up*. They were also evaluated as having a higher understandability as *Heads-up* and better noticeability than *Heads-up* and *Wrist*, regardless of Task. The results suggest that *Subtitle* notifications were relatively robust, providing consistent results in all measurements. They didn't require a large shift of attention away from the focus and had good legibility. Being attached to the user's FOV, it was impossible to move them out of vision.

Heads-up placement had a better reaction time and missed notification count than *Wrist* in the real card Task. Comparing *Heads-up* and *Subtitle* notifications, it can be seen that *Heads-up* did not perform better than *Subtitle* in any of the measurements, while also performing worse in several categories and placing last in the preference ranking. Both feature the advantage of always being in view, but participants stated that they found shifting attention to the top-right caused a higher distraction from the game. However, the *Subtitle* notifications enabled them to keep concentrating on the game, while simultaneously reading the notification. One participant stated "bottom-center was the easiest to view, because you're not completely distracted from the game, but you still notice that there is a notification." *Heads-up* notifications were also more prone to an incorrect fit of the headset. Because they were in the peripheral vision not only vertically but also horizontally, wearing the headset incorrectly caused them to be cut off much more easily.

5.1 Design Recommendations

Use Wrist notifications in scenarios with high amount of interaction with AR content. These notifications can provide great results when the user is interacting with other AR content using their hands, but can be easily missed when the hands are outside of the FOV. If the user is not interacting with any other AR content in the environment, other notifications are better suited. Example: Notifications while the user is modifying a 3D model.

Use World notifications if you know where the user is going to look or for stationary tasks. *World* notifications perform well when placed in close proximity to the current task the user is focusing on, without blocking view of the task. If a user is stationary and focusing on a single area for a longer period of time, world notifications seem to be the preferred type of notification. However, if the user is shifting their attention or moving a lot, they might miss stationary notifications placed in the *World*. For these cases *Subtitle* notifications should be used. Example: Notifications while the user is sitting at a desk working on a computer monitor.

When in doubt, use Subtitle notifications for general notifications. *Subtitle* notifications can be displayed to the user regardless of context, do not require any outside world tracking, and are hardest to miss, while still being comfortable and easy to read. Example: Notifications while the user is moving around.

5.2 Limitations

Rating the number of matches might not be a good indicator of task performance, as playing speed and tactic varied between participants. Some prefer to play very systematically and avoid incorrect matches in favor of taking longer to memorize the card order, while others opted for a more direct approach and simply tried to flip cards as fast as possible. Using the real playing cards caused some participants to glance under the headset's display to flip cards as they weren't forced to look through the display, which caused them to completely miss *Wrist* notifications. This is a limitation of the headset's FOV and not necessarily with the placement. However, it is not clear that a larger FOV would have changed the result. Even though the task we chose was a very near-view environment, we still noticed the impact of the small FOV. Testing this in a bigger environment might increase the impact of the smaller FOV even more. But this means that hardware has to be considered when deciding on a certain notification type. Participants were also instructed to use only their right hand in both Task scenarios, as flipping the cards was only possible using the right index finger in the digital card condition. *Wrist*-notifications also only spawned on the right wrist. Even though participants were told this, some used their left hand or both hands during the real card condition. Another possible limitation was the context of the notifications. Participants were instructed to look at and carry out each instruction, which means that there were no unimportant notifications and participants knew to pay attention. This is reflected in the non-significant differences in urgency. The experiment was also set in a quiet and well-lit environment, largely free of any distractions, which might not necessarily reflect real-world conditions. The notifications also arrived at a fixed rate of every 50 seconds, which might have caused the participants to expect their arrival, although data does not seem to support this. Future work should research notification timing and frequency.

5.3 Future Work

Assisting notification delivery in AR through audio cues or haptic feedback should be researched. We saw that *wrist* notifications were missed more because it was possible that the wrist was outside of the headset's FOV. Using audio could help in noticing if a notification is currently present and draw attention to it. It is also worth researching whether notification indicators in the FOV, telling the user that a notification is pending, could improve the notification experience, as a user could then choose when to look at the notification. Notification placement should also be researched with a non-stationary task, especially as a walking task might feature different focal distances which could affect fatigue and performance. Issues with small FOV might also be increased in a non-near-view or non-stationary task. Another topic of interest is to repeat this study with a video-see-through headset and compare differences in notification perception between virtual- and augmented reality, and also optical-and video-see-through augmented reality. In our study the *Heads-up* notifications did not score significantly worse in Intrusiveness, while it did in a similar study done in VR [33], which might indicate that there are noteable differences between AR and VR notfications. Comparing the two could lead to a general notification design guideline in immersive 3D environments.

6 CONCLUSION

In this experiment we compared four different notification placements (*heads-up*, *subtitle*, *world*, *wrist*) in AR while performing one of two card gaming tasks containing physical playing cards *real cards*), or virtual playing cards (*AR-cards*) and constructed design recommendations for notifications in AR. We found that using notifications located on the *Wrist* should take into account how much interactivity or other content is present in the AR environment. Also when using head-stabilized notification in the user's periphery, bottom-center position should be used over top-right placement. The highest number of correct reactions to a notification, was present with *Wrist* notifications but only with a high amount of other virtual content in the environment. The quickest response to notifications was found with *World* notifications.

ACKNOWLEDGMENTS

This project is funded by: DFG–425868361 (SPP2199), NSF-1948254, NSF-2037417, NSF-2016714, NSF-2106590, ONR-N00014-21-1-2580, ONR-00014-21-1-2949N.

REFERENCES

[1] Utku Acer, Afra Mashhadi, Claudio Forlivesi, and Fahim Kawsar. 2015. *Energy Efficient Scheduling for Mobile Push Notifications*. Vol. 2. https://doi.org/10.4108/eai.22-7-2015.2260067 journalAbbreviation: EAI Endorsed Transactions on Energy Web.

[2] Mohammed Safayet Arefin, Nate Phillips, Alexander Plopski, Joseph L. Gabbard, and J. Edward Swan. 2022. The Effect of Context Switching, Focal Switching Distance, Binocular and Monocular Viewing, and Transient Focal Blur on Human Performance in Optical See-Through Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (May 2022), 2014–2025. https://doi.org/10.1109/TVCG.2022.3150503 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[3] Brian P Bailey, Joseph A Konstan, and John V Carlis. 2000. Measuring the effects of interruptions on task performance in the user interface. In *Smc 2000 conference proceedings. 2000 ieee international conference on systems, man and cybernetics.'cybernetics evolving to systems, humans, organizations, and their complex interactions'(cat. no. 0*, Vol. 2. IEEE, 757–762.

[4] Mark Billinghurst and Hirokazu Kato. 1999. *Collaborative Mixed Reality.* Springer Berlin Heidelberg, 261–284. https://doi.org/10.1007/978-3-642-87512-0_15

[5] Soon Hau Chua, Simon T. Perrault, Denys J. C. Matthies, and Shengdong Zhao. 2016. Positioning Glass: Investigating Display Positions of Monocular Optical See-Through Head-Mounted Display. In *Proceedings of the Fourth International Symposium on Chinese CHI - ChineseCHI2016*. ACM Press, 1–6. https://doi.org/10.1145/2948708.2948713

[6] Saverio Debernardis, Michele Fiorentino, Michele Gattullo, Giuseppe Monno, and Antonio Uva. 2014. Text Readability in Head-Worn Displays: Color and Style Optimization in Video vs. Optical See-Through Devices. *IEEE transactions on visualization and computer graphics* 20 (Jan. 2014), 125–39. https://doi.org/10.1109/TVCG.2013.86

[7] Michael DeGusta. 2012. Are Smart Phones Spreading Faster than Any Technology in Human History? https://www.technologyreview.com/2012/05/09/186160/are-smart-phones-spreading-faster-than-any-technology-in-human-history/

[8] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. ACM, Virtual Event USA, 754–768. https://doi.org/10.1145/3472749.3474784

[9] Nicholas Fitz, Kostadin Kushlev, Ranjan Jagannathan, Terrel Lewis, Devang Paliwal, and Dan Ariely. 2019. Batching smartphone notifications can improve well-being. *Computers in Human Behavior* 101 (Dec 2019), 84–94. https://doi.org/10.1016/j.chb.2019.07.016

[10] Jonathan Flittner*, John Luksas, and Joseph L. Gabbard. 2020. Predicting User Performance in Augmented Reality User Interfaces with Image Analysis Algorithms. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 64, 1 (Dec 2020), 2108–2112. https://doi.org/10.1177/1071181320641511

[11] Joseph L. Gabbard, J. Edward Swan, and Deborah Hix. 2006. The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality. *Presence: Teleoperators and Virtual Environments* 15, 1 (Feb. 2006), 16–32. https://doi.org/10.1162/pres.2006.15.1.16

[12] Sarthak Ghosh, Lauren Winston, Nishant Panchal, Philippe Kimura-Thollander, Jeff Hotnog, Douglas Cheong, Gabriel Reyes, and Gregory D. Abowd. 2018. NotifiVR: Exploring Interruptions and Notifications in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (April 2018), 1447–1456. https://doi.org/10.1109/TVCG.2018.2793698 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[13] Rebecca A. Grier, Aaron Bangor, Philip Kortum, and S. Camille Peres. 2013. The System Usability Scale: Beyond Standard Usability Testing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 57, 1 (2013), 187–191. https://doi.org/10.1177/1541931213571042 arXiv:https://doi.org/10.1177/1541931213571042

[14] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (Oct 2006), 904–908. https://doi.org/10.1177/154193120605000909

[15] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[16] David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. 2008. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision* 8, 3 (Mar 2008), 33. https://doi.org/10.1167/8.3.33

[17] Ching-Yu Hsieh, Yi-Shyuan Chiang, Hung-Yu Chiu, and Yung-Ju Chang. 2020. Bridging the Virtual and Real Worlds: A Preliminary Study of Messaging Notifications in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. https://doi.org/10.1145/3313831.3376228

[18] Samat Imamov, Daniel Monzel, and Wallace S. Lages. 2020. 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)*. 851–858. https://doi.org/10.1109/VR46266.2020.00110 ISSN: 2642-5254.

[19] Shamsi T. Iqbal and Eric Horvitz. 2010. Notifications and awareness: a field study of alert usage and preferences. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work - CSCW '10*. ACM Press, Savannah, Georgia, USA, 27. https://doi.org/10.1145/1718918.1718926

[20] Jacek Jankowski, Krystian Samp, Izabela Irzynska, Marek Jozwowicz, and Stefan Decker. 2010. Integrating Text with Video and 3D Graphics: The Effects of Text Drawing Styles on Text Readability. (2010), 10.

[21] Ernst Kruijff, Jason Orlosky, Naohiro Kishishita, Christina Trepkowski, and Kiyoshi Kiyokawa. 2019. The Influence of Label Design on Search Performance and Noticeability in Wide Field of View Augmented Reality Displays. *IEEE Transactions on Visualization and Computer Graphics* 25, 9 (Sept. 2019), 2821–2837. https://doi.org/10.1109/TVCG.2018.2854737 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[22] Kostadin Kushlev, Jason Proulx, and Elizabeth W. Dunn. 2016. "Silence Your Phones": Smartphone Notifications Increase Inattention and Hyperactivity Symptoms. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 1011–1020. https://doi.org/10.1145/2858036.2858359

[23] Shawn Lehner. 2017. ARMaker - Augmented Reality Marker Generator. https://shawnlehner.github.io/ARMaker/

[24] Chin-Chiuan Lin. 2003. Effects of contrast ratio and text color on visual performance with TFT-LCD. *International Journal of Industrial Ergonomics* 31, 2 (Feb 2003), 65–72. https://doi.org/10.1016/S0169-8141(02)00175-0

[25] Feiyu Lu, Shakiba Davari, Lee Lisle, Yuan Li, and Doug A. Bowman. 2020. Glanceable AR: Evaluating Information Access Methods for Head-Worn Augmented Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Atlanta, GA, USA, 930–939. https://doi.org/10.1109/VR46266.2020.00113

[26] Andrés Lucero and Akos Vetek. 2014. NotifEye: using interactive glasses to deal with notifications while walking in public. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*. ACM, Funchal Portugal, 1–10. https://doi.org/10.1145/2663806.2663824

[27] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My phone and me: understanding people's receptivity to mobile notifications. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 1021–1032.

[28] Microsoft. 2022. Typography - Mixed Reality. https://learn.microsoft.com/en-us/windows/mixed-reality/design/typography

[29] Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2014. Managing Mobile Text in Head Mounted Displays: Studies on Visual Preference and Text Placement. 18, 2 (2014), 12.

[30] Chunjong Park, Junsung Lim, Juho Kim, Sung-Ju Lee, and Dongman Lee. 2017. Don't Bother Me. I'm Socializing! A Breakpoint-Based Smartphone Notification System. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, 541–554. https://doi.org/10.1145/2998181.2998189

[31] Lucas Plabst, Sebastian Oberdörfer, Oliver Happel, and Florian Niebling. 2021. Visualisation methods for patient monitoring in anaesthetic procedures using augmented reality. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*. ACM, 1–3. https://doi.org/10.1145/3489849.3489908

[32] Rufat Rzayev, Susanne Korbely, Milena Maul, Alina Schark, Valentin Schwind, and Niels Henze. 2020. Effects of Position and Alignment of Notifications on AR Glasses during Social Interaction. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. ACM, Tallinn Estonia, 1–11. https://doi.org/10.1145/3419249.3420095

[33] Rufat Rzayev, Sven Mayer, Christian Krauter, and Niels Henze. 2019. Notification in VR: The Effect of Notification Placement, Task and Environment. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Barcelona Spain, 199–211. https://doi.org/10.1145/3311350.3347190

[34] Rufat Rzayev, Paweł W. Woźniak, Tilman Dingler, and Niels Henze. 2018. 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*. ACM, Montreal QC Canada, 1–9. https://doi.org/10.1145/3173574.3173619

[35] Cary Stothart, Ainsley Mitchum, and Courtney Yehnert. 2015. The attentional cost of receiving a cell phone notification. *Journal of experimental psychology: human perception and performance* 41, 4 (2015), 893.

[36] Cary Stothart, Ainsley Mitchum, and Courtney Yehnert. 2015. The attentional cost of receiving a cell phone notification. *Journal of Experimental Psychology: Human Perception and Performance* 41, 4 (Aug. 2015), 893–897. https://doi.org/10.1037/xhp0000100

[37] Daniel J. Velleman and Gregory S. Warrington. 2013. What to Expect in a Game of Memory. *The American Mathematical Monthly* 120, 9 (Nov 2013), 787–805. https://doi.org/10.4169/amer.math.monthly.120.09.787

[38] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Vancouver BC Canada, 143–146. https://doi.org/10.1145/1978942.1978963

5.2 Modality and Location of Notifications

In the first experiment, we did not include audio when a notification was delivered to focus on the location and the impact of clutter. Notifications were also disconnected from the main task, and therefore distracting for it. The following paper[2] investigates the impact of notification modality and further details notification location [80]. We used a virtual cooking task [81] running on the Hololens 2 that participants had to perform, during which they would receive notifications relating to the task. Notifications were either delivered visual only, audio only, or using a combination of both. Directly related to the task, notifications informed the user about events in the environment and were placed in a world-stabilized way, either over the object causing the notification, or on a fixed list in the environment.

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

Aditya Raikwar*
Colorado State University

Lucas Plabst†
Colorado State University

Anil Ufuk Batmaz‡
Concordia University

Florian Niebling§
Advanced Media Institute, TH Köln
University of Applied Sciences, Köln, Germany

Francisco R. Ortega¶
Colorado State University

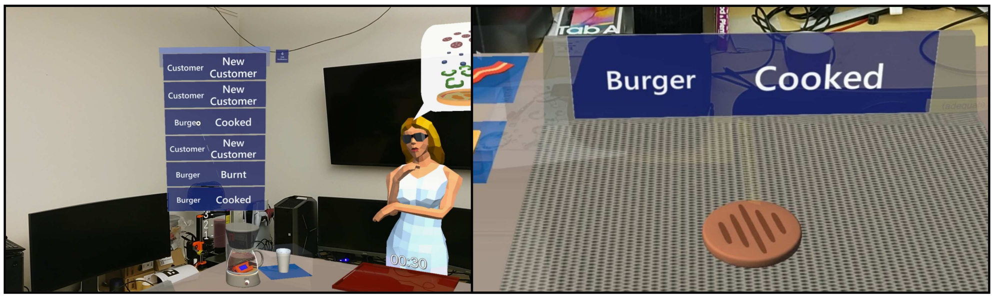


Figure 1: Left: Notifications on Dock; Right: Notification on Object

Abstract

Implementing visual and audio notifications on augmented reality devices is a crucial element of intuitive and easy-to-use interfaces. In this paper, we explored creating intuitive interfaces through visual and audio notifications. The study evaluated user performance and preference across three conditions: visual notifications in fixed positions, visual notifications above objects, and no visual notifications with monaural sounds. The users were tasked with cooking and serving customers in an open-source Augmented-Reality sandbox environment called ARtisan Bistro. The results indicated that visual notifications above objects combined with localized audio feedback were the most effective and preferred method by participants. The findings highlight the importance of strategic placement of visual and audio notifications in AR, providing insights for engineers and developers to design intuitive 3D user interfaces.

Index Terms: Augmented Reality, Human-computer interaction (HCI), Visualization design and evaluation methods, Notification

1 Introduction

Interruptions are a significant research topic in Human-Computer Interaction (HCI), as humans have an innate ability to multitask [17], leading to frequent switching between tasks by choice or due to external interruptions. This multitasking behavior is common in day-to-day life, such as checking emails during meetings. Gould et al. discuss that interruptions can have both positive and negative effects on user attention and task performance [7]. While interruptive notifications can disrupt attention and hinder performance, well-designed notifications can aid in task management and reduce user anxiety by providing timely reminders and updates, particularly in dynamic and time-sensitive situations [13].

Notification windows are commonly used to interrupt users in Head Mounted Displays (HMDs) applications [31]. When designing these notifications, it is essential to avoid obstructing the user's field of view (FoV), particularly during locomotion. In Augmented Reality (AR) HMD systems, designers must consider factors such as the placement of information, acceptable levels of occlusion, timing of the display of information appropriately, and selecting the most effective interaction method for users [12]. Since notifications in AR-HMDs often involve displaying information directly within the user's visual field, conventional frameworks used for smart and desktop devices may not be directly applicable to the presentation of information in AR-HMD systems.

As display systems continue to advance, they are becoming increasingly integrated into users' daily lives. Screens have evolved from desktops to laptops, then to handheld devices, and now AR HMDs are placing screens optimized for clear and sharp images directly in front of our eyes. Moreover, the integration of virtual elements into the user's real-world environment through AR HMDs offers a seamless and immersive experience. All these advancements suggest that AR "glasses" are likely to become ubiquitous in the near future. Therefore, understanding the potential integration of these AR devices into daily life is crucial.

*e-mail: adirar@colostate.com
†e-mail: lucas.plabst@colostate.edu
‡e-mail: ufuk.batmaz@concordia.ca
§e-mail: florian.niebling@th-koeln.de
¶e-mail: fortega@colostate.edu

Understanding the user's context is crucial for delivering relevant information and controlling the pervasive devices [9]. Exploring the placement of notifications in an AR-HMD during demanding tasks, such as cooking and serving food to customers, is, to our knowledge, novel, particularly when considering the incorporation of both visual and audio modalities for notifications. The objective of this paper is to investigate how users can smoothly transition between tasks, guided by the Multiple Resource Theory (MRT) [40], *using visual and audio modalities*. The MRT offers insight into how users manage concurrent tasks and interruptions through different channels.

In this study, we looked at how notifications can help with the primary task as opposed to interrupting them. Previous studies have looked at the negative effects of interruptions on the primary task due to notifications. Studies like [36, 27, 16, 19] state that task-independent notifications will decrease the performance. In this paper, we designed two types of world-fixed visual notifications: Notifications On Dock and Notifications On Object, as shown in Figure 1. The audio notification may or may not accompany the visual notifications in the form of bubble popping sound (refer to the supplementary video for an example). To enhance participants' visibility and attention, we incorporated sound feedback. Our system was developed using the open-source ARtisan Bistro environment [30]. Various parameters in ARtisan Bistro, such as customer frequency and cooking speed, were adjusted to induce greater task-related stress, particularly regarding the participants' required task completion speed. Notifications were added to assist in task completion when user attention was interrupted. Our results showed that placing notifications directly on relevant objects can improve task performance and is preferred by users. Additionally, we demonstrated that the use of sound is an important factor in capturing the attention of users. Audio notifications enhanced the noticeability of visual cues, leading to better performance and higher user satisfaction. These findings highlight the importance of combining visual and audio notifications to optimize user experience in AR applications.

2 Related Work

Unlike conventional smart devices, AR HMDs represent a novel computing environment, which poses unique challenges for implementing notification mechanisms. Understanding how to effectively utilize this new approach to deliver AR notifications on HMDs in real-world scenarios is crucial [28, 21].

2.1 Notifications

One of the ways of getting users' attention is by means of notifications. Notifications can be delivered through visual, audio, and haptic mediums, and they generally fall into two categories [33]. The first category is **action-required notifications**, where the user is required to take some immediate action based on the information provided by the notification. E.g., Windows asking for administrative authentication. The second category is **passive notifications**, where the user is presented with some information that does not require the user to act, e.g., calendar notification.

A typical multimodal notification, such as receiving an email on a smartphone, includes a visual pop-up, a sound cue, and a vibration sequence to alert the user in various situations. Notifications vary based on context and user needs, and determining the optimal moment to interrupt users without human intervention is an ongoing research area. For instance, Kern et al. identified five factors influencing user interruptibility: location, event importance, user's activity, social situation, and social activity [15].

2.2 Visual Cues

Visual cues provide information to users through vision, ranging from simple error Light-Emitting Diode (LED) lights to complex AR notifications. Wallmyr et al. conducted a study on transparent interfaces using mixed reality to display key information to construction site operators. Their findings indicated that users responded more quickly and experienced lower workloads with head-up displays compared to head-down displays [37].

The position of the visual cues is also important. Harrison et al. [10] conducted a study to examine the reaction time performance based on the placement of visual cues, specifically small blinking lights on different body parts. Their study aimed to determine the effectiveness of visual cues in terms of accessibility, stability, comfort, social acceptability, and information conveyance. They found that reaction time performance varied significantly with the position of the lights, ranking from highest to lowest as follows: wrist, arm, brooch, shoulder, thigh, waist, and shoe. The performance was influenced by physical distance, visual accessibility, and external factors such as occlusion by furniture.

The study by Weber et al. [39] investigated the use of notifications on smart televisions (TVs), which are often shared by multiple users. Unlike other smart devices, the notification mechanisms for TVs cannot be personalized. Through three focus groups, the study gathered impressions regarding the duration, amount of information, position, and number of notifications. Based on these findings, the researchers highlighted that the notifications that are truly important should be displayed, privacy is important when multiple people use the TV, and the notifications should be displayed when there is a break.

2.3 Audio Cues

In situations where visual cues can be distracting and harmful, audio cues are recommended. Lee et al. [23] studied a collision detection system with audio and haptic cues to mitigate driver distraction. They found that graded warnings (gradually increasing levels) were preferred and performed better than single-stage warnings. Graded warnings were also less irritating and more trustworthy. While user performance was the same for haptic and audio cues, users preferred haptic cues in terms of 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 an obstacle is in the way [4]. All auditory displays improved object localization compared to silence. Continuous spatial sound and sequential discrete auditory cues proved the most effective. Whereas, echolocation and auditory looming were somewhat less efficient.

2.4 Notifications in Mixed Reality

Mixed reality devices, ranging from smartphones with a rear camera to immersive HMDs like Oculus Quest Pro [34], are primarily output visual information. The key considerations for displaying notifications or interrupting users include where to display information, the acceptable level of occlusion in different contexts, the timing of information display, the behavior of the currently running application, and the interaction methods for the user.

On the other hand, notifications in Virtual Reality (VR) cannot use the same framework as other smart devices without breaking immersion. Zenner et al. [43] proposed a framework where notifications are integrated into the VR scenario, such as a villager delivering a letter in a medieval setting or a drone in a futuristic one, based on the context and urgency of the notification. While this approach maintains immersion, it requires additional work for developers and standardization of notification priorities. Rzayev et al. conducted experiments on the effects of notification positions in VR and AR, finding that top-right placement increased workload and reduced comprehension, and motion negatively impacted comprehension [32]. Imamov et al. conducted a study to determine the optimal placement of interfaces in 3D space [12]. They discovered that central and central-low positions on the display were the fastest, whereas the top left position was the slowest. In terms of depth, the closest position (1m) was the slowest, while positions at 2m and 3m were the fastest for central and central-low placements.

Cidota et al. investigated how different types of notifications impact workspace awareness and task performance in an augmented reality (AR) setting [3]. The study compared three conditions: no notifications, audio notifications, and visual notifications, where a remote user was instructing a local user by the use of these notifications. Key findings reveal that users prefer visual notifications over audio or no notifications. The authors noted that visual notifications caused less cognitive overload, possibly because the game's tasks already required visual attention. An audio signal would have forced participants to divide their attention between two cues (audio and visual) instead of focusing on a single modality (visual). Woodward et al. conducted an extensive systematic literature review of 140 peer-reviewed studies to assess the effectiveness of augmented reality (AR) in enhancing situational awareness (SA) [42]. Key findings indicated a significant gap in the use of specific SA evaluation techniques in the majority of user studies, with only 19% included such methods. While several studies analyzed the color and style of text in AR, there was a notable lack of research on the users' SA, emphasizing the need for future studies to explore these areas to improve AR's efficacy in maintaining SA.

In a study by Lee et al., the efficacy of displaying AR notifications in various positions on an HMD during dual-task performance was investigated [21]. Results showed that notification location significantly impacted performance, with the top-left position causing the highest task load and slowest response time and middle positions yielding the lowest task load and fastest response time. The study concludes that middle positions are optimal for AR notifications in dual-task scenarios, as they offer visibility with minimal disruption. The importance of display location in AR notification design, especially for multitasking users, is emphasized. According to the authors, this position provides sufficient notification visibility while causing the least disruption to the main job. The study emphasizes how crucial it is to consider display location when creating augmented reality notifications, especially when users must carry out several tasks at once.

Ghosh et al.'s NotifiVR extends notifications to visual, auditory, and haptic cues to alert users immersed in virtual environments to external events [6]. They proposed five types of notifications for each cue type. For audio cues, they suggested that the sound played should be at the position of notification in 3D space, metaphorical design like someone entering space as a door opening sound, and gradual increase of sounds to signify increasing urgency. For haptic cues, they recommended metaphorical vibrations like feet vibrating to signify footsteps, vibrations in the direction of notification, and a gradual increase of vibration to signify increasing urgency. For visual cues, they proposed notifications should be displayed in 3D space and not always attached to the viewport, metaphorical representations like clock face for alarms, blocks of notifications placed in 3D space, on the wall or surface, and in dangerous cases stopping the immersion and letting user reorient with the real-world space.

Despite extensive research on AR in academic and industrial settings for specific use cases, there has been limited study on displaying notifications in AR HMDs. Companies like Meta [34], Apple [18], and Xiaomi [35] are working to bring AR to the general public. The number of notification studies, specifically in AR HMDs, is limited. For example, Lucero et al. conducted a study where participants walked down a busy road while tracking notifications using NotifEye [25]. They used a rubbing pad for simple interactions. Participants found that minimalistic information combined with the rub pad created a non-distracting notification system. The main concern was social acceptability, including size, comfort, and fashion.

Plabst et al. [29] found that notifications in AR HMD are most effective when placed in the real world or at the bottom center of the headset's FoV, especially for sustained concentration tasks. Users also preferred these positions. Similarly, Lee et al. [22] observed that bottom FoV placement resulted in higher noticeability and comprehension for both icon- and text-type notifications compared to top placement during an AR walking task. Lazaro et al. [20] recommend using both visual and auditory signals for AR notifications and suggest further research on notification placement.

The findings from Zenner et al. [43] on the positioning of notifications informed our design choices, particularly in placing notifications in context-aware locations to enhance visibility and minimize disruption. Additionally, the study by Ghosh et al. [6] on multimodal notifications

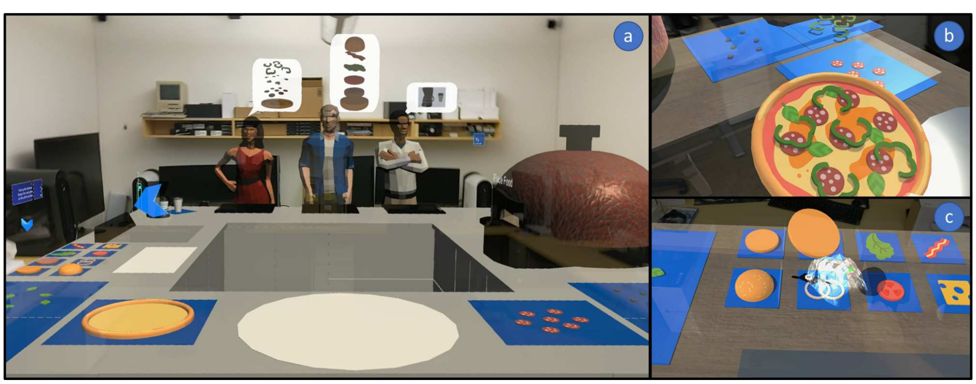


Figure 2: (a) Complete Setup of ARtisan Bistro; (b) Perfectly Cooked Pizza Taken out of the Oven; (c) Making Burger, Picking up Bottom Bun by Hand [30]

and their contextual application provided a foundation for our exploration of integrating visual and audio cues effectively. Specifically, their recommendation to play audio cues at the position of the notification in 3D space was implemented in our study, further enhancing the relevance and impact of our notifications. By building on these foundational works, our results demonstrated that placing notifications directly on relevant objects can improve task performance and is preferred by users while also highlighting the crucial role of sound in capturing user attention.

3 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 [30]. 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[26]. 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. Figure 2 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.

3.1 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.

4 Notification Design

The notifications were designed with two modalities: visual and audio. Depending on the level, participants received notifications that were visual, audio, both visual and audio, or none.

4.1 Visual Notifications

AR-HMD environment offers a dynamic way to display notifications using 3D space. Zenner et al.'s method integrates notifications into the scenario to maintain immersion, but it may be excessive as it limits notification types and relies on a person to send messages [43]. This method also places the responsibility of setting the correct priority level on the sender, which can vary among individuals. To address these issues, two types of visual notifications were designed to handle more general types of notifications.

4.1.1 Visual Notification Design and Properties

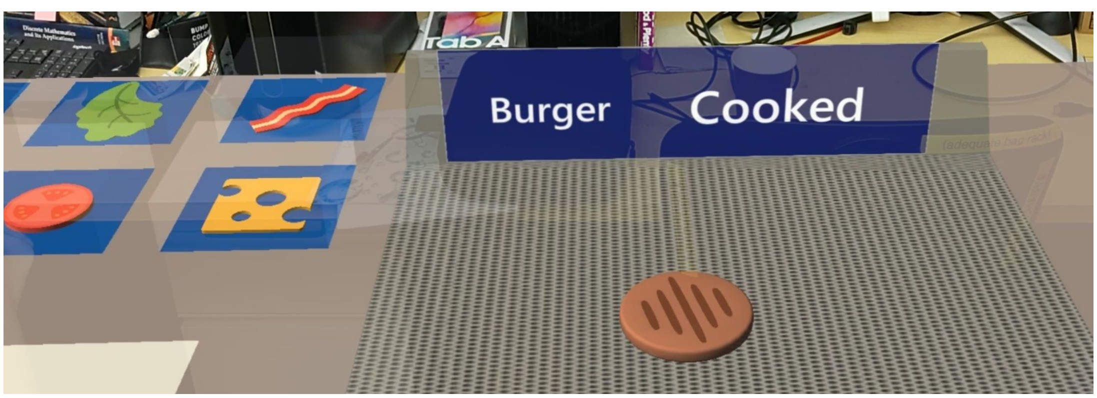


Figure 3: Notification Button

The notification system, as illustrated in Figure 3, displayed information in a 'Station Message' format. The 'Station' indicated the cooking station's name, while the 'Message' provided relevant information for that station. The notifications were cuboid-shaped, navy blue with white text. The notification design is based on the buttons provided and recommended by Hololens 2 developers. They were disabled after seven seconds. This duration, determined during the design phase of the experiment, was found suitable for users to pause their primary task, view the notification, and understand its content. Notifications that went unnoticed for seven seconds were disregarded.

The intervals between the notifications were based on the actions of users and the environment. For example, a customer stayed for 2 minutes after requesting their food. The next customer comes 5 seconds after the previous one leaves. In this case, the notification interval between these 2 customers will be 125 seconds, but if the participant serves the customer in 65 seconds, then the interval will be 70 seconds.

We used 4 texts to notify the participants.

New Customer: When a new customer arrives requesting food.

Cooked: When either burger patty on the grill or pizza in the oven is cooked to the level of customer satisfaction.

Burnt: When either burger patty on the grill or pizza in the oven is cooked beyond the level of customer satisfaction.

Coffee cup added: When the coffee level in the coffee maker crosses a threshold where a complete coffee cup can be filled.

4.1.2 Notification on Object

Figure 4: Notification on Object

Inspired by Zenner et al.'s [43] implementation of immersive notifications, Notification on Object (O) was created to present notifications in a context-aware 3D space. This approach aims to provide users with relevant information without overwhelming them, especially when focusing on specific tasks. For instance, in a cooking scenario, notifications about the pizza's cooking status can be displayed over the oven, allowing users to respond quickly without breaking immersion.

In the experiment, notifications were displayed on objects in a single box format Figure 4. If a new notification from the same station appears before the current one is destroyed, the text is updated with the new information, and the timer is reset.

Although this solution reduces clutter and increases response time to notifications, it can also lead to missed notifications if they are outside the user's field of view or if the user does not look at them in time. One solution is to use audio alerts for new notifications. Another option is to list all notifications in one place, allowing users to periodically check the list for any new notifications.

4.1.3 Notification on Dock

The concept of Notification on Dock (D) is inspired by smartphone notifications, where all notifications are displayed in one accessible place by swiping down on the screen [1, 2]. This list format simplifies access to information and reduces the mental effort required to find notifications. Users might miss notifications if they appear randomly, but a centralized list allows users to periodically check for new notifications. Unlike object-based notifications, a dock cannot provide additional information beyond text (like the location of the object that prompted the notification), making it important to indicate the station that generated the notification.

Building on this idea, the dock is a drop-down list that displays notifications in reverse chronological order, as shown in Figure 5. When a notification is destroyed, the subsequent notifications move up to fill the space. The dock can be positioned anywhere in the 3D space, allowing users to place it where they are most comfortable or move it periodically during the experiment.

Notification on Dock, like Notification on Object, has its drawbacks. Although it helps users locate notifications more easily, users must find an optimal position for the dock in 3D space. Inefficient placement can increase the time to find visual objects in 3D space [38]. Additionally, the design does not inherently provide location information, necessitating extra details in the message to indicate the location.

Figure 5: Notification on Dock

4.2 Audio Notification

Incorporating audio cues serves as a potential solution to mitigate the limitation of visual cues not being feasible as the primary notification method [4]. Different sounds can be assigned to different notifications, similar to how smartphones allow different tones for different applications. This helps users identify and prioritize notifications without needing to look at the application names.

In this experiment, only one localized sound clip played for notifications, allowing users to determine the direction of the sound. The Hololens 2's speakers, despite not offering a full surround sound experience, provide adequate left-right differentiation. Notifications were limited to the XZ plane, making the Hololens 2 speakers sufficient. Localized audio helped users identify the direction and source of the notification, such as different cooking stations, thereby reducing response time. The sound's origin depends on the notification's spawn location: on the dock for dock conditions and above the object for object conditions. In a condition with only sound and no visual notifications, the sound plays as a standard audio clip as if it were in the user's head, alerting participants to an event or task. There are two conditions for audio notification, with sound (S) and without sound (N).

5 Methodology

5.1 Research Questions

Based on the previous work, we investigated the following Research Questions (RQ) in this paper:

RQ1 *How does the inclusion of visual and audio cues influence the performance of users in serving customers?* Prior research shows that visual and audio cues can enhance user attention and reduce task load [21, 13]. By studying this in the context of serving customers, the research aims to understand how these cues help in managing interruptions and multitasking, potentially leading to more efficient and accurate task performance in AR, using MRT[40]. Building on these studies, will users perform better in Notifications on Object with Sound *OS*, Notifications on Object without Sound *O*, Notifications on Object with Dock *DS*, Notifications on Object without Dock *D*, and no Visual Notification with Sound *S* conditions compared to the no Visual Notification without Sound *N* conditions?

RQ2 *What type of notifications are most noticeable?* The notifications should warn the user in a virtual environment when they focus on a task. With this RQ, we seek to determine which notification types are most prominent and immediately noticeable to users, ensuring that critical information is not missed. Which of these conditions (*OS*, *O*, *DS*, *D*) drew the user's attention the best?

RQ3 *What type of visual and audio cues do users prefer?* The previous work indicates that user preferences can significantly affect the perceived effectiveness and acceptance of notifications. For example, Lee et al. found that users preferred graded audio warnings in driving scenarios [23]. Which of the following conditions - *OS*, *O*, *DS*, *D* - was the most desired? By understanding user preferences in the context of AR-HMD notifications, designers can create more user-friendly and accepted notification systems that align with user expectations and needs.

These research questions are crucial for developing effective and user-friendly AR-HMD notification systems, ensuring that users can efficiently manage tasks and interruptions in dynamic and potentially stressful environments.

5.2 Experiment Design

This study used a within-subjects design with two factors: auditory notifications (A_2 =with audio notification and without audio notification) and visual notification (V_3 =notification on object, notification on dock, and no Visual notification), yielding 6 different conditions: Notification on Object Without Sound ***O***, Notification on Object With Sound ***OS***, Notification on Dock Without Sound ***D***, Notification on Dock with Sound ***DS***, no visual notification without sound ***N***, and no visual notification with sound ***S***.

Each participant followed the following steps. The experiment began with participants arriving at the research lab, completing a consent form, and receiving an explanation of the study. They watched a brief video to familiarize themselves with the environment and notification types. After addressing any questions, participants were given a Hololens 2 headset and entered the AR environment. They started with an introduction screen and then proceeded to a tutorial level where they practiced making food items for three customers without a time limit. The ingredient list was consistent for all participants.

After serving three customers, a level was chosen from a shuffled list of conditions. The order of levels was randomized for each participant using a chi-square model. Instructions were provided at the start of each level to inform participants about the type of notification they would encounter. This process was repeated until all six conditions were completed, followed by post-surveys.

Each level followed the script where six customers (maximum three at a time) made requests for a food item. The customers were virtual, asking for virtual food. The participants were tasked to make the item in a set amount of time. The appliances were also virtual, presenting extremely low to no risk to the participants. This scenario consisted of timed tasks where the participants had to worry about not only the customers' patience (tolerance time limit) but also the cooking status of different food items. The cooking status notifications were provided to aid the participants with the tasks and keep track of the cooking status. The notifications were presented using different modalities and at different positions. The pictorial representation of the experiment flowchart is available for reference in the Supplementary section.

During the study design process, it was determined that six customers (two requesting each food type) would be used in this study. The number of ingredients needed for a burger or pizza remained consistent across all trials and participants. The order of food items requested was decided by sampling without replacement. Each customer experienced identical wait times for all food items, ensuring that each participant received the exact same wait times for each food item.

In *O* and *D* conditions, the visual notification stayed active for 7 seconds. In *D* conditions, if a new notification from the same station appeared, it replaced the existing one. Notifications from different stations were stacked. The notification sound was a bubble-popping sound (refer to the supplementary video for an example), chosen for its lightness and lack of food metaphor. An additional number, randomly selected between 0 and 15 and unique among active notifications, was added to the notification text to note observation. Participants acknowledged the notification by calling out this number. The measurements recorded included if a customer departed before 120 seconds, customer waiting time, total number of customers served, notification recognition frequency, and reaction time.

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 [11]. The participants completed the NASA-TLX survey after they completed all the trial conditions. The average overall workload experienced by participants was 60.64 (SD = 10.84). This is above average global workload score reported in [8]. The average raw scores are listed in Table 1.

Table 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

5.3 Participants

The experiment involved 26 participants, equally divided between males and females, aged 23 to 70 (M = 32.73, STD = 11.9). Participants consisted of students and staff from Colorado State University, as well as people who were not affiliated with the university. 84.6% of participants were either students or faculty. 69.23% of the participants reported that they had used 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.

6 Results

To analyze the data gathered in the experiment, we performed a two-way ANOVA. If the assumptions of the ANOVA were not met, the data was transformed using Aligned Rank Transform [41]. If significant effects were found, we conducted a post hoc pairwise analysis using TukeyHSD, or in the case of transformed data, the ART-C procedure [5].

6.1 Performance

Testing for the performance of the participants based on how many customers they served is shown in Figure 6. 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). The error bars represent standard error. Individual means, standard deviations, and standard errors can be found in Table 2. During the ART-C procedure, the degrees of freedom are calculated using the Kenward-Roger Method [14].

We found that there was a statistically significant interaction between the effects of visual cues and audio cues ($F(2,125) = 4.14$, $p < .05$). Simple main effects analysis showed that visual notifications ($F(2,125) = 2.01$, $p = .14$) and audio notifications ($F(1,125) = 0.27$, $p = .61$) did not have a statistically significant effect on the performance.

Contrast tests showed that Notifications on Object with Sound performed significantly better than both Notifications on Dock with Sound ($t(125) = 3.52$, $p < .01$) and Control with Sound ($t(125) = 3.16$, $p < .05$). There were no other significant differences found.

There was no significant effect on the performance of participants in serving the customers based on whether they have worked in a restaurant in some capacity ($t(24) = 1.6$, $p = .12$), despite participants with restaurant experience (M = 5.12, SD = 0.72) serving more customers than those without (M = 4.57, SD = 1.01). This was also true in the case of task load. NASA-TLX results show that there was no significant difference between the overall task load perceived by the participants who have worked in a restaurant (M = 50.15, SD = 13.33) compared to those who haven't (M = 49.62, SD = 11.56), $t(24) = 0.11$, $p = .91$. Just looking at the temporal demand ($t(24) = -1.07$, $p = 0.29$), the participants who have worked in a restaurant (M = 7.08, SD = 2.36) reported a similar level of temporal demand compared to participants who haven't worked in a restaurant (M = 7.92, SD = 1.61).

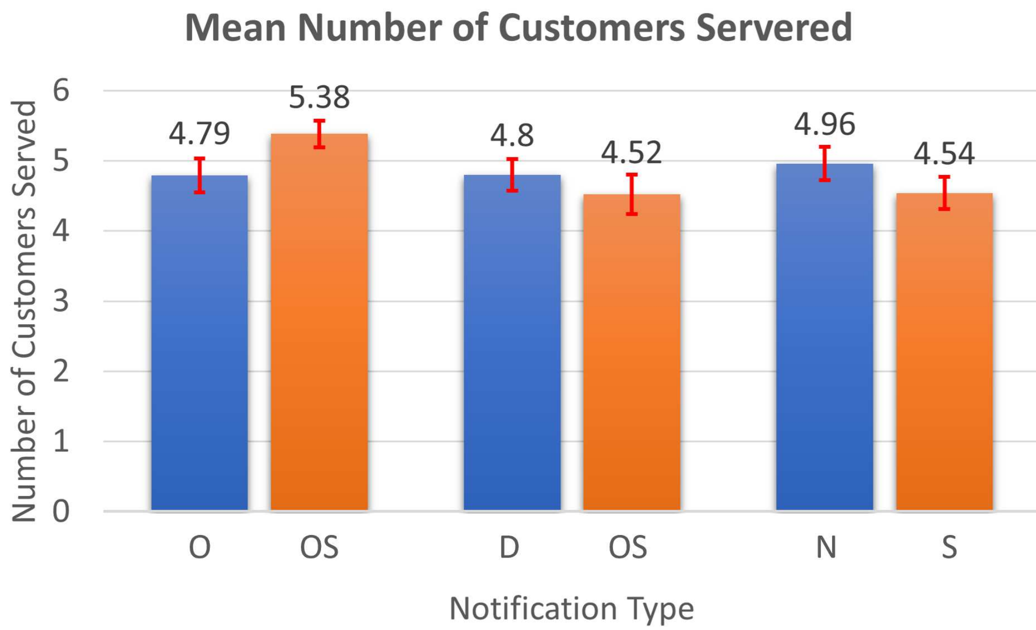


Figure 6: Performance of participants in different notification conditions based on how many customers they served.

Table 2: Mean performance of participants in different notification conditions

Notification Techniques	Mean	STD	SE
O	4.79	1.24	0.24
OS	5.38	0.96	0.19
D	4.8	1.14	0.22
DS	4.52	1.44	0.28
N	4.96	1.19	0.23
S	4.54	1.18	0.23

6.2 Visual Notifications Called

As stated before, in addition to serving the customers, the participants were asked to shout out the number displayed above the text in a visual notification. Testing for how many visual notifications were noticed is shown in Figure 7. The x-axis represents the different conditions (2 conditions of visual notifications and two conditions of audio notifications). The y-axis represents how many customers were served on average in each condition (maximum is 6). The error bars represent standard error. Individual means, standard deviations, and standard errors can be found in Table 3.

Testing revealed that there was no statistically significant interaction between the effects of visual notifications and audio notifications ($F(1,100) = 0.2$, $p = .66$). Simple main effects analysis showed that visual notifications did not have a statistically significant effect on noticeability ($p = .7$). Simple main effects analysis showed that audio notifications did have a statistically significant effect on the noticeability ($p < .001$), with more notifications being called out with sound (M = 6.865, SD = 4.18) than without (M = 4.1, SD = 2,71).

6.3 Visual Notifications Called Reaction Time

Testing for the reaction time of noticing visual notifications is shown in Figure 8. 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 take to shout

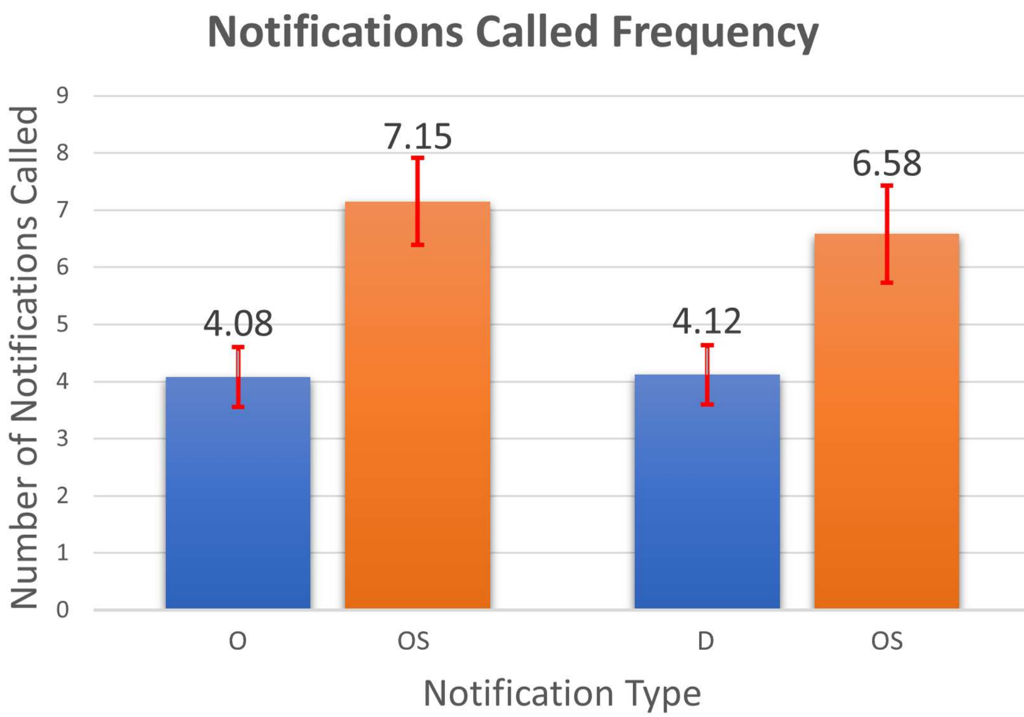


Figure 7: Average number of notifications called in different notification conditions

Table 3: Mean number of notifications called in different notification conditions

Notification Techniques	Mean	STD	SE
O	4.08	2.66	0.52
OS	7.15	3.88	0.76
D	4.12	2.65	0.52
DS	6.58	4.32	0.85

out the number on the notification. The error bars represent standard error. Individual means, standard deviations, and standard errors can be found in Table 4.

Testing revealed no statistically significant interaction between the effects of visual and audio notifications ($F(1,68.08) = 3.28$, $p = .07$). Simple main effects analysis showed that visual notifications did not have a statistically significant effect on reaction time ($p = .36$), and audio notifications did not have a statistically significant effect on reaction time ($p = .34$).

6.4 User Preference

The system's usability was assessed using the Post-Study System Usability Scale (PSSUQ) [24]. All participants completed the PSSUQ after completing all the conditions in the AR environment. There were 18 items that were scored on a Likert scale from 1 (strongly agree) to 7 (strongly disagree). Other than the 18 items, we added a preference question and asked participants which notification condition they preferred among 6 conditions (Notification on Object Without Sound, Notification on Object With Sound, Notification on Dock Without Sound, Notification on Dock With Sound, No Notification without sound, and No Notification with sound)

Overall, participants rated the system favorably on the PSSUQ, with a mean score of 1.23 (SD = 0.44), indicating generally positive perceptions of usability. After examining individual items, the participants agreed strongly with most of the statements. Given the subjective nature of the questionnaire, we have considered any mean agreement rate below 1.75 (25 %) as strongly agreeing. Out of all the items presented in the survey, only item 4, which discussed task completion time (M = 1.85, SD = 1.38), and item 8, which talked about error messages (M = 2.38, SD = 1.88), had varying levels of agreement. For the complete survey and detailed results, refer to section 8.

The last question in the survey asked about the preferred notification type. The preferences were as follows: 50% preferred Notification on Object With Sound, 7.7% preferred Notification on Object Without Sound, 19.2% preferred Notification on Dock With Sound, 0% preferred Notification on Dock Without Sound, 19.2% preferred No Notification With Sound, and 1% preferred No Notification Without Sound.

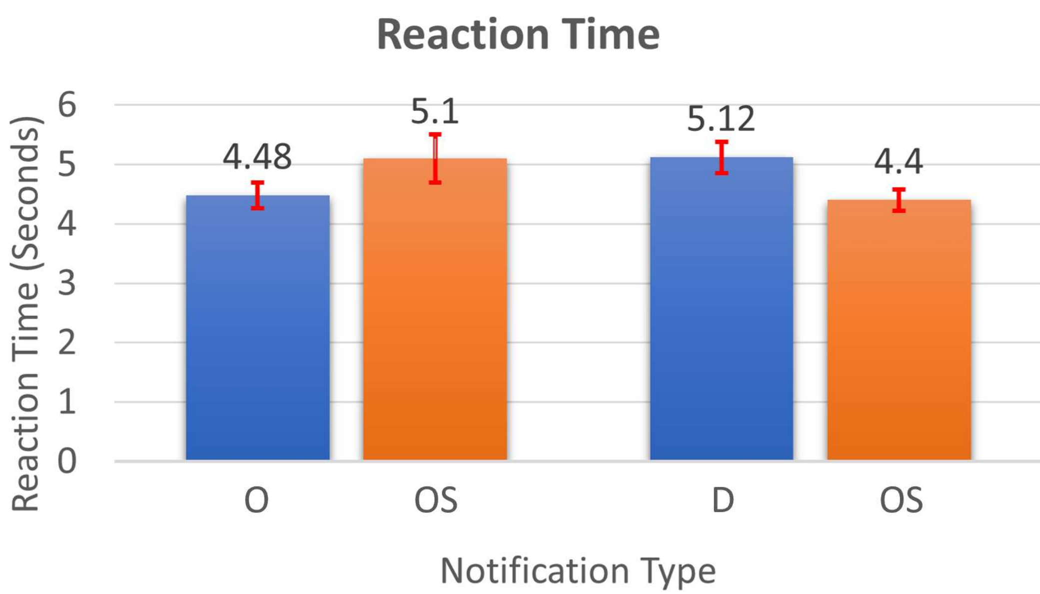


Figure 8: Average reaction time to notice the notification in different notification conditions in seconds

Table 4: Average reaction times to notice the notification in different notification conditions in seconds

Notification Techniques	Mean	STD	SE
O	4.48	1.1	0.21
OS	5.1	2.07	0.41
D	5.12	1.34	0.26
DS	4.4	0.93	0.18

7 Discussion

In this paper, we looked at how two distinct visual and audio notification systems affect the user's performance and preference. The visual notifications were strategically placed in fixed positions or directly above relevant objects, while audio notifications were localized to align with the visual notifications. Our results demonstrated that visual notifications positioned above objects and accompanying audio outperformed alternative configurations. Placing AR notifications in the world anchored to the task context (in our case, on the object) was also found to produce the quickest reaction time, as well as the highest preference and lowest task load in a study by Plabst et al. [29], even though they did not feature audio. Notably, this combination of visual and audio notifications exhibited greater noticeability and was the preferred mode of notification by the participants.

These findings emphasize the importance of carefully choosing where and how visual and audio notifications are used in AR. In our discussion, we explored what these results mean and why they matter for designing AR interfaces.

Looking at how many customers (out of six) were served (RQ1), neither visual nor audio cues had any significant effect on their own, but the combination of the two showed promising results. We contend that the audio notifications prove effective in capturing users' attention while simultaneously asserting that visual notifications above the objects accelerate users' ability to locate and engage with their primary tasks. The post hoc analysis illuminated a noteworthy trend: when audio notifications are present, visual cues above objects aid the users significantly more than visual notifications on a dock and no visual notification. Intriguingly, our findings did not uncover significant distinctions between different visual cue configurations when audio notifications were absent. This observation may suggest that, in the absence of sound notifications, participants relied less on visual notifications as a primary means of task assistance.

This observation aligns with the outcomes of our noticeability assessment (RQ2), which showed that the noticeability of the notifications significantly improves when audio notifications are present. In a 3D environment characterized by distributed information, the likelihood of missing notifications outside the user's immediate FoV is substantial. Audio notifications, where feasible, offer an effective means to augment notification noticeability, corroborating a finding by [20] that audio notifications outperform visual notifications, but the combination of both performs significantly better than either.

Investigating user preferences (RQ3), the results remained consistent with the previous results, with 50% of the participants stating that they prefer notifications on objects with sound. Regarding the PSSUQ survey, the participants generally found the system to be usable and satisfactory. The slightly lower scores on task completion time might stem from the time-sensitive nature of the tasks, intentionally designed to be expedited. The reason for the low average score on the item discussing error messages could be attributed to participants only encountering an error message when presenting food to the customer. This message consisted solely of a brief text indicating an incorrect food item. Providing more specific feedback earlier in the task may improve the score.

While further research is needed in the field of AR HMD notifications, our study has yielded valuable insights. We found that strategically placing visual notifications enhances user performance, with "Notifications on objects with sound" garnering the highest noticeability. Regarding audio notifications, even in environments devoid of ambient noise, the use of sound alone was effective in improving noticeability and was preferred by the participants.

7.1 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.

7.2 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.

8 Conclusion and Future Work

In this study, we focused on the implementation of visual and audio notifications in Augmented Reality (AR) Head Mounted Displays (HMDs), recognizing their pivotal role in crafting user-friendly interfaces. We offered insights into the efficacy of different notification modalities through an empirical evaluation encompassing user performance and preference. Within the immersive context of ARtisan Bistro, an open-source AR sandbox environment, participants engaged in tasks involving cooking and customer service. We systematically manipulated the placement of visual notifications, exploring fixed positions versus locations above objects, and coupled these with localized audio cues.

Our findings unveiled a compelling narrative: the combination of visual notifications above objects, complemented by localized auditory cues, emerged as the most effective condition, outshining alternative configurations in AR HMDs. This approach not only improved user performance but also caught users' attention the most.

This study highlights the importance of judiciously selecting and situating visual and audio notifications within the augmented reality environment. We believe that our findings can help engineers, developers, and others who create 3D interfaces for AR applications consider and design effective notification systems. As we explore mixed-reality tech, these findings guide user-focused innovation for smoother everyday use.

In future research, we aim to examine more effective visual cueing techniques that not only serve as visual notifications but also guide users' attention toward these notifications. Additionally, we intend to transition from a fully virtual kitchen to a setting where physical objects are tangible and physical while retaining the notifications in a virtual form. Another potential area for further investigation is the exploration of different visual notification techniques, such as color-coding and positioning notifications in the FoV, to help users prioritize notifications. For instance, color-coding could be used to indicate the urgency or type of notification, allowing users to quickly assess the importance of incoming alerts. Different shapes, sizes, or animations could also be employed to differentiate notifications based on context or priority. Such enhancements could reduce cognitive load and improve user efficiency in managing multiple notifications in an AR environment.

Supplemental Materials

All supplemental materials are available for reference. These include (1) an Experiment flowchart, (2) the image of the play area, (3) Excel files containing the complete PSSUQ survey and its data and analyses, and (4) a short demo video.

Acknowledgments

This work was supported by National Science Foundation Awards 2327569 and 2238313, and Office of Naval Research grants N00014-24-1-2214 and N00014-21-1-2580.

References

[1] Android and Google. Android notifications. https://developer.android.com/develop/ui/views/notifications. 5

[2] Apple. Apple notifications. https://support.apple.com/en-us/108781. 5

[3] M. Cidota, S. Lukosch, D. Datcu, and H. Lukosch. Comparing the effect of audio and visual notifications on workspace awareness using head-mounted displays for remote collaboration in augmented reality. *Augmented Human Research*, 1:1–15, 2016. 3

[4] K. Crommentuijn and F. 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, p. 255–256. Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1168987.1169042 2, 6

[5] L. A. Elkin, M. Kay, J. J. Higgins, and J. 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, p. 754–768. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3472749.3474784 7

[6] S. Ghosh, L. Winston, N. Panchal, P. Kimura-Thollander, J. Hotnog, D. Cheong, G. Reyes, and G. D. Abowd. Notifivr: exploring interruptions and notifications in virtual reality. *IEEE transactions on visualization and computer graphics*, 24(4):1447–1456, 2018. 3
[7] S. Gould, D. Brumby, A. Cox, V. González, D. Salvucci, and N. 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*, pp. 1189–1192. Association for Computing Machinery, 2012. 1
[8] R. 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. doi: 10.1177/1541931215591373 7
[9] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht. Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE transactions on visualization and computer graphics*, 23(6):1706–1724, 2016. 2
[10] C. Harrison, B. Y. Lim, A. Shick, and S. 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*, pp. 941–944, 2009. 2
[11] S. 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. doi: 10.1177/154193120605000909 7
[12] S. Imamov, D. Monzel, and W. 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)*, pp. 851–858. IEEE, 2020. 1, 3
[13] S. T. Iqbal and E. Horvitz. Notifications and awareness: a field study of alert usage and preferences. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pp. 27–30, 2010. 1, 6
[14] M. G. Kenward and J. H. Roger. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, pp. 983–997, 1997. 7
[15] N. Kern and B. Schiele. Context-aware notification for wearable computing. In *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.*, pp. 223–223. Citeseer, 2003. 2
[16] S.-K. Kim, S.-Y. Kim, and H.-B. 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. doi: 10.1155/2016/5718580 2
[17] I. Koch, E. Poljac, H. Müller, and A. 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. 1
[18] K. Kozuch. Apple glasses updates. `https://www.tomsguide.com/news/apple-glasses`. Published: 2021-09-30. 3
[19] K. Kushlev, J. Proulx, and E. 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, p. 1011–1020. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2858036.2858359 2
[20] M. J. Lazaro, S. Kim, J. Lee, J. Chun, and M.-H. Yun. Interaction Modalities for Notification Signals in Augmented Reality. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, pp. 470–477. ACM, Montréal QC Canada, Oct. 2021. doi: 10.1145/3462244.3479898 3, 10
[21] H. Lee and W. 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)*, pp. 584–585. IEEE, 2022. 2, 3, 6
[22] H. Lee and W. 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)*, pp. 519–529, Mar. 2023. doi: 10.1109/VR55154.2023.00067 3
[23] J. D. Lee, J. D. Hoffman, and E. Hayes. Collision warning design to mitigate driver distraction. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pp. 65–72, 2004. 2, 6
[24] J. 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. doi: 10.1177/154193129203601617 9
[25] A. Lucero and A. 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*, pp. 1–10, 2014. 3
[26] C. S. U. NUILAB. Artisan bistro github repo. `https://github.com/NuiLab/ARtisan-Bistro`. 4
[27] S. Ohly and L. Bastin. Effects of task interruptions caused by notifications from communication applications on strain and performance. *Journal of Occupational Health*, 65(1):e12408, 06 2023. doi: 10.1002/1348-9585.12408 2
[28] C. L. Paul, A. Komlodi, and W. Lutters. Interruptive notifications in support of task management. *International Journal of Human-Computer Studies*, 79:20–34, 2015. 2
[29] L. Plabst, S. Oberdörfer, F. R. Ortega, and F. 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*, pp. 1–11. ACM, Online CA USA, Dec. 2022. doi: 10.1145/3565970.3567701 3, 10
[30] A. Raikwar, L. Plabst, and F. 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)*, pp. 909–910. IEEE, 2022. 2, 3, 4
[31] R. Rzayev, S. Korbely, M. Maul, A. Schark, V. Schwind, and N. Henze. Effects of position and alignment of notifications on ar glasses during social interaction. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, pp. 1–11, 2020. 1
[32] R. Rzayev, P. W. Woźniak, T. Dingler, and N. 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*, pp. 1–9, 2018. 3
[33] K. Salazar. Indicators, validations, and notifications: Pick the correct communication option. `https://www.nngroup.com/articles/indicators-validations-notifications/`. Published: 2015-07-25. 2
[34] M. Team. Met quest pro. `https://www.meta.com/quest/quest-pro/`. 2, 3
[35] X. Team. Xiaomi smart glasses. `https://blog.mi.com/en/2021/09/14/xiaomi-unveils-xiaomi-smart-glasses/`. Published: 2021-09-14. 3
[36] J. D. Upshaw, C. E. Stevens Jr, G. Ganis, and D. 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. 2
[37] M. Wallmyr, T. A. Sitompul, T. Holstein, and R. Lindell. Evaluating mixed reality notifications to support excavator operator awareness. In *IFIP Conference on Human-Computer Interaction*, pp. 743–762. Springer, 2019. 2
[38] A. C. Warden, C. D. Wickens, D. Mifsud, S. Ourada, B. A. Clegg, and F. 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. doi: 10.1177/1071181322661260 5
[39] D. Weber, S. Mayer, A. Voit, R. Ventura Fierro, and N. Henze. Design guidelines for notifications on smart tvs. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, pp. 13–24, 2016. 2
[40] C. D. Wickens, J. S. McCarley, and R. S. Gutzwiller. *Applied attention theory*. CRC press, 2022. 2, 6
[41] J. O. Wobbrock, L. Findlater, D. Gergle, and J. 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, p. 143–146. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963 7

[42] J. Woodward and J. Ruiz. Analytic review of using augmented reality for situational awareness. *IEEE Transactions on Visualization and Computer Graphics*, 29(4):2166–2183, 2023. doi: 10.1109/TVCG.2022.3141585 3

[43] A. Zenner, M. Speicher, S. Klingner, D. Degraen, F. Daiber, and A. Krüger. Immersive notification framework: adaptive & plausible notifications in virtual reality. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–6, 2018. 3, 4, 5

Chapter 6

Notification Interaction in AR

6.1 Unimodal Notification Interaction and Display Methods

In our first experiment as well as other related work [60], participants expressed the wish to dismiss the notifications, so next, we investigated notification interaction [82]. Moreover, in the first experiment, there was only ever one notification at a time, so we also set out to research the display of multiple notifications[3]. We used the same virtual cooking task as in the previous paper. The notifications in that environment could be interacted with using unimodal interaction, in the form of a button on each notification. Interacting with the button caused unimportant informational notifications to be dismissed and important notifications relating to the task to advance the task. The following interaction techniques were used in the experiment:

- **Voice:** Users say the voice command "Message" plus the unique alphanumeric label on the corresponding button.
- **Gaze and Dwell:** Users look at the button for 800ms, after which the button activates.
- **Touch:** Users can push the button with their index fingers.

In our experiment, we also studied how to store multiple notifications at once, so they can be attended at a later time. Notifications stayed in a display-stabilized bottom-center position for 8 seconds, after which they moved into a notification list that was either placed in the environment (world-stabilized) or attached to the left wrist, appearing when the user faced their palm (body-stabilized).

[3] This work was published as Lucas Plabst, Aditya Raikwar, Sebastian Oberdörfer, Francisco R. Ortega, and Florian Niebling. “Exploring Unimodal Notification Interaction and Display Methods in Augmented Reality.” *In 29th ACM Symposium on Virtual Reality Software and Technology*, 1–11. Christchurch, New Zealand: ACM, 2023. HTTPS://DOI.ORG/10.1145/3611659.3615683. © Plabst | ACM 2023. This is the author’s version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in 29th ACM Symposium on Virtual Reality Software and Technology, http://dx.doi.org/10.1145/3611659.3615683.

Exploring Unimodal Notification Interaction and Display Methods in Augmented Reality

LUCAS PLABST, HCI Group, University of Würzburg, Germany
ADITYA RAIKWAR, Computer Science & NUILAB, Colorado State University, USA
SEBASTIAN OBERDÖRFER, HCI Group, University of Würzburg, Germany
FRANCISCO ORTEGA, Computer Science & NUILAB, Colorado State University, USA
FLORIAN NIEBLING, Advanced Media Institute, TH Köln, Germany

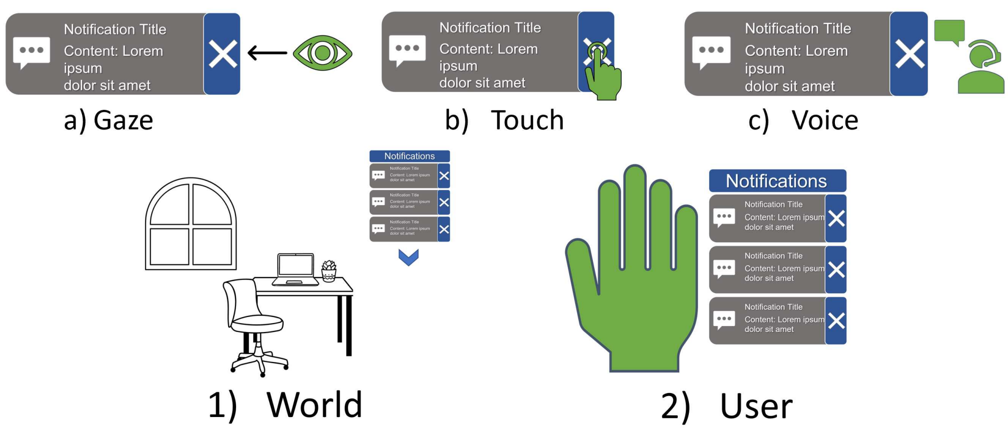


Fig. 1. Techniques for interaction notification: a) Gaze, b) Touch, c) Voice Display methods for notification lists: 1) World 2) User

As we develop computing platforms for augmented reality (AR) head-mounted display (HMDs) technologies for social or workplace environments, understanding how users interact with notifications in immersive environments has become crucial. We researched effectiveness and user preferences of different interaction modalities for notifications, along with two types of notification display methods. In our study, participants were immersed in a simulated cooking environment using an AR-HMD, where they had to fulfill customer orders. During the cooking process, participants received notifications related to customer orders and ingredient updates. They were given three interaction modes for those notifications: voice commands, eye gaze and dwell, and hand gestures. To manage multiple notifications at once, we also researched two different notification list displays, one attached to the user's hand and one in the world. Results indicate that participants preferred using their hands to interact with notifications and having the list of notifications attached to their hands. Voice and gaze interaction was perceived as having lower usability than touch.

CCS Concepts: • **Human-centered computing → Mixed / augmented reality**; *User studies*; Interaction techniques.

Additional Key Words and Phrases: augmented reality, interaction, eye gaze, voice commands, notifications, display methods

ACM Reference Format:
Lucas Plabst, Aditya Raikwar, Sebastian Oberdörfer, Francisco Ortega, and Florian Niebling. 2023. Exploring Unimodal Notification Interaction and Display Methods in Augmented Reality. In *29th ACM Symposium on Virtual Reality Software and Technology (VRST 2023), October 9–11, 2023, Christchurch, New Zealand*. ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3611659.3615683

1 INTRODUCTION

With advancing augmented reality (AR) technology, a future where humans predominantly use these technologies for computing is becoming more likely. Part of daily computing is made up of notifications for events. Interruptions caused by notifications greatly influence attention and responsiveness of users, as these notifications force users to react to newly presented information which might be in conflict with their concentration on a task. In previous studies, Czerwinski et al. have shown disruptive effects of notifications on ongoing computing tasks [13]. Mitigating these disruptions involves considering e.g. timing, placement, and visualization methods of notifications, as users decide how to react based on perceived importance and urgency of the notification information [39].

Research on the presentation of and interaction with notifications has been most prominent on desktop and especially on mobile systems, much less work has been done in Virtual Reality (VR) and AR head-mounted displays (HMDs). In VR, Hsieh et al. found that overlapping use of modalities for delivering alerts, the display locations, and a requirement that the display be moved for notifications to be seen affected the suitability of notifications [22]. Notification presentation and placement also affect response time, noticeability, distraction, and intrusiveness in VR [54], effects that have also been shown to occur in AR [51]. Multimodal presentation of notifications in AR has been explored by Lazaro et al. [31], with the premise that notifications in AR systems must be designed so they capture attention of the user and let them respond to them without disrupting the primary task. They note that in their study, speech input modality was used more frequently over the gesture input modality when acknowledging or confirming [...] notification[s]. The strategy of notification placement seems to be highly dependent on the type of main task, in particular, wrist-attached notifications are beneficial only when hands are in the field-of-view and there is a high amount of interaction with virtual content [51].

The research conducted presents quantitative and qualitative results of a user study comparing different methods of notification placement and interaction modality. We compare world-registered to hand-attached notification placement, as well as gaze-, touch-, and voice-based notification confirmation. Our main task consists of a cooking task that requires frequent interaction with virtual content. Orders as well as tasks necessary to fulfill an order are presented as notifications that have to be acknowledged or dismissed to advance to the next stage of the cooking process, requiring users to interrupt their main task to perform an additional activity.

The **Objective of this research** was to look at correlations between methods for notification placement and interaction methods using different modalities in highly interactive AR environments. We focused on effects of placement and interaction modality on task performance and notification perception. Primary contributions of our research lie in **(1)** gaining a comprehensive understanding of different input modalities for AR notifications and their implications for interaction, and **(2)** exploring and evaluating diverse display techniques for notification storage solutions. By delving into these aspects, we aim to create a foundation for advancements in AR notification design and facilitate seamless integration of notifications into immersive AR environments.

2 RELATED WORK

2.1 Notifications

A notification is a visual cue, auditory signal, or haptic alert generated by an application or service that relays information to a user outside the current focus of attention [24]. Research suggests that individuals receive an average of around 80 notifications per day, with some receiving even higher volumes, reaching up to 200 [2]. Smartphones dominate internet access, contributing to over 60% of internet traffic [38]. As smartphones have become the primary device of computing in their day-to-day life, notifications create a strong emotional response from people. Pielot et al. [47] found that disabling notifications for a day caused participants to feel much less distracted and more productive without them, while simultaneously making them feel worried about missing important information or less connected with their social network. Other studies have also indicated that notifications can disrupt attention-demanding tasks and negatively impact performance [56], or that not receiving notifications can lead to increased frustration and potentially reduce productivity [29].

In response, various approaches have been explored to mitigate attentional costs of notifications. Context-aware delivery systems tailor notifications based on the user's current situation [45]. Grouping notifications into smaller batches and delivering them multiple times throughout the day has been proposed as a strategy to manage their impact [14]. However, certain types of notifications, such as phone call alerts or time-critical safety alerts in critical systems, rule this out as they necessitate immediate delivery. In such cases, HMD use for notification delivery has been explored. It can enhance spatial awareness with minimal impact on performance compared to smartphone-based notifications [43].

2.2 Notifications in 3D/AR

AR has been defined as the supplementation of a real-world environment with digital content [8]. To further develop notification systems for AR-HMDs, we need to understand how information can be presented in those types of systems. According to classifications of Billinghurst et al. [10], there are three ways to display content in Augmented Reality environments: (1) head-stabilized: information is fixed to the user's viewpoint; (2) body-stabilized: information is fixed to the user's body; and (3) world-stabilized: information is fixed to real-world locations.

The positioning of notifications in VR has been studied by Rzayev et al. [54]. Their study revealed no universally preferred placement for notifications in all contexts. Instead, the choice of position should be contingent upon the specific context of the notification and the ongoing task of the user. Plabst et. al. [51] researched the impact of notification position on task performance and perception in AR scenarios and found notifications that were placed in the world, or the bottom center of the field-of-view (FOV) performed better and were preferred by users, but that it depends on the context. This was corroborated by Lee et al. [32] that found that a bottom position in the FOV resulted in a significantly higher noticeability and comprehension for both icon- and text-type notifications compared with a top placement in an AR walking task. Similarly, Ghosh et al. [16] explored interruptions and notifications in VR, employing various modalities such as haptics and audio. They derived design guidelines based on their findings and formulated specific questions to evaluate the perception of notifications.

Furthermore, Rzayev et al. [53] examined the effect of AR notifications during social interactions and found that both the wearer of the headset and the conversation partner would prefer receiving notifications on the headset rather than a smartphone. In a different scenario involving everyday activities like walking and performing pedestrian navigation tasks in a busy city center, Lucero et al. [37] developed and studied notifications on an AR headset. They employed a minimal user interface and a discrete thumb touchpad device for controlling notifications. Their findings indicated that participants faced minimal difficulties in managing notifications while being exposed to potential hazards in an urban environment. However, the increased display of AR content can negatively impact task performance due to clutter [15]; therefore, it should be ensured to not overload the user with notifications.

2.3 AR Interaction

Current AR-HMDs provide a variety of input methods, including controllers, hand gestures or eye-tracking. However, not all methods are suitable for every situation. Direct-touch screen interfaces have been used extensively in research [42] and have shown to be quick for users to learn and engage with [52].

Surale et al. used the advantages of a multi-touch tablet to create an interactive device in VR to perform complex tasks [57]. They exploited the tablet's precise touch input, and metaphorical associations (using the edge as a knife) to make a more intuitive and functional interaction device. Zhang et al. developed a system that used the human body's ability to transfer electrical signals to detect touch input. Combining it with computer vision using headset cameras, they were able to simulate an interface similar to a touch screen with high reliability on the palm of the users [61]. SymbiosisSketch, created by Arora et al. is a hybrid sketching system enabling sketching in 3D space using a mid-air pen as well as in a tablet [7]. Zhu et al. [62] also described different interaction examples using a touch interface and mixed reality.

Satriadi et al. [55] presented a horizontal map navigation system using midair gestures. The gestures used were primarily pinch and move kind, but the application for the gestures were designed to counter different problems, e.g. for manipulation tasks, indirect input to mimic direct manipulation was used to decrease occlusion, and both unimanual and bimanual input were supported. Piumsomboon et al. [50] compared pure gesture based interaction using their G-Shell technique vs gesture-speech multimodal interaction in AR to perform tasks like selection, movement in 3D space, scaling, pushing and flinging. They observed that both pure gesture based interaction and multimodal speech gesture combination had their perks in different context and suggested a combination of both for the best performance. Building on the idea of researching what may be the best interaction techniques for AR, Williams et al. [58–60] conducted elicitation studies to understand what types of gestures and speech prompts are intuitive to most users and provided design recommendations based on their findings.

Eye-tracking has been identified as a viable input method for AR-systems [23] and has been studied by various researchers. Blattgerste et. al. [11] investigated head-gaze and eye-gaze approaches for dwell-time-based target selection tasks in VR and AR and found that eye-gaze outperformed head-gaze in nearly every metric. They also recommend using eye-gaze specifically for AR, where the user's hands might be preoccupied. Parisay et. al. [44] showed that a unimodal dwell-time method outperformed other multi-modal eye-tracking techniques in a target selection task. Relating to specifically notifications, Kosch et al. [28] investigated interacting with notifications during a cycling task. They found that participants preferred using a combination of eye-tracking and a physical button for target selection, but made more mistakes than with an eye-tracking dwell time selection. They partly attributed this to the robust and extended dwell time of 1.8 seconds.

Looking more at general information interfaces, Lu et al. [36] developed an interface that enables users to access concise information in their peripheral vision using different glancing methods, including an eye-tracking based approach.

Another interaction that has been proposed is speech input. Li et. al. [34] researched AR-interaction in a pilot's cockpit and found voice commands improve the perceived workload and situational awareness significantly. In a comparison study by Lee et. al. [33], they found that speech commands in AR worked well for descriptive tasks and that a multimodal voice-gesture interaction did not improve efficiency over the speech commands.

Based on the current state of literature, speech, hand gesture, and eye-tracking interfaces are the most commonly used interaction modalities in AR.

3 METHODS

Notifications are usually considered secondary to a user's primary task, often requiring triage for later time [48]. Unlike smartphone notifications, immersive AR-HMDs can interrupt both digital and physical activities. In a future where AR-HMDs are a pervasive technology, users might not need to specifically be using their headsets when they decide to engage with a notification. While a smartphone can simply be left in the pocket, due to the immersive nature of it, this might not be an option for AR. While there are similarities between AR and VR, we believe that AR-HMD's have the potential to become a pervasive technology like smartphones are now, while VR will stay a technology for specific uses like training or gaming, hence the decision to research AR. To better understand how we can use and interact with notifications in augmented reality environments, we conducted an experiment and set out to answer the three following research questions. **RQ1:** How does the interaction modality of notifications influence task performance and perception of the notifications? **RQ2:** How does the display of the list for multiple notifications influence task performance and perception of the notifications? **RQ3:** Is there any statistical connection between the interaction modality of notifications and the display of the multiple notification list?

Participants were instructed to play a cooking game on the optical see-through AR HMD Hololens 2, during which they received notifications. The headset features a resolution of 1440x936 pixels per eye with a field of view of 43° horizontal, 29° vertical, and 52° diagonal. The cooking environment was developed using Unity Engine 2022.2.10f1, using Microsoft's Mixed Reality Toolkit (MRTK) v2.8.3. The HMDs eye-tracking is refreshed at 30Hz and is predicted to be within 1.5° visual angle around the actual target,according to the manufacturer. These measures have been confirmed by Kapp et al. [27] which found the eye-tracking to be more precise than previously reported. Participants also went through the calibration process by the Hololens before starting the experiment. In the tutorial they were required to test the eye-tracking in order to proceed, confirming the calibration. Using an optical-see through headset, allows the users to see their hands and bodies in real time and also improves depth perception in comparison to VR-headsets [49].

3.1 Experiment Task

In this cooking environment, users were tasked with completing various customer food orders. This cooking environment could act as a metaphor for several different real-world tasks, where the user can be under high- or low-stress, can multitask or only take care of one thing at a time, and can be stationary or moving, depending on the configured layout of the kitchen. For this experiment, the kitchen environment was used as a stand-in for professional and stressful tasks such as work on construction sites or emergency health care, where multiple things simultaneously required user's attention, while being somewhat time-constrained. Overall, the environment's size is ~3 meters by ~3 meters. The cooking environment consists of four stations. Every station (except the station where the customers arrive) consisted of three parts: 1) A "cooking device" for the food; 2) an ingredient supply station in the form of blue plates that spawn the necessary ingredients; and 3) a preparation board on which the finished ingredients shall be assembled. Food and ingredients can be grabbed with each hand, utilizing the full hand-tracking capabilities of the Hololens 2. A trash can is also present where unwanted food can be discarded. A blue hand mesh was placed around the user's real hand to avoid occlusion through the virtual objects. Participants were instructed to grab virtual objects just like physical objects. Grabbing the objects required the index finger and thumb to make contact, so the system was very flexible, as it accommodated several different ways the users could grab objects.

At the **Customer Station**, up to three customers may order food and wait for it. Every customer is represented by a random low-fidelity human avatar with some small idle movements (like looking around or just breathing). In front of every customer, a red tray is placed on the counter, where the prepared food must be served. A notification is sent to the user when a new customer arrives. When the order is accepted by the user (by interacting with the notification), a two-minute timer is displayed behind the food tray, indicating the time left to complete the order.

At the **Burger Station**, the user can take infinite ingredients from the blue spawning plates and put them on the preparation board. A burger always consists of at least a bottom bun, a grilled patty, and a top bun, with up to three other random ingredients. The burger station also features a grill where a burger patty must be cooked. The patty starts out raw and changes color to reflect its cooked state after 10 seconds on the grill. The user also receives a notification that the patty is ready. If it is left on the grill for another 40 seconds, it will burn and not be accepted by the customer.

At the **Pizza Station**, pizza ingredients are also placed on blue spawning plates. Each pizza consists of a pizza base that already has marinara sauce and cheese on it, with three other random ingredients. When the pizza base is assembled with the ingredients, it has to be placed in the pizza oven. After 10 seconds, the pizza changes its color and the user is sent a notification that the pizza is finished. If it is left in the oven for another 40 seconds, it will burn and not be accepted by the user.

The **Coffee Station** features a filter coffee machine with a coffee pot in it. The user has to turn on the machine by pressing a big button on the front. Coffee will start flowing if the pot is inserted into the machine. When enough coffee for a single cup is brewed by the machine, a notification is sent to the user. There is also a gauge beside the machine indicating how many cups of coffee are in the pot. The user can take a cup and pour coffee from the pot to the cup. When the cup is full, a lid appears on the cup, indicating that the coffee is ready to be served.

3.2 Notifications

The notifications were designed with a rectangular form (see Figure 3), resembling the alerts most commonly seen on mobile and desktop operating systems. Each featured a bold title showing the source of the notification. Underneath was a text block with the content of the notification. On the left side, a white icon was displayed, relating to the station and content of the notification, so the user could quickly identify the reason for the notification. For example, if the notification was sent from the coffee machine, a cup of coffee icon was displayed. A sound was played when a notification was delivered, as it has been shown that a combination of visual and audio notifications leads to better performance measures and was generally preferred over unimodal notifications [31].

To ensure high legibility of text, a dark gray background was chosen, along with white text color, keeping in line with the findings of Jankowski et al. [26] and the design recommendations by Microsoft [40]. The font size was set to a minimum of 20pt up to 40pt, putting it above the minimum comfortable range of Microsoft's guidelines for text legibility in AR for near interactions. All notifications automatically aligned to face the user, with the exception of the z-axis(roll), therefore ignoring head tilting.

Plabst et. al. [51] identified that notifications in AR should be placed in the bottom center of the user's FOV for mixed use-cases or when the user is moving around. This was also suggested by Chua et al. [12] for dual-task scenarios that require high noticeability on the secondary stimuli. Hololens 2 system notifications are also displayed using the bottom-center placement. Consequently, in this experiment, notifications are displayed in the bottom-center portion of the display.

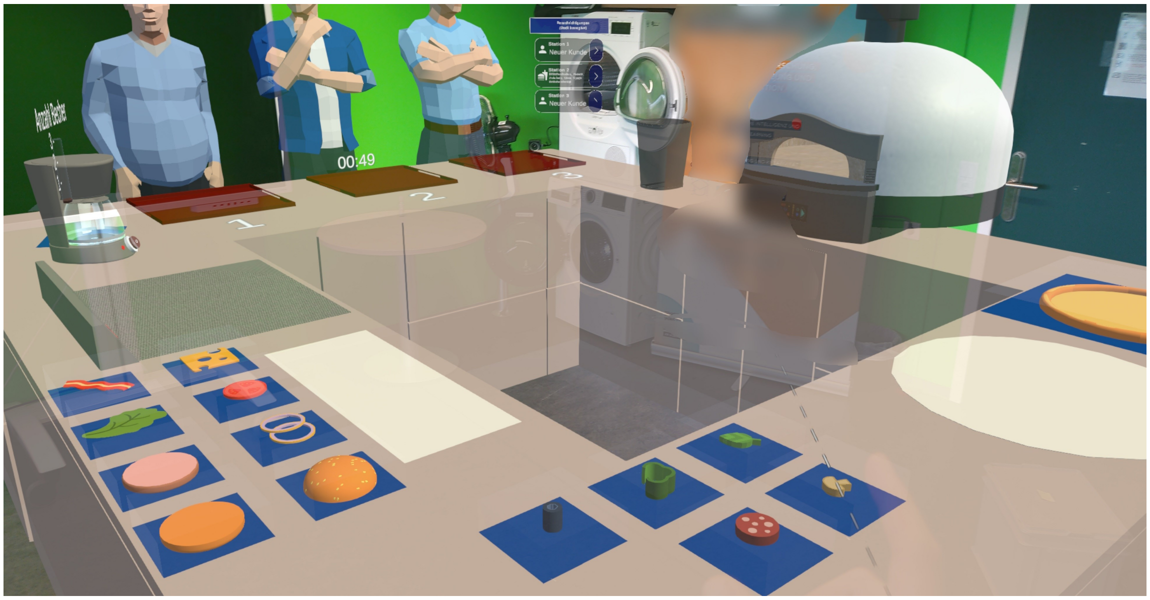


Fig. 2. The virtual cooking environment used in the experiment.

Fig. 3. Notification in the experiment.

When using HMDs, special attention needs to be paid to the vergence-accommodation-conflict [21]. The Hololens 2 display is fixed at an optical distance of approximately two meters and Microsoft recommends not placing any information closer than 40cm [41]. The notifications were spaced 75cm away, so near-interaction with the hands is still possible comfortably. The notifications do not move in-depth, the inter-pupillary distance is measured, and the system is calibrated accordingly, which lessens the potential discomfort caused by the vergence-accommodation-conflict.

3.3 Interaction

According to current user experience design guidelines [3, 4], notifications may be categorized in one of two categories: **actionable notifications**, where the notification is followed by a user action, or **informational notifications**, whose aim is to pass information to the user. In the cooking environment, the notifications from the cooking stations would be considered informational since they did not require follow-up actions, while the order notifications are actionable since they require interaction to advance the task.We specifically chose a combination of actionable and inactionable notifications, since if all notifications are critical for the task, they might pay unrealistically high attention to them, but if none are, they might ignore them all together. We wanted to create a task where all notifications were important to the user, but only some required attendance from the user, in order to create a more realistic scenario. In intensive health care for example, not all alarms from machines are critical, with some being important alerts and others just being status notifications. In a more day-to-day scenario, user might receive notifications for a phone call which requires immediate attention, but might also receive system notifications for i.e. a completed download, which does not require immediate attention.

All notifications in the cooking environment had a button. Upon button activation, informational notifications would be dimissed while actionable notifications would advance to the next stage. The activation was dependent on the interaction method. For this experiment, we specifically chose unimodal interaction methods to first establish a baseline for notification interaction and to further expand on multimodal interaction in future work. Much of the previous work is centered around using a user interface for a certain period of time, and not around short bursts of interaction, like with notifications. Gaze, touch, and voice interaction techniques used in this study are described next.

Gaze. Users had to gaze upon the button and dwell on it for 800ms in order to trigger it. This dwell time lessens the often-found "Midas-touch problem", by requiring intentionality when looking at user interface elements. This dwell time was selected based on the ideal dwell time range of 600ms or 800ms for target selection found by Paulus and Remijn [46]. The button color shifted during the dwell time from dark blue to white, indicating its state. Looking away reset the color and dwell time. Along with voice interaction, this method is completely hands-free. Eye Tracking was implemented using the eye-tracking capabilities of the Hololens 2, along with the eye-tracking features of the MRTK.

Touch. Buttons with the touch interaction can be pressed with either index finger. The button animates to show it is being pressed by compressing and decompressing depending on the push state. Hands are the main interaction type used in the cooking environment, as users prepare the meals with their hands, so touch interaction for notifications could benefit from the lack of modality-switching. As touch-input is the main interaction technique used in smartphones and tablets, taking into account the legacy bias from these devices, users might already feel comfortable with this method [6, 30].

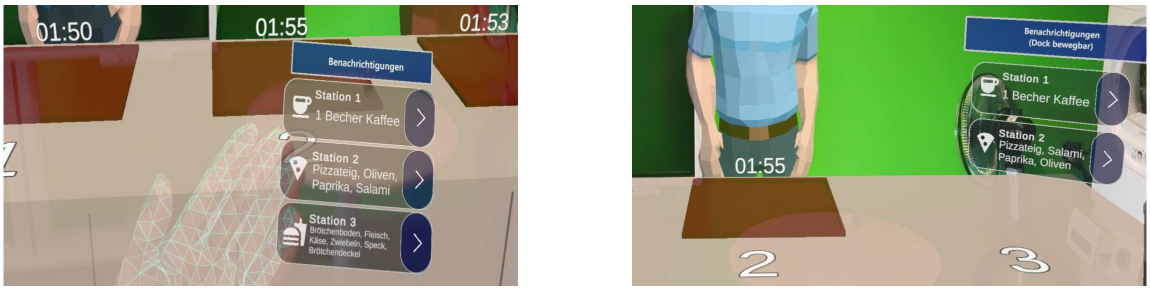
Voice. Every notification with a voice-activated button has a single symbol on the button. When the user says "Message X" (with X being a placeholder for the symbol on the button), the button with the corresponding button would trigger. Notifications outside of the list were labeled with letters, starting with A and then increasing, depending on the amount of other notifications. If the notification is in the list, the buttons are labeled with numbers, starting with 1 and incrementing depending on the position in the list. If a notification in the middle of the list is deleted, the numbers would refresh to form a sequential list again. The button did not need to be in the user's FOV to be activated, as long as the voice command was understood by the device. Voice commands were implemented using the *SpeechInputHandler* from the MRTK and the built-in voice detection from the HMD. When a voice command was detected, a small pop-up would appear confirming the command.

3.4 Notification Lists

Every notification is first displayed in the user's FOV. To avoid clutter in the FOV and to allow later handling of the notification, it disappears into a notification list after 8 seconds of being displayed, acting like the standard behavior on most mobile and desktop operating systems [5, 17]. This ensures that unless a notification is specifically attended to by the user, no notification can disappear. The notification list is sorted chronologically, with new notifications added to the bottom. For this list, we devised two display types based on Billinghurst et. al.'s [10] classification: 1) A body-stabilized list attached to the user's left hand and 2) a floating world-stabilized list above the kitchen counter.

Since notifications themselves were already head-stabilized, we did not want to clutter the FOV and did not also provide a display-stabilized list option. Both list types featured a blue box with white text that read "Notifications."

The **hand list** (see Figure 4a) was attached to the user's left hand. When the user held up their left hand and rotated their palm to face them, the list would appear. If the palm was not oriented towards the user, the list was hidden, making opening the list a deliberate action. This way of accessing notifications is very reminiscent of using a smartwatch to check recent messages. Notifications were scaled down, while still adhering to font size best practices.



(a) The list of notifications fixed to the left hand.

(b) The list of notifications located in the environment.

Fig. 4. Two notification list types: (a) Hand list fixed to the user's left hand and (b) World list placed in the virtual environment.

The **world list** (see Figure 4b) was placed in the room with the other objects in the virtual environment. Alongside the "Notifications" text on the top, it also stated that the list is movable. The list could be grabbed at the blue box and moved to where the user wished. The world list would always face the user to improve legibility independent of the user's position with the exception of the z-axis(roll). Notifications did not change size when moving from the user's FOV to the world list.

3.5 Procedure

The experiment took place in a large room (see Figure 2) measuring ~5x5 meters. The blinds in the room were closed and the lights turned on to control the lighting. The cooking environment was manually placed in the room, guided by a visual marker for precise placement. Upon entering the room, participants were instructed to fill out a demographics questionnaire. They were then introduced to the Hololens 2 headset and instructed on proper wear. Every participant then calibrated the Hololens to their eyes, ensuring optimal clarity and precise eye-tracking. Participants proceeded through a tutorial which involved reading explanations about the experiment and the cooking setting. They would then be shown notifications with every interaction type and list type, ensuring they understood the interaction. Subsequently, an interactive section started, where every cooking station was explained and participants had to prepare every type of food. After completion, they were allowed to continue practicing for a maximum of five minutes or start the experiment independently by pressing a button. Every experiment consisted of six runs, one for every combination of interaction methods and lists. The order of conditions was randomized [1]. During each run, participants were required to complete six customer orders. Customers would always order one type of meal with randomized ingredients and every run required the participants to make two of each food, in a random order. When a new customer appeared, a notification was displayed. To acknowledge the order, participants had to interact with the button on the notification. A two-minute countdown would then be displayed in front of the customer, and the notification text would change to show the ordered food. When the food was prepared, participants had to place the food on the tray in front of the customer and press the button on the notification again. If the food was correct, a "success" tone would play, and the notification would display "correct order." After a few seconds, the notification along with the customer and tray would disappear, making way for a new customer. If the food was wrong, the notification would display "wrong order." This had to be acknowledged by the user by interacting with the button, after which the notification would show the ingredients again. If the correct food was not prepared within the two-minute countdown, the customer would leave without their order. When six customers had been completed, by giving them the correct food or by letting them expire, a notice would be given to the participants to flip the display of the Hololens up and fill out a questionnaire on the laptop. When all six conditions had been completed and the questionnaire had been filled out, the experiment ended.

3.6 Measurements

3.6.1 Performance. During the experiment, all events were logged on the Hololens for further analysis. Using this log we were able to measure several variables relating to task performance of the order fulfillment. We measured the *total time* per experiment run to understand overall performance. We measured the *time needed per customer*, to better account for breaks participants might have taken between orders. When it comes to the process of order preparation, we measured the amount of *wrong orders prepared* and the amount of *customers expired*. Lastly, concerning the notifications, we measured the *time until order was accepted* by the participants.

3.6.2 Perception. After each experiment cycle, participants had to complete a set of questionnaires. To assess overall usability of the notification interaction, the System Usability Scale (SUS) questionnaire [18] was deployed, along with a NASA Task-Load-Index questionnaire (NASA-TLX) [20] used for assessing task load. Hart et. al.[19] found that not weighing sub-scales on the TLX does not impact the results we chose to also not weigh the sub-scales (so-called Raw-TLX). When the experiment ended, participants were asked to rank the interaction techniques and lists by preference, sorting them from highest to lowest and then explain their ranking. They were instructed to specifically evaluate the interaction with the notifications and the type of notification list.

3.7 Participants

Participants were recruited from a pool of university students studying human-computer-systems or media communication. They are required to gather experiment hours for their coursework and were rewarded with 1.25 hours of participation time. In total, 29 participants were recruited (10 Male and 19 female). Age ranged from 19 to 29 years (M = 21.3, SD = 2.6). All either had a normal or corrected-to-normal vision. Of those participants, all 29 stated that they used smartphones and internet daily, and computers daily (25) or weekly (4). 27 had either never used AR-technology before or only in experiments (19 in experiments, 8 never) and 28 participants had experienced virtual reality before (19 in experiments). Additionally, 12 participants said they played video games regularly (at least weekly) and 26 were right-handed.

4 RESULTS

To analyze the results, we used R 4.3.0 and Visual Studio Code 2023 running R-compatible plugins. We calculated a two-way ANOVA to measure the main effects and interaction effects. For every significant effect we found, we used TukeyHSD-tests for pairwise analysis. The assumptions for ANOVA were met.

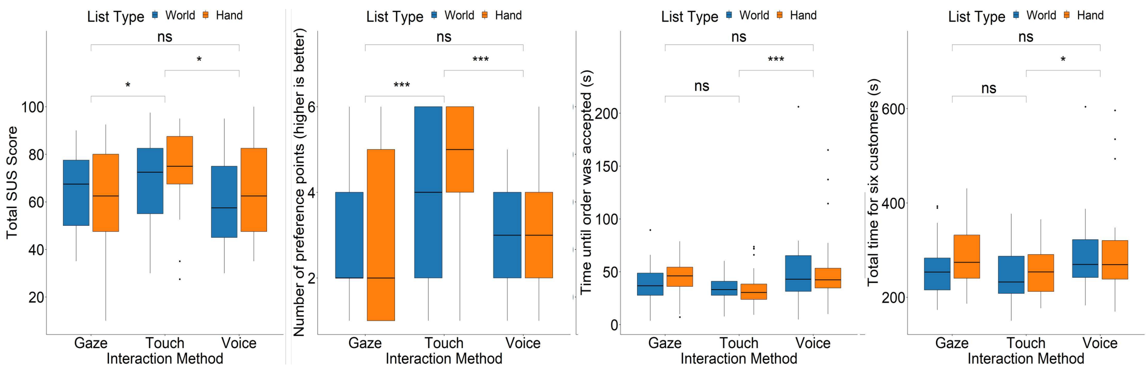


Fig. 5. Results for the SUS-Score, Preference ranking, Time until orders were accepted and Total Time.
***: p < 0.001; **: p < 0.01; *: p < 0.05; ns: p > 0.05

Table 1. Descriptive statistics for significant measurements (SUS, preference score, time until order was accepted and total time.)

Condition	SUS (0-100)			Preference 1 → 6			Order acceptance in s			Total time in s		
	Mean	Median	Sd	Mean	Median	Sd	Mean	Median	Sd	Mean	Median	Sd
Gaze - World	63	68	17	2.8	2	1.5	38	37	18	260	253	60
Gaze - Hand	65	68	18	3	2	1.8	46	46	18	283	274	57
Touch - World	67	70	15	4.1	4	1.8	33	33	14	251	232	56
Touch - Hand	72	71	17	4.9	5	1.3	34	30	18	258	254	52
Voice - World	61	60	15	2.9	3	1.3	50	43	37	290	270	82
Voice - Hand	68	68	20	3.3	3	1.5	53	42	35	301	270	108

4.1 Subjective Measures

System Usability Scale. We found a significant main effect on the SUS-score by Interaction ($F(2, 168) = 4.540$, $p = 0.012$). There was no effect for List ($F(1, 168) = 0.934$, $p = 0.335$) or the interaction between the two ($F(2, 168) = 0.540$, $p = 0.584$).

Pairwise comparisons revealed a significant improvement for *Touch* over *Voice* ($p = 0.029$). A significant improvement was also found for *Touch* over *Gaze* ($p = 0.024$). No significant difference was observed between *Voice* and *Gaze* ($p = 0.753$).

RAW Task Load Index. We found no significant effect of Interaction on the *TLX-score*, ($F(2, 168) = 1.194$, $p = 0.306$). The List showed no significant effect on the *TLX-score*, ($F(1, 168) = 0.023$, $p = 0.879$). The interaction between Interaction and List also did not yield a significant effect on the *TLX-score*, ($F(2, 168) = 1.947$, $p = 0.146$).

Ranking. Participants were asked to rank all conditions according to preference, going from six points (highest) to one(lowest). There was a significant main effect of the Interaction on the *preference* ($F(2, 168) = 18.347$, $p < 0.001$). The List showed a significant effect on *preference* ($F(1, 168) = 5.279$, $p = 0.023$). The interaction between Interaction and List did not reach statistical significance ($F(2, 168) = 0.533$, $p = 0.588$).

Pairwise comparisons revealed a significant improvement for *Touch* over *Gaze* ($p < 0.001$). A significant improvement was found for *Touch* over *Voice* ($p < 0.001$). No significant difference was observed between *Voice* and *Gaze* ($p = 0.75$). A significant improvement of the List was found for *Hand* over *World* ($p = 0.023$).

4.2 Performance

Time until Order was accepted. We found a significant effect of Interaction on the *Time until order was accepted* ($F(2, 161) = 7.340$, $p < .001$). The List variable did not show a significant effect on the Time until order was accepted ($F(1, 161) = 1.121$, $p = 0.291$). The interaction between Interaction and List also did not yield a significant effect on the *Time until order was accepted* ($F(2, 161) = 0.284$, $p = 0.752$).

Pairwise comparisons revealed a significant improvement for *Touch* over *Voice* ($p < .001$). No significant differences between *Touch* and *Gaze* ($p = 0.164$) and *Voice* and *Gaze* ($p = 0.113$) were observed.

Total Time. We found a significant effect of Interaction on *total time* ($F(2, 161) = 3.866$, $p = 0.023$). The List variable did not show a significant effect on *total time* ($F(1, 161) = 1.015$, $p = 0.315$). Interaction between Interaction and List did not yield a significant effect on *total time* ($F(2, 161) = 0.300$, $p = 0.741$).

Pairwise comparisons revealed a significant improvement for *Touch* over *Voice* ($p = 0.0168$). No significant differences were found for *Touch* and *Gaze* ($p = 0.407$), and also *Voice* and *Gaze* ($p = 0.293$).

Time needed per Customer. We found no significant effect of Interaction ($F(2, 161) = 1.092$, $p = 0.338$), List ($F(1, 161) = 0.596$, $p = 0.441$), or the interaction between Interaction and List ($F(2, 161) = 0.951$, $p = 0.389$) on the *time needed per customer*.

Wrong orders prepared. We found no significant effect of Interaction ($F(2, 161) = 1.444$, $p = 0.2390$) List ($F(1, 161) = 2.959$, $p = 0.087$) or the interaction between Interaction and List ($F(2, 161) = 2.465$, $p = 0.088$) on the amount of *wrong orders prepared* . The p-values of the List and the interaction are both approaching significance, possibly suggesting an effect.

Customers expired. We found no significant effect of Interaction ($F(2, 161) = 0.122$, $p = 0.886$),List ($F(1, 161) = 0.638$, $p = 0.425$) or the interaction between Interaction and List ($F(2, 161) = 0.127$, $p = 0.881$) on the amount of *customers expired* .

4.3 Interview Feedback

Participants could give an explanation on their ranking preference. These answers were then structured and analyzed using an affinity diagram [9]. The question specifically was related to how much the users liked the technique that they used to interact with notifications.

Gaze. The main criticism participants had was the perceived reliability of eye-tracking. P3 said that *"with the eye controls I selected something by accident multiple times that I didn't want to."*. Similarly, P29 reported that *"Eye selection was cool but it either took too long for the button to be pressed or it was selected by accident"*, alluding to the next criticism: Gaze- and dwell- taking too long. P13 said that *"selecting with the gaze was comfortable, but the time seemed a bit too long"* and P17 said *"I think that the eye button took too long or was too imprecise"*. This hints at the Midas-touch [25] still being a problem with gaze, even when utilizing an optimal dwell time. Some participants also considered this dwell time of 800ms too long, like P21 that said: *"Eye selection takes too much time and too much concentration"*. Lastly, participants found it strenuous to use, reporting things like *"I thought it was hard to focus the notification with my eyes, I often jumped somewhere else and had to restart"* (P5) or *"With the eye controls I needed to concentrate much more and it was much harder to complete the tasks. It also did not work as reliable"* (P19).

Touch. Justifications for the *Touch* ranking were mostly positive, with P5 stating *"The touch button was the easiest to use, it was generally the fastest and safest - it was easy to keep an overview."* or P24: *"I liked touch the most because I always knew that the thing I wanted to select was actually selected"*. Positive comments focused on ease of use, reliability, and speed of the interaction. Criticism was mentioned in regards to incorrect inputs: *"I was afraid that I would trigger something while grabbing ingredients."* (P1).

Voice. The biggest issue with *Voice* interaction seemed to be the symbols on the notifications, and the resulting voice command. P25 said *"The voice selection was very confusing because I often mixed up the numbers and letters. For example, Message A which was sent by Station 3"* with P24 saying the same. Another issue participants had, was changing of numbers on the notifications when a notification in the list was dismissed. *"With the voice I had the issue that the numbers would refresh when I dismissed one, So I had to wait for the list to refresh before I could dismiss another one. I got confused and then said the wrong number."* (P5).

Hand-List. Almost half of all participants (13/29) stated that they preferred the *Hand-List* display interface because they did not need to turn around towards the list and had the information with them at all times. P6 stated that *"On the hand was always handier because I always had the messages with me"*, and P27 wrote *"The list on the hand was the most practical because I could access it quickly and I did not have to change [the] location for it"*. An issue mentioned by 4 participants was that the list was a little small and could get crowded quickly. P16 stated: *"The hand list was too small and it was straining to look at it. I also needed to tilt my head down for it."*

World-List. Much of the praise for the *Hand-list* was directly critiquing the *World-list*, as participants disliked needing to turn to see the notifications. Still, some participants found it easier to read and said: *"Generally I found the list in the room easier because I had a specific location and it wasn't as 'wobbly' "* (P5). Another participant (P25) found that *"having the list in the room was the most comfortable because you are not overwhelmed by all the displays and have all the info in one spot."*

5 DISCUSSION

The study found multiple significant effects on the Interaction and a main effect on List. Looking closer at the main effects of Interaction, we found that *Touch* performed better than *Voice* or *Gaze*. We were not able to measure any significant difference, where a different Interaction method performed better than *Touch*. It performed equally or better in task performance metrics, featured higher usability, and was chosen as the most preferred option. This answers **RQ1**, by showing that just the interaction modality of notifications alone has an impact on task performance and perception of the notification. Surprisingly, the hands-free solutions did not outperform the *Touch* method. *Voice* took longer in total and in order acceptance time. This leads to the conclusion, that participants did not multitask as much here as in the touch condition, as they did not accept multiple orders quickly but decided to complete them sequentially. A reason for this could be the re-assignment of symbols on the notification for voice selection, which was mentioned in the interview responses. When the notification was dismissed at any other position than the first or last, its number would be reassigned to the following notification in the list. For example, if the third notification in a list with the symbol 3 was dismissed, the notification in position 4 would move up and get the symbol 3 instead of 4. This seemingly caused confusion, as participants wanted to rapidly dismiss notifications, but had to think about what the new number for each notification would now be. To differentiate between notifications in the FOV and in the list, we chose to label them differently, with numbers in the list and letters in the FOV. Lastly, the stations were numbered from one to three, but as the symbol for *Voice* interaction was determined by the list position, there might have been a mismatch between the numbers on the notification. Identifiers for voice commands should be chosen depending on notification context and kept static.

As for Gaze, participants seemed to have the most issues with the reliability of the eye-tracking and the dwell time. Participants specifically mentioned issues with the *Hand-List*, as the buttons were already somewhat small, but were also not static, as the list could be moved with hand-movement. This also resulted in a potential behavior, where the quality of the eye-tracking could have been influenced by the distance the participants extended their arms, as very close proximity to the eye-tracking button could have negatively impacted registration. This could also explain the issues with the dwell time. If the eye-tracking did not instantly register correctly, the dwell time would have been longer than intended.

Regarding Lists, Hand-List did not perform better, but was ranked more preferential as the *World-List*. Answering **RQ2**, we see that the display of the list did not influence task performance, but impacted the participants' preference for the system. The main benefit mentioned for *Hand-List*, was that the user did not have to turn around to look at the notifications, and that information was always within reach. While the *World-List* was moveable, because of the need to move around in the environment for the task, it was still out of reach and sight on many occasions. While we did not measure the number of times the list was moved, based on observation during the experiment and participant's feedback, it seems like most did not move the list to another location. Where the *world list* to be moved, it likely would not have remained in participants line of sight due to the kitchen's station layout and the HoloLens' FOV. Relating to **RQ 3**, we did not find any connection between the interaction modality and the display of the list.

Overall we can conclude, that participants preferred using *Touch*-interaction, performed better with it, and rated it as having a higher usability in the kitchen environment. When it comes to interacting with notifications, we suggest *Touch*-interaction as an easy and reliable method. As users preferred having notifications close to them, we suggest using a body-stabilized way of displaying the list.

Limitations. None of the tasks truly required two hands to complete and it was possible to pause the current task to interact with a notification; however users might sometimes, for example, be carrying something or using instruments in both hands they cannot put down, like in a study conducted by Li et. al. [35], where voice interaction in AR increased pilots' spatial awareness and monitoring performance while also decreasing mental and physical demands. While *Touch* interaction was favorable in this scenario, this might not apply to heavily two-handed tasks. Also, the touch interactions required to complete tasks within this system may have biased participants towards favoring touch interactions with the notifications. The task was seen as a metaphor for other manual labor tasks such as construction or health care, which are also heavily hands-focused, so while we believe our findings would apply to tasks like that, this may not hold true for all types of tasks and requires further investigation. For example, operating machinery where the hands are in constant use could favor other input modalities like voice or gaze. Also, non-touch modalities could for example be better suited for people with motor disabilities or for more mobile scenarios, such as walking and navigating a public space.

Another limitation was the size of the experiment environment, as it was small and although users did need to move within it, they never had to cover a long distance. Therefore, the List in the *World* was always within reach. In larger environments that require more movement, the *world* list may not be viable.

Also, participants did not have to specifically participate in this experiment, but to receive course credit they had to gather experiment hours. This skews the participant pool to a group of college students studying a certain subject.

Additionally, the Hololens does not have a large FOV, with only 50°on its widest side, limiting visibility of content outside of the FOV. Specifically, placement of world-stabilized content like the *World* list might benefit from advancements in AR technology, specifically increases in FOV, as it will be easier to see a wider range of content in the environment. Another limitation was the reliability of the eye-tracking. Kapp et. al. [27] found the eye-tracking of the Hololens 2 comparable to other state-of-the-art mobile eye trackers, so the problems with registration are most likely not a technical limitation of the device, but rather a limitation of the interaction.

Lastly, the simulation of the environment was not entirely realistic, as interaction with ingredients was not like in the real world, where you would for example use a spatula to put patties on the grill. Application of our findings to real-world tasks in an entirely physical environment may produce different results, which is something we want to research in the future.

6 CONCLUSION

We evaluated three different techniques (*Gaze*, *Touch* and *Voice*) to interact with notifications and two display types for multiple notifications (*World* and *Hand*). We found that while participants preferred having the notification list attached to their *Hand*, the numbers did not show any improvements over the list in the *World*. For the interaction, *Touch* was preferred, while also boasting higher usability and task completion time. Future directions may include multimodal interaction techniques and the construction and evaluation of a physical task environment.

ACKNOWLEDGMENTS

This project is funded by: DFG–425868361 (SPP2199),ONR N00014-21-1-2949, ONR N00014-21-1-2580, ONR N00014-23-1-2298.

REFERENCES

[1] Random.org . 2023. RANDOM.ORG - True Random Number Service. https://www.random.org/

[2] Utku Acer, Afra Mashhadi, Claudio Forlivesi, and Fahim Kawsar. 2015. *Energy Efficient Scheduling for Mobile Push Notifications.* Vol. 2. https://doi.org/10.4108/eai.22-7-2015.2260067 Journal Abbreviation: EAI Endorsed Transactions on Energy Web Publication Title: EAI Endorsed Transactions on Energy Web.

[3] alison@uxmag.com. 2020. Designing Notifications for Apps. https://uxmag.com/articles/designing-notifications-for-apps

[4] Apple. [n. d.]. Declaring your actionable notification types. https://developer.apple.com/documentation/usernotifications/declaring_your_actionable_notification_types

[5] Apple. 2022. Use notifications on your iPhone or iPad. https://support.apple.com/en-us/HT201925

[6] Shaikh Shawon Arefin Shimon, Courtney Lutton, Zichun Xu, Sarah Morrison-Smith, Christina Boucher, and Jaime Ruiz. 2016. Exploring Non-touchscreen Gestures for Smartwatches. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, San Jose California USA, 3822–3833. https://doi.org/10.1145/2858036.2858385

[7] Rahul Arora, Rubaiat Habib Kazi, Tovi Grossman, George Fitzmaurice, and Karan Singh. 2018. SymbiosisSketch: Combining 2D & 3D Sketching for Designing Detailed 3D Objects in Situ. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3173574.3173759

[8] Ronald T Azuma and I Malibu. 1997. A Survey of augmented reality. (1997), 31.

[9] Hugh Beyer and Karen Holtzblatt. 1997. *Contextual design: defining customer-centered systems.* Elsevier.

[10] Mark Billinghurst and Hirokazu Kato. 1999. Collaborative Mixed Reality. In *Mixed Reality*, Yuichi Ohta and Hideyuki Tamura (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 261–284. https://doi.org/10.1007/978-3-642-87512-0_15

[11] Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2018. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the Workshop on Communication by Gaze Interaction*. ACM, Warsaw Poland, 1–9. https://doi.org/10.1145/3206343.3206349

[12] Soon Hau Chua, Simon T. Perrault, Denys J. C. Matthies, and Shengdong Zhao. 2016. Positioning Glass: Investigating Display Positions of Monocular Optical See-Through Head-Mounted Display. In *Proceedings of the Fourth International Symposium on Chinese CHI - ChineseCHI2016*. ACM Press, San Jose, USA, 1–6. https://doi.org/10.1145/2948708.2948713

[13] Mary Czerwinski, Edward Cutrell, and Eric Horvitz. 2000. Instant messaging and interruption: Influence of task type on performance. In *OZCHI 2000 conference proceedings*, Vol. 356. 361–367.

[14] Nicholas Fitz, Kostadin Kushlev, Ranjan Jagannathan, Terrel Lewis, Devang Paliwal, and Dan Ariely. 2019. Batching smartphone notifications can improve well-being. *Computers in Human Behavior* 101 (Dec. 2019), 84–94. https://doi.org/10.1016/j.chb.2019.07.016

[15] Jonathan Flittner*, John Luksas, and Joseph L. Gabbard. 2020. Predicting User Performance in Augmented Reality User Interfaces with Image Analysis Algorithms. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 64, 1 (Dec. 2020), 2108–2112. https://doi.org/10.1177/1071181320641511 Publisher: SAGE Publications Inc.

[16] Sarthak Ghosh, Lauren Winston, Nishant Panchal, Philippe Kimura-Thollander, Jeff Hotnog, Douglas Cheong, Gabriel Reyes, and Gregory D. Abowd. 2018. NotifiVR: Exploring Interruptions and Notifications in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (April 2018), 1447–1456. https://doi.org/10.1109/TVCG.2018.2793698 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[17] Google. 2023. Notifications overview. https://developer.android.com/develop/ui/views/notifications

[18] Rebecca A. Grier, Aaron Bangor, Philip Kortum, and S. Camille Peres. 2013. The System Usability Scale: Beyond Standard Usability Testing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 57, 1 (Sept. 2013), 187–191. https://doi.org/10.1177/1541931213571042 Publisher: SAGE Publications Inc.

[19] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (Oct. 2006), 904–908. https://doi.org/10.1177/154193120605000909 Publisher: SAGE Publications Inc.

[20] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9 ISSN: 0166-4115.

[21] David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. 2008. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision* 8, 3 (March 2008), 33. https://doi.org/10.1167/8.3.33

[22] Ching-Yu Hsieh, Yi-Shyuan Chiang, Hung-Yu Chiu, and Yung-Ju Chang. 2020. Bridging the Virtual and Real Worlds: A Preliminary Study of Messaging Notifications in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. https://doi.org/10.1145/3313831.3376228

[23] Tobias Höllerer and Steven K. Feiner. 2003. 9 Chapter Nine Mobile Augmented Reality. https://www.semanticscholar.org/paper/9-Chapter-Nine-Mobile-Augmented-Reality-H%C3%B6llerer-Feiner/8dbdd436b2a6aa43dee83dbd21dcaf6970e98a0d

[24] Shamsi T. Iqbal and Eric Horvitz. 2010. Notifications and awareness: a field study of alert usage and preferences. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work - CSCW '10*. ACM Press, Savannah, Georgia, USA, 27. https://doi.org/10.1145/1718918.1718926

[25] Robert J. K. Jacob. 1991. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems* 9, 2 (April 1991), 152–169. https://doi.org/10.1145/123078.128728

[26] Jacek Jankowski, Krystian Samp, Izabela Irzynska, Marek Jozwowicz, and Stefan Decker. 2010. Integrating Text with Video and 3D Graphics: The Effects of Text Drawing Styles on Text Readability. (2010), 10.

[27] Sebastian Kapp, Michael Barz, Sergey Mukhametov, Daniel Sonntag, and Jochen Kuhn. 2021. ARETT: Augmented Reality Eye Tracking Toolkit for Head Mounted Displays. *Sensors* 21, 6 (March 2021), 2234. https://doi.org/10.3390/s21062234

[28] Thomas Kosch, Andrii Matviienko, Florian Müller, Jessica Bersch, Christopher Katins, Dominik Schön, and Max Mühlhäuser. 2022. NotiBike: Assessing Target Selection Techniques for Cyclist Notifications in Augmented Reality. *Proceedings of the ACM on Human-Computer Interaction* 6, MHCI (Sept. 2022), 1–24. https://doi.org/10.1145/3546732

[29] Kostadin Kushlev, Jason Proulx, and Elizabeth W. Dunn. 2016. "Silence Your Phones": Smartphone Notifications Increase Inattention and Hyperactivity Symptoms. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 1011–1020. https://doi.org/10.1145/2858036.2858359

[30] Anne Köpsel and Nikola Bubalo. 2015. Benefiting from legacy bias. *Interactions* 22, 5 (Aug. 2015), 44–47. https://doi.org/10.1145/2803169

[31] May Jorella Lazaro, Sungho Kim, Jaeyong Lee, Jaemin Chun, and Myung-Hwan Yun. 2021. Interaction Modalities for Notification Signals in Augmented Reality. In *Proceedings of the 2021 International Conference on Multimodal Interaction*. ACM, Montréal QC Canada, 470–477. https://doi.org/10.1145/3462244.3479898

[32] Hyunjin Lee and Woontack Woo. 2023. 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)*. 519–529. https://doi.org/10.1109/VR55154.2023.00067 ISSN: 2642-5254.

[33] Minkyung Lee, Mark Billinghurst, Woonhyuk Baek, Richard Green, and Woontack Woo. 2013. A usability study of multimodal input in an augmented reality environment. *Virtual Reality* 17, 4 (Nov. 2013), 293–305. https://doi.org/10.1007/s10055-013-0230-0

[34] Wen-Chin Li, Jingyi Zhang, Samuel Court, Peter Kearney, and Graham Braithwaite. 2022. The influence of augmented reality interaction design on Pilot's perceived workload and situation awareness. *International Journal of Industrial Ergonomics* 92 (Nov. 2022), 103382. https://doi.org/10.1016/j.ergon.2022.103382

[35] Yuan Li, Feiyu Lu, Wallace S Lages, and Doug Bowman. 2019. Gaze Direction Visualization Techniques for Collaborative Wide-Area Model-Free Augmented Reality. In *Symposium on Spatial User Interaction*. ACM, New Orleans LA USA, 1–11. https://doi.org/10.1145/3357251.3357583

[36] Feiyu Lu, Shakiba Davari, Lee Lisle, Yuan Li, and Doug A. Bowman. 2020. Glanceable AR: Evaluating Information Access Methods for Head-Worn Augmented Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Atlanta, GA, USA, 930–939. https://doi.org/10.1109/VR46266.2020.00113

[37] Andrés Lucero and Akos Vetek. 2014. NotifEye: using interactive glasses to deal with notifications while walking in public. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*. ACM, Funchal Portugal, 1–10. https://doi.org/10.1145/2663806.2663824

[38] Connor Mahon. 2023. 20 Mobile Vs. Desktop Usage Statistics [2023]: What Percentage Of Internet Traffic Is Mobile? https://www.zippia.com/advice/mobile-vs-desktop-usage-statistics/ Section: Research.

[39] Daniel C McFarlane. 2002. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-computer interaction* 17, 1 (2002), 63–139.

[40] Microsoft. [n. d.]. Typography - Mixed Reality. https://docs.microsoft.com/en-us/windows/mixed-reality/design/typography

[41] Microsoft. 2021. Comfort - Mixed Reality | Microsoft Learn. https://learn.microsoft.com/en-us/windows/mixed-reality/design/comfort

[42] Thomas Moeslund and Lau Nørgaard. 2003. A Brief Overview of Hand Gestures Used in Wearable Human Computer Interfaces. (Jan. 2003).

[43] Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2014. Managing Mobile Text in Head Mounted Displays: Studies on Visual Preference and Text Placement. 18, 2 (2014), 12.

[44] Mohsen Parisay, Charalambos Poullis, and Marta Kersten-Oertel. 2021. EyeTAP: Introducing a multimodal gaze-based technique using voice inputs with a comparative analysis of selection techniques. *International Journal of Human-Computer Studies* 154 (Oct. 2021), 102676. https://doi.org/10.1016/j.ijhcs.2021.102676

[45] Chunjong Park, Junsung Lim, Juho Kim, Sung-Ju Lee, and Dongman Lee. 2017. Don't Bother Me. I'm Socializing! A Breakpoint-Based Smartphone Notification System. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 541–554. https://doi.org/10.1145/2998181.2998189

[46] Yesaya Tommy Paulus and Gerard Bastiaan Remijn. 2021. Usability of various dwell times for eye-gaze-based object selection with eye tracking. *Displays* 67 (April 2021), 101997. https://doi.org/10.1016/j.displa.2021.101997

[47] Martin Pielot and Luz Rello. 2017. Productive, anxious, lonely: 24 hours without push notifications. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, Vienna Austria, 1–11. https://doi.org/10.1145/3098279.3098526

[48] Martin Pielot, Amalia Vradi, and Souneil Park. 2018. Dismissed!: a detailed exploration of how mobile phone users handle push notifications. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, Barcelona Spain, 1–11. https://doi.org/10.1145/3229434.3229445

[49] Jiamin Ping, Yue Liu, and Dongdong Weng. 2019. Comparison in Depth Perception between Virtual Reality and Augmented Reality Systems. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Osaka, Japan, 1124–1125. https://doi.org/10.1109/VR.2019.8798174

[50] Thammathip Piumsomboon, David Altimira, Hyungon Kim, Adrian Clark, Gun Lee, and Mark Billinghurst. 2014. Grasp-Shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 73–82. https://doi.org/10.1109/ISMAR.2014.6948411

[51] Lucas Plabst, Sebastian Oberdörfer, Francisco Raul Ortega, and Florian Niebling. 2022. 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*. ACM, Online CA USA, 1–11. https://doi.org/10.1145/3565970.3567701

[52] R. L. Potter, L. J. Weldon, and B. Shneiderman. 1988. Improving the accuracy of touch screens: an experimental evaluation of three strategies. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '88*. ACM Press, Washington, D.C., United States, 27–32. https://doi.org/10.1145/57167.57171

[53] Rufat Rzayev, Susanne Korbely, Milena Maul, Alina Schark, Valentin Schwind, and Niels Henze. 2020. Effects of Position and Alignment of Notifications on AR Glasses during Social Interaction. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. ACM, Tallinn Estonia, 1–11. https://doi.org/10.1145/3419249.3420095

[54] Rufat Rzayev, Sven Mayer, Christian Krauter, and Niels Henze. 2019. Notification in VR: The Effect of Notification Placement, Task and Environment. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Barcelona Spain, 199–211. https://doi.org/10.1145/3311350.3347190

[55] Kadek Ananta Satriadi, Barrett Ens, Maxime Cordeil, Bernhard Jenny, Tobias Czauderna, and Wesley Willett. 2019. Augmented Reality Map Navigation with Freehand Gestures. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 593–603. https://doi.org/10.1109/VR.2019.8798340

[56] Cary Stothart, Ainsley Mitchum, and Courtney Yehnert. 2015. The attentional cost of receiving a cell phone notification. *Journal of Experimental Psychology: Human Perception and Performance* 41, 4 (Aug. 2015), 893–897. https://doi.org/10.1037/xhp0000100

[57] Hemant Bhaskar Surale, Aakar Gupta, Mark Hancock, and Daniel Vogel. 2019. TabletInVR: Exploring the Design Space for Using a Multi-Touch Tablet in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300243

[58] Adam S. Williams, Jason Garcia, and Francisco Ortega. 2020. Understanding Multimodal User Gesture and Speech Behavior for Object Manipulation in Augmented Reality Using Elicitation. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (Dec. 2020), 3479–3489. https://doi.org/10.1109/TVCG.2020.3023566 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[59] Adam S. Williams and Francisco R. Ortega. 2020. Understanding Gesture and Speech Multimodal Interactions for Manipulation Tasks in Augmented Reality Using Unconstrained Elicitation. *Proceedings of the ACM on Human-Computer Interaction* 4, ISS (Nov. 2020), 1–21. https://doi.org/10.1145/3427330

[60] Adam S. Williams and Francisco R. Ortega. 2022. The Impacts of Referent Display on Gesture and Speech Elicitation. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (Nov. 2022), 3885–3895. https://doi.org/10.1109/TVCG.2022.3203090 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[61] Yang Zhang, Wolf Kienzle, Yanjun Ma, Shiu S. Ng, Hrvoje Benko, and Chris Harrison. 2019. ActiTouch: Robust Touch Detection for On-Skin AR/VR Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1151–1159. https://doi.org/10.1145/3332165.3347869

[62] Fengyuan Zhu and Tovi Grossman. 2020. BISHARE: Exploring Bidirectional Interactions Between Smartphones and Head-Mounted Augmented Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376233

6.2 Multimodal Notification Interaction

We then turned our attention to multi-modal interaction [4].We repeated the previous experiment but with different combinations of modalities. The interaction techniques we used are split into two blocks: Indication and Confirmation. Bowman et al. [7] classify Indication and Confirmation as two steps in a 3D manipulation task. The last step in their classification is Feedback, which is provided in our system by visual and auditory confirmation of the selection. One modality is used to indicate the notification the user wishes to interact with, whereas the other confirms the interaction, in our case the pressing of a button. The two modalities are used *in complementarity* [83], meaning that they are processed individually but are merged for a single interaction, leading to faster interaction. For this study, we chose the following techniques:

- **Point and Pinch (PP):** Users point at a notification button with a ray extending from their wrist and activate it by pinching their index finger and thumb.
- **Point and Speech (PS):** Users point at the button with the same ray and say "Select" to activate it. The keyword was chosen for being short, easily recognized by the system, and easy to remember, aligning with prior multimodal interaction research.
- **Gaze and Speech (GS):** Users select the button by looking at it (causing it to highlight) and say "Select" to activate it. This approach avoids the Midas touch problem by requiring deliberate speech confirmation.
- **Gaze and Pinch (GP):** Users select a button by looking at it and activate it by pinching their thumb and index finger.
- **Touch:** Since in our previous work, Touch interaction was found to be ideal, we have included it for comparison to better understand the difference between uni- and multi-modal interaction.

[4]© 2025 IEEE. Reprinted, with permission, from Lucas Plabst, Florian Niebling, Sebastian Oberdörfer, and Francisco R. Ortega, "Order up! Multimodal Interaction Techniques for Notifications in Augmented Reality," in the IEEE Transactions on Visualization and Computer Graphics, pp. 1–10, 03-2025, DOI: 10.1109/TVCG.2025.3549186.

Order Up! Multimodal Interaction Techniques for Notifications in Augmented Reality

Lucas Plabst, Florian Niebling, Sebastian Oberdörfer, and Francisco Ortega

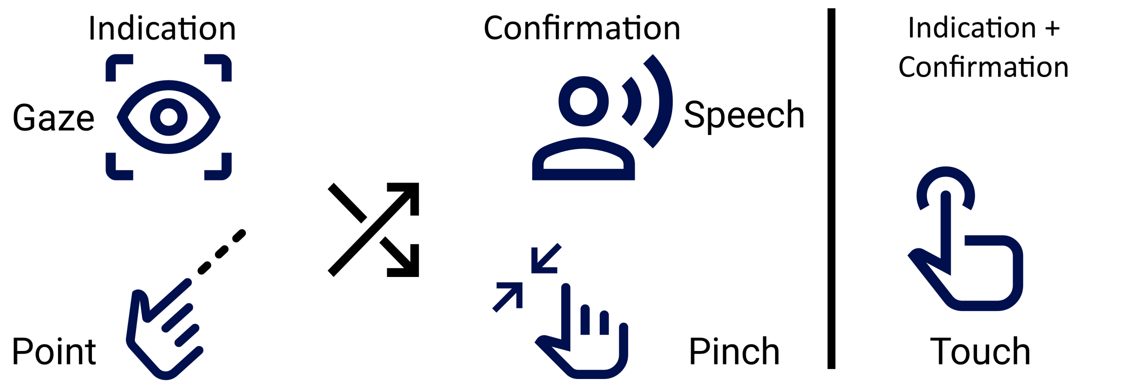


Fig. 1: Modalities used for indication and confirmation of notifications in the experiment. Modalities consist of Gaze, Hand-Pointing, Speech, and Touch. The resulting interactions are Gaze and Pinch, Gaze and Speech, Point and Pinch, Point and Speech and Touch.

Abstract—As augmented reality (AR) headsets become increasingly integrated into professional and social settings, a critical challenge emerges: how can users effectively manage and interact with the frequent notifications they receive? With adults receiving nearly 200 notifications daily on their smartphones, which serve as primary computing devices for many, translating this interaction to AR systems is paramount. Unlike traditional devices, AR systems augment the physical world, requiring interaction techniques that blend seamlessly with real-world behaviors. This study explores the complexities of multimodal interaction with notifications in AR. We investigated user preferences, usability, workload, and performance during a virtual cooking task, where participants managed customer orders while interacting with notifications. Various interaction techniques were tested: Point and Pinch, Gaze and Pinch, Point and Voice, Gaze and Voice, and Touch. Our findings reveal significant impacts on workload, performance, and usability based on the interaction method used. We identify key issues in multimodal interaction and offer guidance for optimizing these techniques in AR environments.

Index Terms—augmented reality, multimodal interaction, eye gaze, speech commands, notifications, gestures, 3D user interfaces

1 Introduction

Notifications are a unique aspect within user interfaces (UI), defined by their dynamic and temporary nature [24]. Notifications possess a unique state, capable of conveying critical information and presenting trivial updates. Their significance stems from their real-time connection to events, messages, or updates, rendering them inherently time-sensitive. Yet, their temporary nature sets them apart, as they can quickly lose relevance if the notification-producing event is no longer relevant. Whether serving as ignorable reminders or vital alerts, notifications try to grab the user's attention. Unless they are ignored until they expire, most notifications are interacted with [52], even if it is just a dismissal, as they do not provide relevant information at that time. Due to their invasive nature, the sudden appearance of a notification and the subsequent interaction with it could cause distraction or frustration. With a smartphone, ignoring a notification while the phone is not in use is as simple as leaving the phone in your pocket until you want to attend it [26]. But what if the display is no longer in your pocket, but is instead worn on your head? In a future where AR glasses could become our means of interacting with the digital and the physical world, finding a way to deal with these distractors is essential. AR can, for example, be a useful tool in surgery [18]. So imagine a future surgeon operating on a patient while wearing an AR head-mounted display (HMD). Patient alarms or other time-sensitive alerts could be directly displayed in the headset without the surgeon needing to turn away from the patient. However, these alerts cannot be detrimental to the current task, and it is crucial that the surgeon can interact with the notifications, such as acknowledging a patient alarm, to ensure the information has been processed and the situation addressed. Even simple urban navigation could be impacted by poorly designed pop-ups, since it has been shown that notifications are highly interruptive, even if ignored [65]. To systematically explore the complexities of AR notifications and their impact on user performance and experience, we first turned to a more general task: a simulated cooking environment. This setting allowed us to control variables and observe the fundamental interactions between users and AR notifications in a broadly applicable way, before extending insights to more specialized and high-stakes scenarios. Our primary contributions in this work are as follows:

- *Lucas Plabst is with Computer Science & NUILAB at Colorado State University*
  E-mail: lplabst@colostate.edu
- *Florian Niebling is with Advanced Media Institute at TH Köln*
  E-mail: florian.niebling@th-koeln.de
- *Sebastian Oberdörfer is with HCI Group at University of Würzburg*
  E-mail: sebastian.oberdoerfer@uni-wuerzburg.de
- *Francisco Ortega is with Computer Science & NUILAB at Colorado State University*
  E-mail: fortega@colostate.edu

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

- We provide an analysis of various interaction modalities for AR notifications, demonstrating how their combinations influence user perception and performance.
- Our findings show that compared to using each modality on its own, integrating multiple interaction modalities yields superior results.
- We identify the ongoing challenges and limitations faced by multimodal interactions in AR environments.
- Based on our findings, we offer suggestions for designing future AR systems that effectively leverage multimodal interactions.

2 Related Work

2.1 Notifications in Mixed Reality

Notifications have been studied in desktop environments [14, 15] and more recently in the smartphone field, given their ubiquitous use, with an estimated 90% of American adults owning and using smartphones daily [50]. Working adults receive an average of between 45 to 80 notifications per day, with some groups like college students receiving over 400 [1, 60, 39]. Research by Pielot et al. [52] revealed that disabling notifications for a day led to participants feeling less distracted and more productive, though they also expressed concern about missing important information and feeling disconnected from their social networks. With Mixed Reality (MR) anticipated to grow at a rate of 30% annually until 2032 [10], and more companies releasing MR-HMDs, a future where these headsets become our primary means of computing is plausible. AR-HMDs have a great potential to become pervasive in day-to-day life, as they do not isolate the user from their surroundings and can supplement the real world with helpful information. In this work, we refer to AR as seeing the real-world with virtual objects superimposed onto or composited with, following Azuma et al. [6]. We use MR as an umbrella term referring to a part of the Reality-Virtuality Continuum [63], containing both AR and VR. Janaka et al. [26] have demonstrated that messaging using an AR-HMD showed better multitasking capabilities and faster response times than using a smartphone, and suggest that even now the usage of an HMD as a supplemental device to the phone can improve messaging during multitasking. Research on notification modalities in VR by Ghosh et al. [19] and in AR by Lazaro et al. [35] found that combining visual and auditory notifications improved preference and performance. Participants expressed not only wanting to receive notifications but also to interact with them.

Another output modality that has been researched is the position of notifications in AR. Plabst et al. [53] looked at notification placement in AR and found a bottom-center position in the user's field of view to be optimal in general use cases, world placement in stationary tasks, and body placement in tasks with a lot of digital content. Similar conclusions were reached in a study about VR notification placement [59], as well as in an AR experiment about dual-task scenarios [13]. Imamov et al. [23] also researched glanceable information in simulated AR and found that a bottom-center position close to the primary task was more comfortable and faster than other placements. It is however worth noting that because of the inherent differences between AR and VR, it is unclear how the results of VR experiments apply to AR. Rzayev et al. [58] also showed the effect notification position in AR can have in a social interaction setting. They found that most participants preferred receiving notifications on an AR headset rather than a smartphone and that the position affected the person wearing the headset and the person not wearing the headset differently. Lee et al. [36] found that the position of notifications in AR affected response time, as well as inversely affecting walking speed, leading to the conclusion that the priority of the notification content needs to be considered when deciding where to place them. Another position for glanceable information in AR was researched by Satkowski et al. [62] who found that the ceiling or the floor of a physical space are well suited to display secondary non-critical information in AR without being obtrusive. Looking more at presentation as well as location, Lucero et al. [41] proposed a minimal notification interface in AR that did not distract participants from walking in public while still giving them updates. Notifications were presented as butterflies moving across the field of view, that could be opened and dismissed using swipe gestures on a finger-mounted touchpad.

2.2 Augmented Reality Interaction

In their study on general interactions in AR, Wang et al. [66] explored the fusion of various input modalities. They discovered that a combination of Gaze, Gesture, and Speech yielded the most efficient results, although users preferred the combination of Gesture and Speech. Conversely, unimodal interactions using solely Gaze or Gestures generally underperformed compared to the multimodal options. Lee et al. [38] similarly investigated multimodal versus unimodal inputs in AR, noting that while participants favored multimodal input over sole Speech or Gaze, they did not demonstrate objective performance improvements with it. For accessing generic information on an AR headset, Lu et al. [40] devised a system enabling users to summon information like weather updates or sports scores through either head movement or gaze. They observed that users preferred and performed best with gaze-based interactions.

Kosch et al. [31] evaluated three AR notification selection techniques while cycling: gaze with a dwell time, gestures, and gaze for indication with a physical button press for confirmation. They found that gaze with dwell time resulted in the lowest error rates, but required more attention and focus, whereas the gaze with button approach was preferred by users and had the lowest task completion time. Also using a dedicated input device for AR, Cai et al. [11] used a circular UI on an HMD that was controlled with a ring mouse worn on the hand, which they called *ParaGlassMenu*. In that experiment, Cai et al. compared this with more linear displays, voice assistants, and a smartphone in a conversational setting. They found that the *ParaGlassMenu* outperformed the other interactions in performance as well as subjective measures. Plabst et al. [54] investigated various display types for multiple notifications in AR, along with unimodal interaction techniques. Their findings indicated that Touch outperformed Voice or Gaze, and participants favored having a notification list attached to their hand rather than in the surrounding environment.

3 Experimental Setup: Kitchen Environment

In this work, we researched how we can interact with notifications in AR using multimodal interaction. Especially because notifications are usually short-lived, interaction needs to be quick, easy, and not distracting. For this reason, we set out to answer two research questions:

RQ 1: Is the user's perceived usability of notifications influenced by the interaction technique?

RQ 2: Is the primary task performance influenced by the interaction technique of notifications?

Usability has been defined by the International Organization for Standardization (ISO) as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [25]. We built upon the open-source research environment proposed by Raikwar et al. [57] to conduct our experiment. While in the cooking environment, participants engaged with various notifications using the Hololens 2, an optical see-through (OST) AR-HMD. The OST design of the headset not only grants users real-time visibility of their hands and bodies but also enhances depth perception, which distinguishes it from VR headsets [29], while also being more accurate for indication and confirmation tasks [33]. In comparison with video see-through AR, OST was found to have better text legibility [16] and better depth perception [2]. The HMD's eye-tracking functionality demonstrated precision within a 1.5° visual angle around the intended target, a finding by Kapp et al. [30]. Hand-interactions, eye-tracking, and voice recognition were implemented using the Mixed-Reality-Toolkit 2.8.3 [46]. Performance logs show that the environment runs at a consistent 60 frames per second, with infrequent drops to around 40.

3.1 Task

In this simulation environment, participants were tasked with fulfilling customer food orders, representing different real-life situations. These

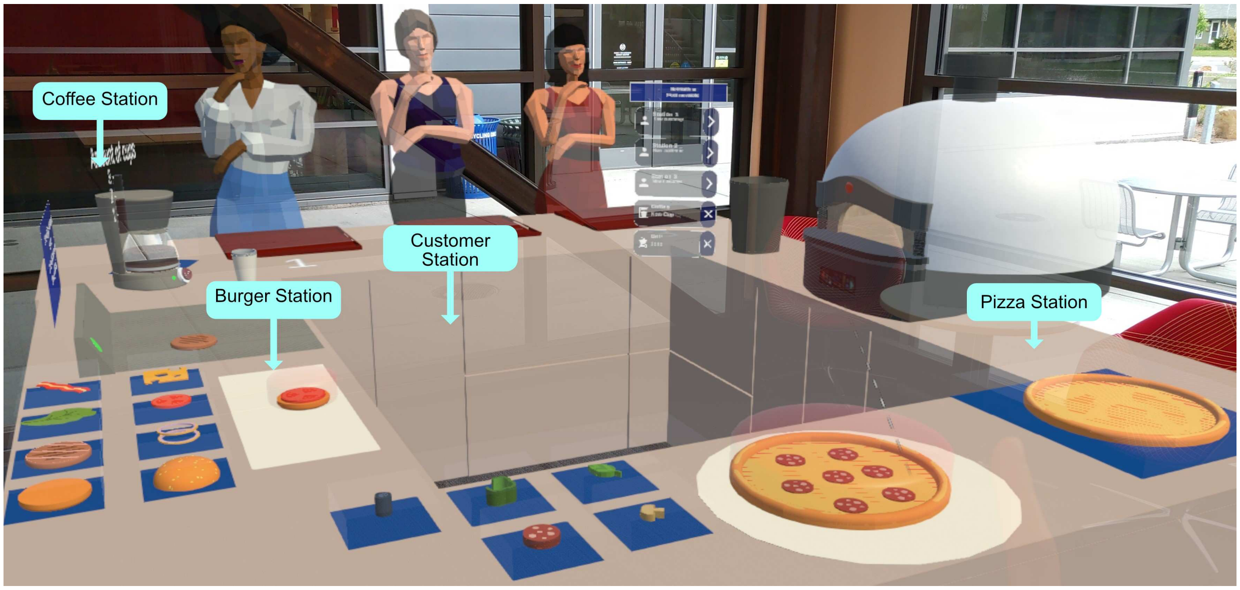


Fig. 2: Virtual cooking environment displayed in a large room in Augmented Reality

scenarios can range from high-stress to low-stress environments, allowing users to handle multiple tasks at once or focus on one thing at a time, moving around or staying in one spot, depending on how the kitchen was set up. Stress can be regulated by several factors, including but not limited to the number of customers, the amount of ingredients on meals, the waiting time for customers, or the time until food is burnt and needs to be recooked. For our work the kitchen task is used as a general use case that people could encounter in their daily life, trying to perform multiple things at once with some time sensitivity. The use of a virtual cooking environment in our study serves as a model for understanding interaction techniques with AR notifications. This environment mirrors several essential elements found in professional settings, allowing us to draw broader conclusions while also exploring specific nuances relevant to high-stakes scenarios. Both cooking and professional environments involve complex task management, where users juggle multiple tasks simultaneously, requiring attention to detail, and prioritization. The cooking environment also demands real-time decision-making, similar to professional settings. A cook must respond promptly to a timer, much like a mechanic reacts to machinery changes. This need for timely responses to notifications is common across different contexts. The choice of a kitchen reflects the need for a versatile, general-use scenario that is easy for participants to understand and engage with, while still being representative of the cognitive and physical challenges encountered in complex, time-sensitive work environments.

The kitchen (see Figure 2) was around 3 meters by 3 meters in size and had four stations. Participants could use both hands to handle food and ingredients and were instructed to grab the virtual objects just like they would with physical ones. At the **Customer Station**, up to three customers could order food. Each customer was represented by a simple human avatar with slight movements. A red tray indicated where the prepared food had to be served. When an order was accepted by the user by interacting with a notification, a two-minute timer appeared, showing how much time was left to finish. No direct interaction with the avatars occurred. The **Coffee Station** had a coffee machine with a pot. Users had to press a button to start the machine, and when enough for a cup was made, a notification alerted the user. They could then pour coffee into a cup, and a lid appeared when it was full, signaling it was ready. The **Pizza station** and **Burger station** had ingredients to assemble the meal and a cooking device. The patty or pizza dough had to be cooked for at least 10 seconds to be done, upon which a notification was sent. After 40 seconds, the food would burn and would not be accepted by the customers. Next to the Pizza station, participants could find a trash can to discard incorrectly prepared food items. A movable list was next to the trash can, where notifications would move after being in the field of view for 8 seconds.

3.2 Notifications

Participants received several notifications throughout their task. Every notification featured a title indicating its source, with the content of the notification underneath. Next to this text was an icon corresponding with the notification's title for quicker information access [27]. We limited notifications to task-related ones for privacy reasons, to control the notification amount, and to ensure participants paid attention to them, as they directly related to their task.

The notification text was set within the comfortable range for text size [44], making it easy to read and allowing all notification texts to fit inside the panel without cutting off the text or decreasing font size to fit. Text was displayed with a white font and a dark gray background for optimal readability in most situations [28], as the changing background in AR environments can lead to contrast issues depending on color choice. The moment a notification was delivered, a sound was played to alert the user of its presence, as this leads to better recognition and performance [35], and ensures that attention is drawn towards the notifications, so that missing notifications is unlikely. Based on previous findings [53, 37, 13], the notifications were positioned in the bottom center of the device's field of view. This position was found to be ideal as it does not distract the user from their task too much while still being noticeable, while also allowing the user to move around without losing sight of the notification. Notifications were placed at a distance of 75 - 90 cm, keeping them within the acceptable range recommended in the development guidelines for the headset [44]. We chose this distance because placing the notifications closer than 75cm could increase eye-fatigue due to the vergence-accommodation conflict [32], whereas placing them too far away would make touch interaction impossible, and could introduce text legibility issues.

Notifications had a blue button on the right side that was used for interaction. Interacting with the button caused an action, depending on whether it was an order or food notification. Looking at user experience design patterns [4, 3], they describe two types of notifications: actionable notifications, which the user has to react to, or informational notifications, whose purpose is to relay information to the user. The main difference between the two types is the expectation of interac-

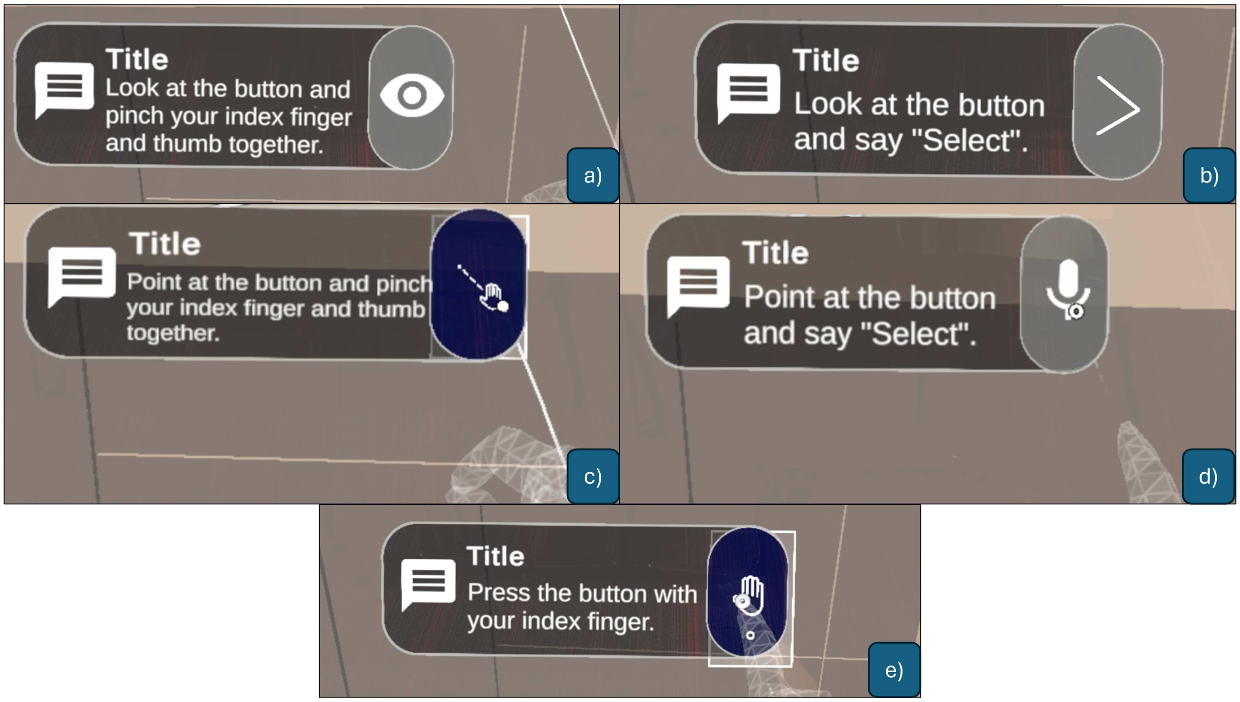


Fig. 3: Interaction techniques used in the experiment: a) Gaze and Pinch; b) Gaze and Speech; c) Point and Pinch; d) Point and Speech; e) Touch

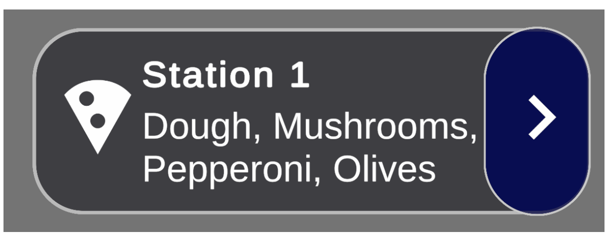


Fig. 4: A notification for a pizza order.

tion. Whereas informational notifications do not require a response, actionable notifications at least expect to be attended. For example, a notification informing a user about today's weather does not expect a response, whereas a notification about an incoming phone call or direct message does expect a response. Since users deal with notifications differently depending on their context [52], it was important to not only have a single type of notification in this experiment. The system we used deploys two types of notifications: order notifications and food notifications.

Order notifications: An order notification was sent for each new customer. This notification had to be interacted with to see what the order was, therefore making it an actionable notification. Upon accepting the order, the notification content would change to show the customer's order (see Figure 4). This would also start a visible two-minute countdown, showing the time remaining to prepare the food, after which the customer. When the meal was placed on the tray before the customer, the notification had to be interacted with again to confirm the order. If the ingredients matched, a "success" tone would play and the notification would display "correct order" before disappearing. If the food was incorrectly prepared, a "failure" tone would play and the notification would display "incorrect order". Activating the button again would cause the "incorrect order" text to disappear and the notification would display the ingredients for the food ordered again.

Food notifications: Every station produced notifications when the desired cooking state was reached. Activating the button dismissed the notifications, making them informational notifications since no direct action was required by the user as a result of the notification, and their purpose was strictly to convey information. Interaction with this notification was not necessary for completing the task, but it could be dismissed for better organization.

3.3 Interaction Techniques

Plabst et al. [54] researched notification interaction. However, their work is limited by the fact that they only looked at unimodal interaction techniques. In our work, we wanted to research multimodal interaction for notifications, as they might offer more flexibility and performance improvements, as well as being preferred by users over unimodal interaction techniques [61, 12]. In the work by Plabst et al. [54], touch interaction was unilaterally found to be the optimal technique, so we used it as a control condition to compare against multimodal interaction techniques (see Figure 3). They found several problems with Gaze and Voice, such as participants' frustration with the artificial slow-down in Gaze interaction due to the dwell-time, or issues with selecting the correct notification in Voice interaction. Our multimodal approach has the potential to address these issues by allowing users to combine complementary input methods, eliminating the need for dwell-time or labels for specific voice commands.

The interaction techniques we used are split into two blocks: Indication and Confirmation. Bowman et al. [9] classify Indication and Confirmation as two steps in a 3D manipulation task. The last step in their classification is Feedback, which is provided in our system by visual and auditory confirmation of the selection. One modality is used to indicate the notification the user wishes to interact with, whereas the other confirms the interaction, in our case the pressing of a button. The two modalities are used *complementarity* [42], meaning that they are processed individually but are merged for a single interaction, leading to faster interaction. We found that the use of hands, eyes, and voice are all extensively studied techniques for MR devices [49]. These modalities are also commonly found and used in state-of-the-art MR headsets like the Hololens 2, Meta Quest Pro, or Apple Vision Pro. For this study, we therefore chose the following techniques:

Point and Pinch (PP): When pointing at something in the environment, a ray appears extending from the wrist of the user's hand. Notifications could be interacted with by pointing at the button and then pinching the index finger and thumb together. Correctly pointing at the button caused it to be highlighted, confirming the target. Notifications using the pointing interaction were placed at a distance of 90 cm instead of 75 cm like the others, as the shorter distance would require the arm to be held in an uncomfortable position. This technique is used as the primary means of interaction by the Hololens 2 [45], as well as the Meta Horizon OS (Meta Quest) [43].

Point and Speech (PS): Utilizing the same ray, participants could point at the notification button and say the command "Select" out loud while still pointing. Through informal piloting and based on previous literature showing a preference for short voice commands in multimodal interaction [48], we chose this keyword, as it was short, easily recognizable by the Hololens voice recognition, and easy for the participants to remember. This interaction technique, introduced by Bolt et al [8], has since been used extensively in 3D environments.

Gaze and Speech (GS): Subjects could use the device's eye-tracking capabilities to select the button by looking at it, causing it to turn gray to reflect the state of being looked at. They could then say the "Select" keyword to activate the button. Using any confirming action with Gaze also alleviates the Midas touch problem, where systems cannot differentiate between basic eye functions like looking from deliberate interaction.

Gaze and Pinch (GP): Like GS, the button could be highlighted by looking at it and then activated by pinching the thumb and index finger together on one hand without needing to point using the hands. This interaction technique is also used as the primary interaction with headsets like the recently released Apple Vision Pro [5].

Touch: Each notification button was selected and subsequently activated by pressing it with the index finger of either hand. Pushing the button was animated to give the user the appropriate feedback since mid-air touches cannot provide haptic feedback.

As participants needed to grab objects in the environment using their hands, there was a potential mismatch between modalities in the main task and the interaction with the notifications. Our task is a simulation of a real-world task that requires the use of hands, and in the real world, the only interaction with digital content would be the notifications. Also, switching modalities between the UI and the main task is a common practice in 3D applications like video games, where the game's controls and the UI's controls often differ [55, 56]. For example, a VR video game might use the controllers as virtual hands to interact with the environment but use a ray for UI interaction.

4 Experiment

We used a within-subjects design with one independent variable: *interaction technique* (GS, GP, PP, PS, and Touch), yielding five total conditions. Latin-square counterbalancing was applied to set the order of conditions. Overall, 30 participants were recruited from a university campus; however, due to technical issues three of them did not complete the experiment. We analyzed data of the remaining 27 participants (17 male, 9 female, 1 gender-fluid) aged between 19 and 34 (M = 24.4, SD = 3.7). They were given a $30 US-dollar equivalent gift card as compensation in local currency. All participants self-reported normal or corrected to normal vision and English proficiency. Only two participants reported using a Hololens 2 (or any OST HMD) before. The experiment was carried out under the supervision of the Institutional Review Board (IRB) of Colorado State University.

4.1 Procedure

The experiment occurred in a spacious lecture room, approximately ∼10x6 meters in size. Blinds were lowered to minimize sun-blinding, and lights were turned on to ensure consistent lighting for all participants. The cooking environment was positioned in the room, with a marker for consistent placement among participants. To ensure consistency for comparison, we replicated the procedure used by Plabst et al. [54].

Upon arrival, participants filled out a demographics questionnaire and gave their informed consent after reading the consent form. They were then instructed on how to use and wear the Hololens 2. After calibration for their eyes, ensuring correct display placement, and eye-tracking setup, participants entered the virtual cooking environment. Here, they underwent a tutorial explaining system functions, followed by a training segment covering all possible notification interactions. This made sure that the systems worked with the participant, and that the participant understood the interactions. Subsequently, participants engaged in a task where they prepared various food items and served customers at their own pace. Once they felt ready, they informed the experimenter to start the experiment. Participants had to fulfill six orders per trial, each comprising a single food type with randomized ingredients for burgers and pizzas. They had to prepare each food type twice in a random sequence. Upon completing all orders or letting them expire, participants were prompted to fill out questionnaires on a tablet. This process repeated for all five conditions before participants could remove the headset and complete the post-questionnaire.

4.2 Measures

When an experiment trial was completed, the participants were asked to fill out questionnaires about their experience with the notifications. A Raw NASA-Task-Load-Index (TLX) questionnaire [22, 21] was used to understand the perceived task load. Besides this, we deployed the System Usability Scale (SUS) questionnaire [20] to understand the system's overall usability. When all trials finished, participants were asked to rank the interaction techniques according to their usage preference and explain their ranking. For performance measurements, the device logged and measured the following metrics during the experiment: For general task performance, we measured the **total time per trial** and the **time** needed **per customer**. To understand the accuracy of task completion, we measured the amount of **incorrect orders** and **expired orders**. **Time until an order notification was accepted** was measured to understand the multitasking performance.

5 Results

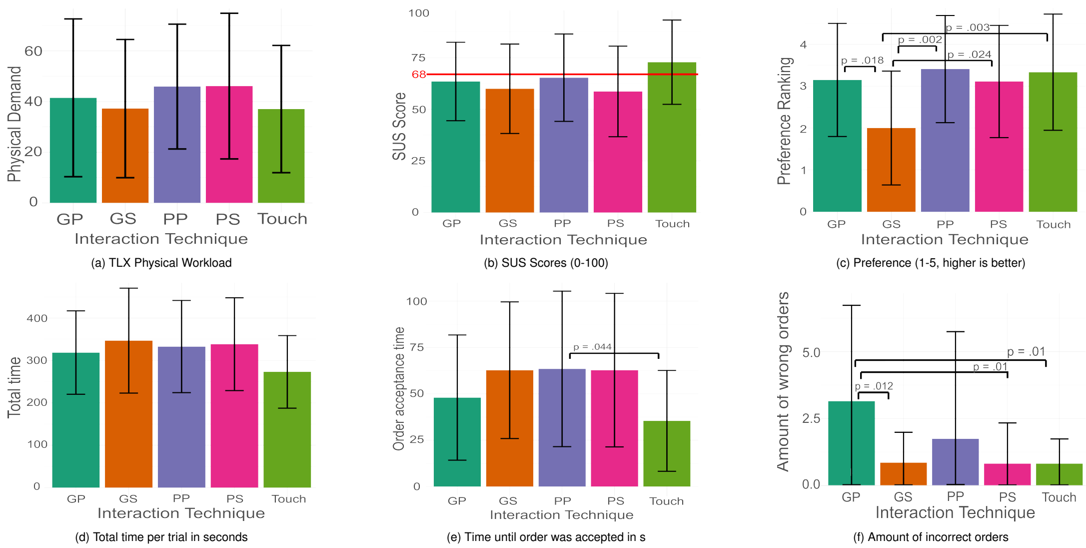
For all measures, we calculated a repeated-measures ANOVA for the independent variable *interaction technique* and conducted post-hoc analysis by running Tukey pairwise tests. Almost all measures met assumptions for ANOVA and were tested using Levenes tests (homogeneity of variance) and Shapiro-Wilk tests (normal distribution). Time until order was accepted was found to be not normally distributed, so we used a Kruskal-Wallis Test for main effect and Wilcoxon rank sum tests for pairwise comparisons. Overall, we found participants preferred using Gaze and Speech (GS) less than the other techniques, significant differences in physical workload as well as usability, significant differences in trial time, and that Touch performed better than Point and Speech (PP) in the time it took to start an order.

5.1 Subjective Measures

Task-Load-Index: We found no significant effect on the overall TLX score ($F(4,104) = 1.964$, $p = 0.122$). Looking at the individual subscales, we found a significant effect on the physical demand ($F(4,104) = 3.082$, $p = 0.019$). In the post-hoc analysis, we found no significant differences between individual groups. We found no significant differences for mental demand ($F(4,104) = 2.343$, $p = 0.06$), frustration ($F(4,104) = 2.364$, $p = 0.058$), effort ($F(4,104) = 1.361$, $p = 0.253$), and performance ($F(4,104) = 1.488$, $p = 0.211$).

System Usability Scale: We found a significant effect on the SUS score ($F(4,104) = 5.643$, $p < 0.001$). When conducting post-hoc analysis, we found no significant differences between individual groups. When analyzing the SUS, a value of 68 would place a score into the 50th percentile, with any value above being considered good usability. Looking at the mean values in Figure 2 shows us that only Touch falls into this range of good.

Preference Ranking: We assigned points for each rank using the Borda-Count method [17]. If a technique was ranked as the most



(a) TLX Physical Workload
(b) SUS Scores (0-100)
(c) Preference (1-5, higher is better)
(d) Total time per trial in seconds
(e) Time until order was accepted in s
(f) Amount of incorrect orders

Fig. 5: Performance and subjective measures, bars show standard error

Table 1: TLX Scores with subscales (0-100), SUS Scores, and Preference Rankings (values are Mean, *SD*, `Median`).

Condition	TLX Score			Phys. Demand			Frustration			Ment. Demand			SUS Score			Pref. Rank		
	Mean	*SD*	`Med.`	Mean	*SD*	`Med.`	Mean	*SD*	`Med.`	Mean	*SD*	`Med.`	Mean	*SD*	`Med.`	Mean	*SD*	`Med.`
GP	43.09	*23.93*	`45.0`	41.48	*31.16*	`40`	38.52	*28.21*	`35`	42.59	*28.02*	`40.0`	63.4	*19.0*	`67.5`	3.15	*1.35*	`3.0`
GS	41.67	*23.01*	`45.0`	37.22	*27.29*	`30`	32.96	*25.47*	`30`	41.30	*27.82*	`40.0`	59.9	*21.7*	`65.0`	2.00	*1.36*	`2.0`
PP	41.98	*21.70*	`45.0`	45.93	*24.65*	`55`	32.22	*29.59*	`25`	44.44	*24.94*	`45.0`	65.3	*21.2*	`70.0`	3.41	*1.28*	`3.0`
PS	47.13	*24.76*	`50.0`	46.11	*28.77*	`60`	45.19	*31.55*	`40`	48.33	*27.87*	`50.0`	58.6	*22.0*	`55.0`	3.11	*1.34*	`3.0`
Touch	37.65	*20.25*	`41.6`	37.04	*25.13*	`35`	29.44	*25.73*	`20`	37.22	*24.03*	`35.0`	72.8	*20.4*	`72.5`	3.33	*1.39*	`3.0`

preferable, it received 5 points, whereas the least preferable ranking was 1 point. We found a significant effect of *interaction technique* on the preference ranking score ($F(4,104) = 3.924$, $p = 0.005$). When conducting post-hoc analysis for pairwise comparisons, we found significant differences between GS (M = 2.00, SD = 1.36) and Touch (M = 3.33, SD = 1.39) p = 0.003, GP (M = 3.15, SD = 1.35) p = 0.018, PP (M = 3.41, SD = 1.28) p = 0.002, and PS (M = 3.11, SD = 1.34) p = 0.024.

5.2 Performance

Trial Time: We found a significant effect on the time per trial ($F(4,104) = 3.617$, $p = 0.008$). In the post-hoc analysis, we found no significant differences between individual groups.

Time until order accepted: This measurement was started when the notification was initially sent and stopped when the participant acknowledged the order. We found a significant effect on the time until the order was accepted (Chi square = 14.457, p = 0.006, df = 4). In the post-hoc analysis, we found significant differences between Touch (M = 35.40, SD = 27.20) and PP (M = 63.46, SD = 41.96) p = 0.036 and GS (M = 62.7, SD = 36.9) p = 0.011.

Time per customer: The time per customer began to be measured when the participant accepted the order. We found no significant effect on the time spent per customer ($F(4,104) = 0.602$, $p = 0.662$).

Incorrect Orders: We found a significant effect on the wrong order number ($F(2.28,59.18) = 5.575$, $p = 0.004$). In the post-hoc analysis, we found significant differences between GP (M = 3.15, SD = 3.58) and Touch (M = 0.81, SD = 0.92) p = 0.01, GS (M = 0.85, SD = 1.13) p = 0.012, and PS (M = 0.81, SD = 1.52) p = 0.01.

Expired Orders: We found no significant effect on the number of expired orders ($F(4,104) = 0.673$, $p = 0.612$).

5.3 Preference Responses

Participants were asked to explain their preference ranking. These interview responses were then collected into an Affinity-Diagram [7], where key points were identified from the responses.

Generally, participants had problems with the reliability of some modalities. Starting with eye-tracking, participants found it to be sometimes unreliable and expressed that it took the system too long to recognize that they were looking at the target. P1 said "I did not find the eye tracking as easy to use because often the headset didn't recognize that I was looking at the button," with P13 expressing that "eye-tracking is also hard as it needed time and was hard on my eyes to wait until the system recognizes me."

Participants uttered the same issues about the voice recognition, as they felt that the system did not understand them consistently, with P16 saying, "Sometimes the speech recognition did not recognize my command well, and it made the cooking experiment a bit frustrating." In addition to the unreliable detection, participants were sometimes frustrated by needing to keep saying the keyword repeatedly, which was amplified by needing to repeat a word when it was not recognized. P1 said that they "also didn't like the voice commands that much, because it seemed kind of annoying to have to repeat the word " with P13 adding that "Voice command made me self aware." Most of the negative comments surrounding voice have to do directly with the system not understanding them well enough. However, this did not affect all users, as PS did not score lower on the preference rankings, with P7 stating that they "felt like they had the most control" when using PS. This highlights that detection accuracy is a key issue holding back speech interaction.

Lastly, reliability was also a common theme in the pinching (PP and

Table 2: Performance Results (values are Mean, *SD*, `Median`).

Condition	Trial time in seconds			Time until order was accepted (s)			Time per customer (s)			Expired orders			Incorrect orders		
	Mean	*SD*	`Med.`	Mean	*SD*	`Med.`	Mean	*SD*	`Med.`	Mean	*SD*	`Med.`	Mean	*SD*	`Med.`
GP	318.5	*98.3*	`303.93`	48.0	*33.8*	`40.46`	66.1	*19.2*	`61.58`	1.63	*1.78*	`1`	3.15	*3.58*	`2`
GS	346.7	*123.8*	`327.51`	62.7	*36.9*	`55.20`	68.4	*16.0*	`64.72`	1.25	*1.51*	`1`	0.85	*1.13*	`0`
PP	332.7	*108.7*	`313.34`	63.5	*42.0*	`52.19`	63.1	*13.7*	`60.56`	1.15	*1.26*	`1`	1.74	*4.01*	`1`
PS	338.3	*109.4*	`296.57`	62.8	*41.4*	`50.80`	69.5	*23.1*	`62.89`	1.26	*1.81*	`0`	0.81	*1.52*	`0`
Touch	273.0	*85.6*	`254.17`	35.4	*27.2*	`30.83`	66.5	*17.1*	`63.21`	1.11	*1.57*	`0`	0.81	*0.92*	`1`

GP) conditions, as participants had issues getting the headset to detect the pinching gesture reliably. P27 attributed this to the headset's small detection field, saying, "The pinch felt inconsistent because I really had to make sure my hand was in front of me," and P5 said that it was "always hard to register the pinch." The registration of the gesture seems to be a trade-off between accuracy and false positives. In their study evaluating Gaze and Pinch interaction, Pfeuffer et al. [51] observed many issues with the system falsely registering small hand movements as a pinch and suggest a more explicit pinch gesture like the one the Hololens uses. However as we saw in our experiment, if the gesture is too explicit, the user frustration is still there, as instead of being too easy to perform, the gesture is now too hard to perform.

We can now also identify an inherent problem of the multimodal interaction techniques used in this experiment: the latency between indication and confirmation. Since both recognition systems take time to register, the interaction relies on both systems registering simultaneously so as not to cause a mismatch. If there are underlying delays or reliability issues with both indication and confirmation, the resulting interaction can cause frustration. This highlights that even if the interaction itself is well-liked and appropriate, a perceived lack of reliability can dramatically decrease the user's experience with the system.

Point and Pinch (PP): Participants felt that PP was effective overall but that it was sometimes hard for them to get the action correct, "as there are some chances of pointing something else while pinching" (P9).

Point and Speech (PS): Participants mentioned that while they found PS convenient to use, they also thought it was uncomfortable and "annoying, with the disadvantage of having to talk and physically point" (P10).

Gaze and Pinch (GP): Using GP, participants stated that they felt there was a disconnect between the indication and confirmation, as "it felt very unnatural to look at the option and select it with my hand" (P15) and "I needed to concentrate on both while managing both, it felt a little bit difficult" (P23).

Gaze and Speech (GS): Comments about GS mostly focused on the difficulty of maintaining one's gaze on the target while saying the keyword. P9 stated: "Eye tracking and voice command [are] very difficult for me because while talking my eye pointing changes and it makes me concentrate more on the selection", with P26 echoing the complaint: "I found it difficult to look in the specific spot for the button while telling it to select."

Touch: Over half of the participants directly expressed liking Touch, as they felt it was the most intuitive and easy interaction method. P14 said that "We touch things in real life, so it is the most interactive and realistic." However, some participants had problems pressing the button when it popped up initially: "Touch didn't work well for me. Sometimes it seemed that buttons were too far" (P2), or when they wanted to interact with the notifications and the list was not within reach: "Touch is easiest to use but when there is a list it is difficult to reach it" (P13).

6 Discussion

Comparing our findings to the results of Plabst et al. [54], we can see that multimodal interaction techniques mostly close the gap to the unimodal "winner," Touch, when compared to unimodal interaction techniques. In their study [54], touch was rated significantly higher in preference than the unimodal techniques, while also performing better and being rated higher in usability. In our study, however, Touch was not rated significantly higher than the multimodal interactions in preference, nor did it have higher ratings or performance benefits besides outperforming PP and GS in order of acceptance times. A challenge we and our participants encountered with touch interaction is the balance required in positioning user interface elements. If placed too close, users may experience discomfort, partly due to the vergence-accommodation conflict, which remains a concern in MR devices [32]. Given that the Hololens' focal distance is approximately two meters [44], only distant content is unaffected by this issue. While positioning elements farther away could enhance comfort and mitigate the conflict's impact, it would render direct touch interaction impossible. Mentioning input legacy bias is also still important [34, 67]. Touch interaction is a natural part of human behavior, spanning both physical and digital environments. From how we communicate and connect with others through physical touch to how we interact with touchscreen devices, it's a fundamental aspect of our daily lives. Whether shaking hands or swiping on a smartphone, touch plays a central role in our experiences, making this interaction more intuitive and efficient. While the task in the experiment was performed using hands, most general daily activities also require the use of hands, making this a well-suited metaphor for daily activities. The other techniques we investigated in this experiment are novel and not commonly found or used in other computing tasks. This is highlighted by P1: "My favorite options were just the standard touch or point and pinch, probably because that is more like what I'm used to doing in real life and in my past experiences with VR/AR systems." Exposing subjects to these multi-modal interaction techniques for a longer time might enhance the user's abilities with them.

To answer RQ1 ("Is the user's perceived usability of notifications influenced by the interaction technique?"), our findings show that GS was rated as being significantly less preferable than the others. We also found effects on SUS and physical workload overall but were not able to find any differences between individual groups. When explaining their rankings, participants often didn't directly refer to the multimodal technique but rather to the individual modalities. More often than not, they directly mentioned issues with the overall modalities, causing them to rank techniques containing that modality as less preferable. Participants frequently encountered challenges in coordinating two modalities. They reported looking or pointing away before completing actions like pinching or speaking, resulting in misalignment. This underscores a challenge in multimodal systems, where latency or recognition errors can significantly impact interaction quality, leading to frustration and errors. Despite utilizing state-of-the-art eye tracking in the Hololens 2 [30] in well-controlled environments, reliability issues persisted, indicating potential challenges in real-world scenarios with less-than-ideal conditions. Prioritizing the reliability of recognition is essential in designing effective interactions. Implementing mechanisms like predicting the selected target based on dwell or pointing time before executing a manipulation command, or smoothing out pointing movements, may help alleviate these challenges, but further research is needed into optimizing complimentary modalities.

Lastly, while both hands could be used in the experiment, the task could be completed entirely with one hand only. Entirely hands-free interaction like GS could be more appropriate in situations where both hands are busy constantly and might also influence the preference rating of this specific interaction, as participants always had a hand-free to interact with. However, tasks requiring both hands could also demand the user's full attention, locking their gaze onto the task and necessitating a break in focus to interact, making the benefits of hands-free interaction less clear. Since the preference rating did not differ greatly besides this significantly lower preference for GS, allowing the user to choose their preferred interaction is something we recommend.

Answering RQ2 ("Is the primary task performance influenced by the interaction technique of notifications?"), we can see that the number of incorrect orders, representing the accuracy of the task completion, was significantly higher with GP than with GS, PS, and Touch. The higher number can either be attributed to the wrong execution of the food order or a misinput if there was no food on the tray, where the user wanted to choose a different notification or none at all but still confirmed the completion of an order. This means that in more general multitasking scenarios, the GP method is either a more unreliable interaction method, or it distracts the user enough to cause a negative performance impact on whatever they are doing primarily. We also found that the time until an order was accepted was lower with Touch than with PP and to an extent PS and GS, telling us that users using Touch were quicker to accept orders, either because they were multitasking more than in the other conditions, or because the general interaction was quicker. Participants using touch interaction were also 46 seconds faster than the 2nd fastest (GP, 318.5s) and 73 seconds than the slowest (GS, 346.7s) in overall trial time (Mean 320s), which supports this reasoning.

6.1 Design Implications

Based on our findings we want to give some design considerations for multimodal notification interactions in AR:

- Prioritize touch interaction for tasks requiring speed and efficiency, while allowing user customization of distance and size of UI elements.
- Give the user options to choose their preferred interaction technique while giving them a quick way to switch depending on context.
- Focus on reliability of individual modalities as well as the fusion of multiple to minimize errors.
- Provide safeguards to ensure successful action completion, such as briefly maintaining the indication of a target.

6.2 Limitations & Future Work

As participants expressed issues with the reliability of some of the interactions, a limitation of this experiment is, consequently, the systems we used. Even though we and probably many other developers used implementations provided by the manufacturer that were optimized for the device, results may still be impacted by the limitations of current AR systems. For instance, the HoloLens 2's limited camera angle may have impacted hand-tracking accuracy and required exaggerated hand movements. Although these device-specific constraints may have influenced the perceived reliability of some techniques vs. others, the TLX-score and the frustration sub-scale did not indicate differences between techniques. The observed trade-offs in multimodal interaction still show important usability considerations, which we have highlighted in our implications. For a satisfactory experience with multimodal interaction, latency must be minimized and perceived reliability prioritized. Future research on devices with improved capabilities could help isolate the impact of hardware limitations and validate the generalizability of these results.

Another limitation of our task was that notifications had only one action that the user could trigger. In operating systems like Windows or Android, which run on over 70% of all electronic devices connected to the internet [64], notifications have at minimum two interactions depending on program and version: One to dismiss the notification and one to open the corresponding program. However, our notifications only had one action, the dismissal or in case of the order-notifications (which could not be dismissed), the completion of the order. Future work should therefore look at notification interaction with more complex actions like quick-replies or image previews, which would increase the number of interaction targets on each notification. Having more targets could lead to differences in interaction technique performance.

Another possible future direction is to conduct an experiment where uni- and multimodal notification interaction techniques can be used at the same time, to get a better understanding of the context in which each interaction technique is used, since users might not want to always interact multimodally all the time [47]. Another future direction is to look at notification problems in-the-wild.

7 Conclusion

We explored different interaction techniques (four multi-modal: Gaze and Speech, Gaze and Pinch, Point and Pinch, and Point and Speech; one uni-modal: Touch) to interact with notifications in a virtual cooking environment. Our findings show that participants did not like using Gaze and Speech, but had no clear preferences between the other interactions. We also found differences in usability and physical demand, as well as worse task accuracy with the Gaze and Pinch method compared to the other techniques. Based on our findings, we recommend supporting multiple multimodal interaction techniques or allowing the user to choose their preferred interaction if the situation allows it.

Acknowledgments

This work was partially supported by the Office of Naval Research N00014-24-1-2214, Office of Naval Research N00014-21-1-2580, National Science Foundation 2327569, and National Science Foundation 2238313.

References

[1] U. Acer, A. Mashhadi, C. Forlivesi, and F. Kawsar. *Energy Efficient Scheduling for Mobile Push Notifications*, vol. 2. Aug. 2015. Journal Abbreviation: EAI Endorsed Transactions on Energy Web Publication Title: EAI Endorsed Transactions on Energy Web. doi: 10.4108/eai.22-7-2015.2260067

[2] H. Adams, J. Stefanucci, S. Creem-Regehr, and B. Bodenheimer. Depth Perception in Augmented Reality: The Effects of Display, Shadow, and Position. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 792–801, Mar. 2022. ISSN: 2642-5254. doi: 10.1109/VR51125.2022.00101

[3] alison@uxmag.com. Designing Notifications for Apps, Mar. 2020.

[4] Apple. Declaring your actionable notification types, 2023.

[5] Apple. Learn basic gestures and controls on Apple Vision Pro, 2024.

[6] R. T. Azuma. A Survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, pp. 355–385, 1997.

[7] H. Beyer and K. Holtzblatt. Contextual design. *Interactions*, 6(1):32–42, Jan. 1999. doi: 10.1145/291224.291229

[8] R. A. Bolt. "Put-that-there": Voice and gesture at the graphics interface. *SIGGRAPH Comput. Graph.*, 14(3):262–270, July 1980. doi: 10.1145/965105.807503

[9] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., USA, June 2004.

[10] F. Business Insights. Virtual Reality [VR] Market Size, Growth, Share by 2032, 2024.

[11] R. Cai, N. N. P. K. Janaka, S. Zhao, and M. Sun. ParaGlassMenu: Towards Social-Friendly Subtle Interactions in Conversations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pp. 1–21. Association for Computing Machinery, New York, NY, USA, Apr. 2023. doi: 10.1145/3544548.3581065

[12] P. Chojecki, D. Strazdas, D. Przewozny, N. Gard, D. Runde, N. Hoerner, A. Al-Hamadi, P. Eisert, and S. Bosse. Assessing the Value of Multimodal Interfaces: A Study on Human–Machine Interaction in Weld Inspection Workstations. *Sensors*, 23(11):5043, May 2023. doi: 10.3390/s23115043

[13] S. H. Chua, S. T. Perrault, D. J. C. Matthies, and S. Zhao. Positioning Glass: Investigating Display Positions of Monocular Optical See-Through Head-Mounted Display. In *Proceedings of the Fourth International Symposium on Chinese CHI - ChineseCHI2016*, pp. 1–6. ACM Press, San Jose, USA, 2016. doi: 10.1145/2948708.2948713

[14] E. Cutrell, M. Czerwinski, and E. Horvitz. Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. 2001.

[15] M. Czerwinski, E. Cutrell, and E. Horvitz. Instant Messaging and Interruption: Influence of Task Type on Performance.

[16] S. Debernardis, M. Fiorentino, M. Gattullo, G. Monno, and A. Uva. Text Readability in Head-Worn Displays: Color and Style Optimization in Video vs. Optical See-Through Devices. *IEEE transactions on visualization and computer graphics*, 20:125–39, Jan. 2014. doi: 10.1109/TVCG.2013.86

[17] P. Emerson. The original Borda count and partial voting. *Social Choice and Welfare*, 40(2):353–358, Feb. 2013. doi: 10.1007/s00355-011-0603-9

[18] B. Fida, F. Cutolo, G. di Franco, M. Ferrari, and V. Ferrari. Augmented reality in open surgery. *Updates in Surgery*, 70(3):389–400, Sept. 2018. doi: 10.1007/s13304-018-0567-8

[19] S. Ghosh, L. Winston, N. Panchal, P. Kimura-Thollander, J. Hotnog, D. Cheong, G. Reyes, and G. D. Abowd. NotifiVR: Exploring Interruptions and Notifications in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1447–1456, Apr. 2018. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2018.2793698

[20] R. A. Grier, A. Bangor, P. Kortum, and S. C. Peres. The System Usability Scale: Beyond Standard Usability Testing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1):187–191, Sept. 2013. Publisher: SAGE Publications Inc. doi: 10.1177/1541931213571042

[21] S. 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, Oct. 2006. Publisher: SAGE Publications Inc. doi: 10.1177/154193120605000909

[22] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock and N. Meshkati, eds., *Human Mental Workload*, vol. 52 of *Advances in Psychology*, pp. 139–183. North-Holland, 1988. ISSN: 0166-4115. doi: 10.1016/S0166-4115(08)62386-9

[23] S. Imamov, D. Monzel, and W. 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)*, pp. 851–858, Mar. 2020. ISSN: 2642-5254. doi: 10.1109/VR46266.2020.00110

[24] S. T. Iqbal and B. P. Bailey. Oasis: A framework for linking notification delivery to the perceptual structure of goal-directed tasks. *ACM Transactions on Computer-Human Interaction*, 17(4):1–28, Dec. 2010. doi: 10.1145/1879831.1879833

[25] D. Iso. 9241–210: 2010: ergonomics of human-system interaction—part 210: human-centred design for interactive systems (formerly known as 13407). *Switzerland: International Standards Organization*, 2010.

[26] N. Janaka, J. Gao, L. Zhu, S. Zhao, L. Lyu, P. Xu, M. Nabokow, S. Wang, and Y. Ong. GlassMessaging: Towards Ubiquitous Messaging Using OHMDs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(3):1–32, Sept. 2023. doi: 10.1145/3610931

[27] N. Janaka, S. Zhao, and S. Sapkota. Can Icons Outperform Text? Understanding the Role of Pictograms in OHMD Notifications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–23. ACM, Hamburg Germany, Apr. 2023. doi: 10.1145/3544548.3580891

[28] J. Jankowski, K. Samp, I. Irzynska, M. Jozwowicz, and S. Decker. Integrating Text with Video and 3D Graphics: The Effects of Text Drawing Styles on Text Readability. p. 10, 2010.

[29] J. A. Jones, J. E. Swan, G. Singh, E. Kolstad, and S. R. Ellis. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, vol. 7, pp. 9–14. ACM, Los Angeles California, Aug. 2008. doi: 10.1145/1394281.1394283

[30] S. Kapp, M. Barz, S. Mukhametov, D. Sonntag, and J. Kuhn. ARETT: Augmented Reality Eye Tracking Toolkit for Head Mounted Displays. *Sensors*, 21(6):2234, Mar. 2021. doi: 10.3390/s21062234

[31] T. Kosch, A. Matviienko, F. Müller, J. Bersch, C. Katins, D. Schön, and M. Mühlhäuser. NotiBike: Assessing Target Selection Techniques for Cyclist Notifications in Augmented Reality. *Proceedings of the ACM on Human-Computer Interaction*, 6(MHCI):1–24, Sept. 2022. doi: 10.1145/3546732

[32] G. Kramida. Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(7):1912–1931, July 2016. doi: 10.1109/TVCG.2015.2473855

[33] M. Krichenbauer, G. Yamamoto, T. Taketom, C. Sandor, and H. Kato. Augmented Reality versus Virtual Reality for 3D Object Manipulation. *IEEE Transactions on Visualization and Computer Graphics*, 24(2):1038–1048, Feb. 2018. doi: 10.1109/TVCG.2017.2658570

[34] A. Köpsel and N. Bubalo. Benefiting from legacy bias. *Interactions*, 22(5):44–47, Aug. 2015. doi: 10.1145/2803169

[35] M. J. Lazaro, S. Kim, J. Lee, J. Chun, and M.-H. Yun. Interaction Modalities for Notification Signals in Augmented Reality. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, pp. 470–477. ACM, Montréal QC Canada, Oct. 2021. doi: 10.1145/3462244.3479898

[36] H. Lee and W. 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)*, pp. 519–529. IEEE, Shanghai, China, Mar. 2023. doi: 10.1109/VR55154.2023.00067

[37] H. Lee and W. 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)*, pp. 519–529, Mar. 2023. ISSN: 2642-5254. doi: 10.1109/VR55154.2023.00067

[38] M. Lee, M. Billinghurst, W. Baek, R. Green, and W. Woo. A usability study of multimodal input in an augmented reality environment. *Virtual Reality*, 17(4):293–305, Nov. 2013. doi: 10.1007/s10055-013-0230-0

[39] U. Lee, J. Lee, M. Ko, C. Lee, Y. Kim, S. Yang, K. Yatani, G. Gweon, K.-M. Chung, and J. Song. Hooked on smartphones: an exploratory study on smartphone overuse among college students. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2327–2336. ACM, Toronto Ontario Canada, Apr. 2014. doi: 10.1145/2556288.2557366

[40] F. Lu, S. Davari, L. Lisle, Y. Li, and D. A. Bowman. Glanceable AR: Evaluating Information Access Methods for Head-Worn Augmented Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 930–939. IEEE, Atlanta, GA, USA, Mar. 2020. doi: 10.1109/VR46266.2020.00113

[41] A. Lucero and A. 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*, pp. 1–10. ACM, Funchal Portugal, Nov. 2014. doi: 10.1145/2663806.2663824

[42] J.-C. Martin. TYCOON: Theoretical Framework and Software Tools for Multimodal Interfaces. 1997.

[43] Meta. Getting started with Hand Tracking on Meta Quest headsets, 2024.

[44] Microsoft. Comfort - Mixed Reality | Microsoft Learn, Oct. 2021.

[45] Microsoft. Getting around HoloLens 2, Aug. 2021. GettingHoloLens2021.

[46] Microsoft. Releases · microsoft/MixedRealityToolkit-Unity, 2022.

[47] S. Oviatt. Ten myths of multimodal interaction. *Communications of the ACM*, 42(11):74–81, Nov. 1999. Publisher: Association for Computing Machinery (ACM). doi: 10.1145/319382.319398

[48] S. Oviatt, A. DeAngeli, and K. Kuhn. Integration and synchronization of input modes during multimodal human-computer interaction. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. ACM, Atlanta Georgia USA, Mar. 1997. doi: 10.1145/258549.258821

[49] T. Papadopoulos, K. Evangelidis, T. H. Kaskalis, G. Evangelidis, and S. Sylaiou. Interactions in Augmented and Mixed Reality: An Overview. *Applied Sciences*, 11(18):8752, Sept. 2021. doi: 10.3390/app11188752

[50] R. Pew Research Center. Americans' Use of Mobile Technology and Home Broadband, Jan. 2024.

[51] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen. Gaze + pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pp. 99–108. ACM, Brighton United Kingdom, Oct. 2017. doi: 10.1145/3131277.3132180

[52] M. Pielot, A. Vradi, and S. Park. Dismissed!: a detailed exploration of how mobile phone users handle push notifications. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 1–11. ACM, Barcelona Spain, Sept. 2018. doi: 10.1145/3229434.3229445

[53] L. Plabst, S. Oberdörfer, F. R. Ortega, and F. 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*, pp. 1–11. ACM, Online CA USA, Dec. 2022. doi: 10.1145/3565970.3567701

[54] L. Plabst, A. Raikwar, S. Oberdörfer, F. R. Ortega, and F. Niebling. Exploring Unimodal Notification Interaction and Display Methods in Augmented Reality. In *29th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11. ACM, Christchurch New Zealand, Oct. 2023. doi: 10.

1145/3611659.3615683

[55] C. Power, P. Cairns, and T. DeHaven. Mapping Virtual Reality Controls to Inform Design of Accessible User Experiences. In J. Abdelnour Nocera, M. Kristín Lárusdóttir, H. Petrie, A. Piccinno, and M. Winckler, eds., *Human-Computer Interaction – INTERACT 2023*, pp. 89–100. Springer Nature Switzerland, Cham, 2023. doi: 10.1007/978-3-031-42280-5_5

[56] K. Raaen and H. Sørum. Survey of interactions in popular vr experiences. In *Norsk Informatikkonferanse*, 2019.

[57] A. Raikwar, L. Plabst, and F. 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)*, pp. 909–910, Oct. 2022. ISSN: 2771-1110. doi: 10.1109/ISMAR-Adjunct57072.2022.00200

[58] R. Rzayev, S. Korbely, M. Maul, A. Schark, V. Schwind, and N. Henze. Effects of Position and Alignment of Notifications on AR Glasses during Social Interaction. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, pp. 1–11. ACM, Tallinn Estonia, Oct. 2020. doi: 10.1145/3419249.3420095

[59] R. Rzayev, S. Mayer, C. Krauter, and N. Henze. Notification in VR: The Effect of Notification Placement, Task and Environment. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pp. 199–211. ACM, Barcelona Spain, Oct. 2019. doi: 10.1145/3311350.3347190

[60] A. Sahami Shirazi, N. Henze, T. Dingler, M. Pielot, D. Weber, and A. Schmidt. Large-scale assessment of mobile notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3055–3064. ACM, Toronto Ontario Canada, Apr. 2014. doi: 10.1145/2556288.2557189

[61] A. Saktheeswaran, A. Srinivasan, and J. Stasko. Touch? Speech? or Touch and Speech? Investigating Multimodal Interaction for Visual Network Exploration and Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 26(6):2168–2179, June 2020. doi: 10.1109/TVCG.2020.2970512

[62] M. Satkowski, R. Rzayev, E. Goebel, and R. Dachselt. ABOVE & BELOW: Investigating Ceiling and Floor for Augmented Reality Content Placement. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 518–527, Oct. 2022. ISSN: 1554-7868. doi: 10.1109/ISMAR55827.2022.00068

[63] R. Skarbez, M. Smith, and M. C. Whitton. Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Frontiers in Virtual Reality*, 2:647997, Mar. 2021. doi: 10.3389/frvir.2021.647997

[64] S. G. Stats. Operating System Market Share Worldwide, 2024.

[65] C. Stothart, A. Mitchum, and C. Yehnert. The attentional cost of receiving a cell phone notification. *Journal of Experimental Psychology: Human Perception and Performance*, 41(4):893–897, Aug. 2015. doi: 10.1037/xhp0000100

[66] Z. Wang, H. Wang, H. Yu, and F. Lu. Interaction With Gaze, Gesture, and Speech in a Flexibly Configurable Augmented Reality System. *IEEE Transactions on Human-Machine Systems*, 51(5):524–534, Oct. 2021. Conference Name: IEEE Transactions on Human-Machine Systems. doi: 10.1109/THMS.2021.3097973

[67] A. S. Williams, J. Garcia, F. De Zayas, F. Hernandez, J. Sharp, and F. R. Ortega. The Cost of Production in Elicitation Studies and the Legacy Bias-Consensus Trade off. *Multimodal Technologies and Interaction*, 4(4):88, Dec. 2020. doi: 10.3390/mti4040088

Chapter 7

"In-the-wild" Notifications on smartglasses

7.1 Introduction

In addition to the control studies performed in the previous part of the dissertation (see Chapters 5 & 6), it was important to my work to understand what users perceived if they wore smart glasses for a longer period of time. This is critical because most notifications users receive are messages or notifications from social media websites [18]. This makes notifications something inherently personal since they directly relate to the user. Pielot and Rello [28] showed that not receiving notifications made participants anxious that they would miss important updates, showing that one of the essential factors behind notifications is that they relate to the user personally. This relation is very difficult to replicate in a lab environment, as their controlled nature makes it challenging to send personal notifications to the participants that could be important to them at that exact moment. The context of the current task also plays an essential part in notification delivery. In our experiments, participants always pursued a specific controlled task, during which they would receive notifications. However, in their daily lives, notifications are delivered throughout the day, with possibly dozens of different usage scenarios. Also, in laboratory settings, the exposure to the stimulus is limited, as most experiments last only up to an hour or two. To gain a better understanding of notifications on smartglasses, we conducted an "in-the-wild" experiment [68], where participants used a pair of smartglasses for five days in their daily life while receiving visual notifications on it. To the best of our knowledge, this work is the first time real notifications on smartglasses have been evaluated, especially over a longer period.

7.2 Related Work

In this dissertation, we have presented previous research on notifications on XR devices (see chapter 4). However, that work largely focuses on exploring different UI methods to create no-

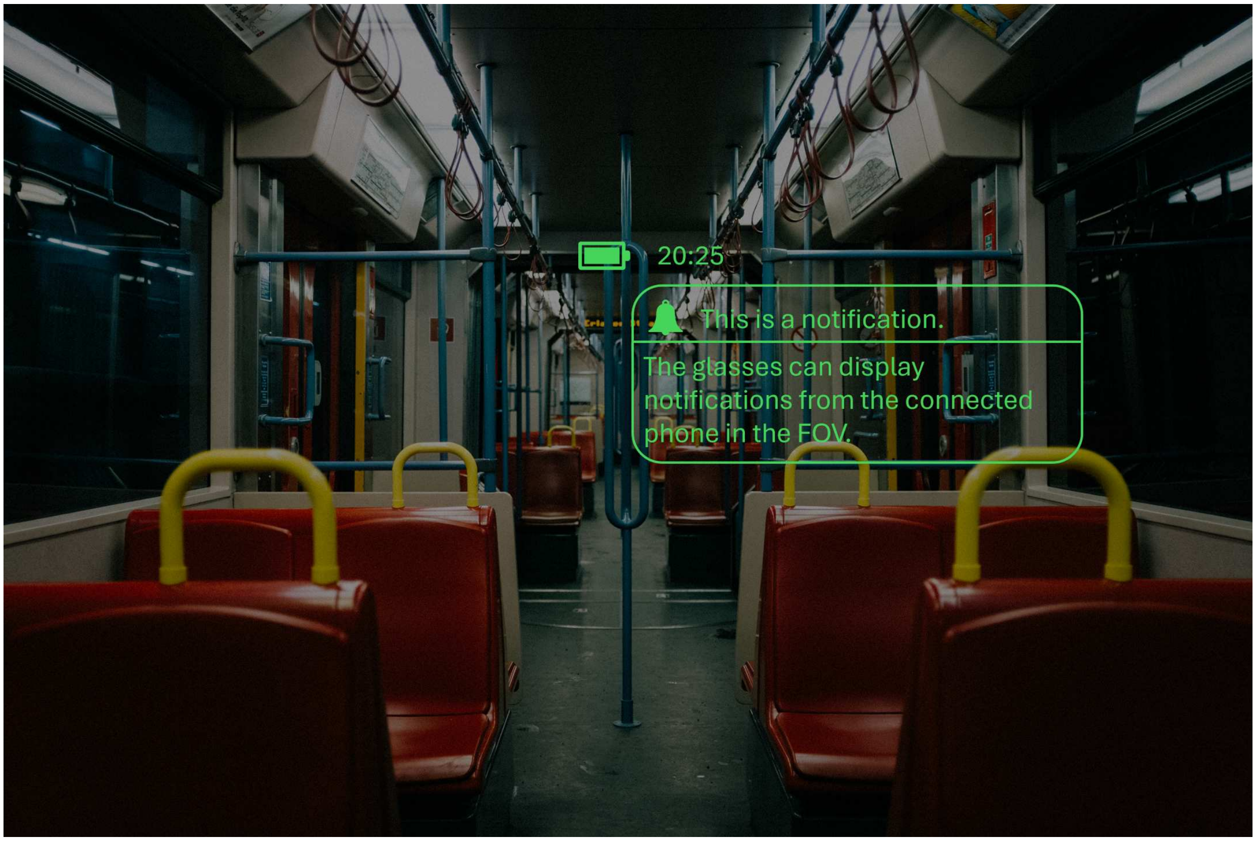
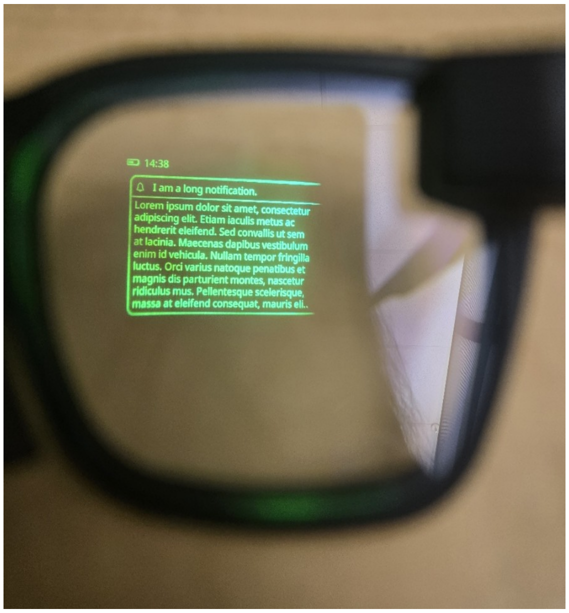


Figure 7.1: A mock-up of what the notification display of the smartglasses looks like from the point of view of someone wearing the glasses.

tifications, and not on the subjective experience of receiving notifications on head-worn displays. Smartglasses like the glasses used in this experiment are not yet commonplace. This means we run into the Collingridge dilemma [84]. The dilemma states that when a technology is not yet widespread, it is easy to control. However, its impact on society cannot be predicted. When a technology is widespread, we know its impact, but we can no longer easily control it. So before smartglasses become ubiquitous, it is important to research their impact, even on a small scale.

Hofmann et al. [85] identify privacy issues as a major concern people have about these devices in their meta-analysis of applicable literature surrounding smartglasses. Most of these concerns relate to the perceived intrusion of privacy because of cameras mounted on the glasses. With more recent devices like the Meta Ray Ban smartglasses [86] that are almost indistinguishable from normal glasses, these privacy concerns could be amplified since onlookers might not even be aware that the glasses are a piece of technology that could potentially record them [87]. Rauschnabel et al. [88] confirm this in a survey about technology adoption, which found that while all the benefits of smartglasses drive the decision to buy a device like this, the only downside driving the purchase decision is the risk to the privacy of *others*. In a second experiment consisting of unstructured interviews, they found that participants weren't concerned with their own privacy as much since they felt they had nothing worth hiding in case of unauthorized access or because a breach of their own privacy would not be as immediate as an onlooker being uncomfortable around smartglasses. They also found that many respondents reported that they accepted living in a world where privacy doesn't exist anymore.

While Rauschnabel et al. [88] findings were on a hypothetical level only, they are supported by McNaney et al. [89] that have conducted five-day field trials, where four patients with Parkinson's disease were asked to wear a pair of Google Glass smartglasses [43] at home and in public to aid with their disease. Three out of four participants reported feeling uncomfortable because of the attention they received from wearing the glasses. Besides concerns with this attention, there were some concerns about the reliability of the interaction with the device, but opinions about it were otherwise largely positive. This indicates that concerns were mostly not centered around the user but about the onlookers. Besides this work, there are hardly any studies on longer exposure to XR. Grubert et al. [90] investigated the use of AR in an industrial task, where subjects were exposed to smartglasses for four hours, while completing a search and pick task. The task was completed twice, split across two days, with participants using either AR or no AR on the first or second day. Participants using AR performed significantly better in the task, but half of the participants reported to have experienced visual fatigue. This experiment, however, was also conducted in a laboratory using students and professionals who were not trained in their everyday routine. Looking at VR exposure, Steinicke et al. [91] exposed one HCI researcher to an immersive virtual environment for 24 hours, with 10-minute breaks after 2 hours of exposure. They found significant issues with simulator sickness and eye strain because of the lower resolution of the device. In a similar but broader study, Biener et al. [92] gave VR headsets to their participants, with a virtual work environment that would resemble a normal desktop environment. They were asked to complete a five-day,



A long text notification on the smartglasses.

A short text notification on the smartglasses.

Figure 7.2: Notifications displayed on the glasses. Please note that capturing this view with a camera is challenging and not entirely representative. The right side of the view is slightly cut off in this capture.

eight-hour workweek in this VR environment instead of their usual desks with monitors. They then compared results between the participants who received headsets and participants who continued to work in a physical environment. They found that the use of VR led to severe simulator sickness symptoms and below-average usability ratings compared to the almost identical but physical work environment. Two of their 16 participants had to drop out because of severe nausea and migraine.

7.3 Methods

We conducted an experiment where participants received Vuzix Z100 [93] smartglasses including a heads-up-display for a period of five days, to better understand the experience with real notifications appearing in the FOV, for more time than what is possible in a controlled experiment. The glasses paired with the participants' phone and display any notifications received on the paired phone. They have a monochromatic green display in the center of the right lens with a 640 by 480 resolution and a FOV of 30 degrees. Over the course of five days, participants were asked to wear the glasses for at least one hour before noon and one hour before the evening. While the glasses are also available in a tinted version, we provided participants with the non-tinted version, which we paired with an optional sunglasses clip, that could be mounted on the glasses for use in bright settings. We validated this to work beforehand. Participants also installed an open-source research app on their phones that logs their notifications [94] without any of its content. We modified the code to indicate whether the glasses were currently being worn when the notification was sent, to understand the frequency of notifications in general and on the glasses. Every day at noon and in the evening, participants were asked to write a note about their usage of the device. At the end of the five-day period, participants were asked about their experience with the glasses in a semi-structured interview. Each interview included all the following questions, with more questions depending on their responses:

- How would you rate the experience overall?
- Can you tell me a bit about your typical day?
- Can you describe a situation where the glasses helped you achieve a goal, and what made it effective?
- What was your first impression when you started using the glasses, and how has that changed over time?
- In what kind of situations did you decide to take off the glasses?
- Did the glasses make you change your thoughts on your phone notifications?

The notes and interviews are then reviewed using thematic analysis [95], where the data is first coded and then grouped into themes to extract the main points and common meanings that recur throughout the data. Coding was done using affinity mapping [96]–[99] (see Appendix B). Interview transcripts have been made using the Distil-large model [100] running locally on the author's machine. Transcripts were then double-checked for mistakes and cleaned for easier analysis. Cleaning involved removing duplicate words and filler words such as “umm” or “like”. In the notification logs, notifications are filtered not to include hidden system notifications or notifications

Table 7.1: Notification average per day

Day	Notifications average	Notifications average with smartglasses
Monday	63	50
Tuesday	185	127
Wednesday	180	87
Thursday	177	106
Friday	191	126
Total	159	99

from media playback. The category of notifications depends on the category the app's creator gave it. Not all apps producing notifications have included a category; therefore, some notifications will be categorized as "NA". We have also cleaned the log from duplicates if multiple notifications were logged by the same app in the same second. Since the notification logging app would only work with Android, participants had to own and use a smartphone running Android OS 12 or newer. It has been found that there are no significant differences in personality depending on mobile operating system choice (Android or iOS) [101] and both platforms handle notifications in generally similar ways, so this population constraint should not restrict our results.

7.4 Results

For the experiment, we recruited eight participants aged from 22 to 29 years old (M = 26.8, SD = 2.3), with four males, three females, and one declining to answer. No other genders were reported. Vision was either normal (20-20) or corrected to normal with contact lenses during the experiment (two participants). Participants were asked to turn off their smartwatch notifications if they had one. Only two participants reported using a smartwatch, which is in line with the age demographic of smartwatch ownership [102].

7.4.1 Notification Logs

One of the participants encountered issues with the notification logging app, so their data has been excluded from the quantitative analysis, leaving 7 participants' log files. Overall, after notifications were filtered, 5863 notifications were logged. We found that our participants received

Table 7.2: Categories of notifications

Category	Number total	Smartglasses Number
Message	2820	1831
Other	1566	989
Email	624	378
Social	389	195

an average of 159 notifications daily, with an average of 99 received while the glasses were connected and turned on daily. Looking at the categories of notifications, we see that our results align closely with similar work, with messages (51.1%) and emails (10%) being the two most frequent notification categories. Several applications did not provide a category, so we cannot be sure what category the notifications were in. As for the phone settings, we saw that 46% were received with the phone on silent, 11% on vibrate, and 43% with sound. With the smartglasses connected, the numbers change slightly to 48% silent, 11% vibrate, and 41% sound. Looking at this setting on a per-user basis, we see that users largely did not change their ringer mode, and most notifications were delivered in one setting throughout the experiment.

7.5 Interviews & Note Results

Our interviews and the journal notes the participants sent were combined in an affinity diagram with 406 notes. These notes were clustered into more minor themes, and then those clusters were again clustered into seven larger topics, which we will summarize in the following. Asked to give a rating from 1 to 10, with one being the worst and ten being the best, participants rated the smartglasses a 6.3 on average.

7.5.1 Approach for Notifications on Smartglasses

Interaction

Participants had many thoughts about how they would like to receive notifications on the smartglasses. First, participants mentioned they would like to reply to certain notifications on the glasses directly, inline with several previous experiments in laboratory settings [60], [79]. P5 reported

“I definitely think there needs to be an interactive component a bit more where I can actually essentially handle the messages using just the glasses”. Participants expressed a preference for hands-free interaction, especially voice commands, to handle notifications. However, others were skeptical, with P6 stating they “couldn’t imagine a way of how to do that in a way that’s not terrible”. Another concern about interaction was the dismissal. The glasses had a button on the frame that could be used to hide a notification that was currently on the screen, but participants reported that they were often in situations where they couldn’t or didn’t want to press the button to dismiss. P5 was explaining a situation in which they were presenting something and would have had to constantly press the button to dismiss notifications to not be distracted. Other participants reported that they would have liked a hands-free way to get rid of distracting notifications, when they couldn’t use their hands but were distracted by the smartglasses.

Presentation

A second area of response was the way the notifications are presented. Participants thought that they would often prefer not always to see the entire text of a notification, but would prefer an icon or just a short snippet, with the option to then access the rest of the notification when it was convenient to them. This access to notifications also extends to older notifications, since participants wanted to be able to read all pending notifications on the smartglasses without needing to get their phones to do that. Lastly, some participants liked that they could read the entire notification immediately but noted that lingering notifications introduced unnecessary visual clutter and affected their concentration

Importance

Probably the most frequently mentioned area was the importance of notifications or their content. Participants reported that the priority or importance of notifications should dictate how they are presented, or if they are presented at all. To begin with, it was reported that notifications for personal messages like instant messages or emails were the most important to them. This is reflected in the log files, where these categories accounted for most notifications received. Five out of eight participants stated that they do not want the smartglasses to display any notifications that are not relevant to them and keep those on the phone. One key factor for them to determine importance was whether a notification had any time-critical information in it. Participants felt that those notifications provided the most utility on the smartglasses since it helped them address things immediately, like reminders about calendar events that were coming up or messages that required quick responses.

7.5.2 Improvement Opportunities

From the participant interviews, we found several opportunities where the smartglasses could be improved. A large part of the participants (five out of eight) expressed that they felt the smartglasses should look even more like a pair of real glasses. P2 said, "I probably would be more open to wearing them a lot if they just look like normal glasses." Three participants, including one that thought they should look more like glasses, said that while they didn't like the looks, it didn't stop them from wearing them. In addition, participants expressed that they would want more features from the smartglasses. A commonly named feature was the inclusion of especially audio input and output (I/O) as well as more display capabilities, like glanceable information about upcoming calendar events or timers. Some participants also mentioned that connection with digital assistants paired with audio I/O would be preferred.

7.5.3 Benefits from Smartglasses

Participants have expressed that they often preferred using the smartglasses over their phones, and that using the smartglasses had helped them to use their phone less often. They reported that using their phones would introduce even more potential for distraction, with P4 saying "I'll go pull out my phone to check a notification, and then I'll be like: "Ooh what's this meme my sibling sent me on Instagram," and before I know it, I've killed 20 minutes there". Another reason for distractions seemed to be that participants would have noticed that they received a notification and wanted to check what kind of notification it was. With the smartglasses, they immediately saw what a notification was and whether it required action. Six participants said that they used the smartglasses notifications to determine if a notification was important enough to pause their current task to deal with it. The smartglasses also helped some participants not to worry whether they missed something important, which was another big reason that they liked using the glasses. Four participants said they noticed notifications and especially phone calls only because of the smartglasses, which they would have missed otherwise because their phone is usually on silent mode (which 48% of notifications in the log files were delivered in). P2 said that "it reminded me of my parking expiring, which is a notification I usually miss on my phone".

While there were some reports about the glasses being too intrusive, participants said that the glasses did interrupt them but did not distract them from their task, as they found it easy to get back into their tasks. Especially in tasks where the hands were preoccupied, participants (five out of eight) saw the biggest advantage of the smartglasses. P5 said, " [The smartglasses] feels like it's the most beneficial in situations where [. . .] I'm doing something that requires both hands and I can't actively use my phone, but I would still like to know if there's something that has come up."

Turning to social situations, some participants liked using the glasses more than using their phone because they felt they could check their notifications without the other person noticing or without interrupting the conversational flow by *phubbing* (see section 2.3). Lastly, we saw some general positive sentiments towards the glasses. Two participants stated that they would miss using the smartglasses and would like to own one going forward if they liked the looks. Four participants said that the notification feature alone was enough for them to be interested in such a product, with three of them finding the smartglasses reasonably comfortable.

7.5.4 Notification Ecosystem

One identified theme was what we call the Notification Ecosystem. First, participants reported that the experiment changed their relationship with their smartphone notifications. They reported feeling surprised at how many notifications they receive daily and how many of those they do not need or that stress them. P6 said, "I never knew how much phone notifications really played a role in my life. [. . .] not in my life, in the way I go through my day [. . .] just how much they stress me."

Participants also stated that they will turn off notifications from certain apps after the experiment is over, since they never look at them. The second point of interest was the relationship between the smartglasses and the smartphone. Participants were divided whether they wanted to dismiss phone notifications through the smartglasses, or whether both devices should be treated independently when dismissing notifications. This synchronous connection between phone and smartglasses was also highlighted, with some participants saying that they thought that some notifications were redundant since they were using their phone anyway and disliked getting the notification on their phone and smartglasses simultaneously. Occasionally, the smartglasses would be a little delayed in displaying the notification, which caused frustration. Lastly, it was mentioned that the smartglasses notifications are redundant when you need to respond and have to use your phone anyway for that.

7.5.5 Obstacles for Adoption of Smartglasses

Participants also expressed their issues with the smartglasses, which would keep them from owning them. While some initially found them novel or convenient in certain situations, they felt that the benefits did not outweigh the frustrations. Some noted that they do not get enough important notifications to warrant wearing the glasses. Overall, four participants said they wore the smartglasses primarily because of the experiment but would not choose to use them in daily life. Besides this, participants experienced general discomfort using the smartglasses. An issue that four participants reported was that looking at monitors or televisions while using the glasses made them experience eyestrain or headaches. They also noted that wearing them made the monitors or televisions harder to read. Another issue was the brightness of the display. Since the device lacks auto-brightness, participants found themselves in situations where the display was uncomfortably bright in a dark environment or where it was not bright enough for outdoor settings, even with the sunglasses attachment. Some participants just felt that the frame design itself was not comfortable, however a larger obstacle for the smartglasses is the fact that all but two participants do not wear vision-correcting glasses in their daily life. Four participants said they were uncomfortable just wearing any type of glasses daily, not specifically the smartglasses. One of the two participants who regularly wear glasses stated that they would absolutely wear an smartglasses instead, if it looked like their normal glasses. Lastly, there were some situations that participants encountered where they had problems with the smartglasses. A common challenge that AR faces is the contrast between the real world and the display; our participants also ran into this issue. Four participants reported that they frequently had to face a uniformly colored wall to be able to read the text on the smartglasses. Besides issues reading the screen, some participants stated that there are occasions where they would not want to have any kind of notifications, causing them to take off the glasses entirely. Some also reported that receiving many notifications in a short time from for example group chats would cause them frustration. The smartglasses's hardware was also mentioned as a problem. Three participants had liked to wear them while cooking but reported that the smartglasses fogged up more than normal glasses during cooking or that the lenses smudged more easily. Both of these problems probably occurred because the smartglasses are made of plastic, not glass like normal lenses. There were also concerns about durability, as participants feared they would break the borrowed device. Lastly, some participants encountered small software glitches, like animation freezes or notifications re-appearing after they had already been dismissed, though these were rare occurrences.

7.5.6 Societal Challenges

A significant theme found was the societal challenges that participants faced. The first area of concern is privacy. It is worth noting that the smartglasses used in the experiment does not have any recording capabilities since it lacks a microphone or camera. However, participants still felt uncomfortable in some situations because other people asked about being recorded. This caused some participants not to want to wear the smartglasses around other people. While some participants would rather not wear them in public because of their looks, most participants expressed not feeling comfortable about wearing them around others because of the attention they were getting or because they felt self-conscious about others thinking they were being recorded (six out of eight).

Participants said that they felt rude using the smartglasses in conversations or that they were drawn out of conversations when they received a notification during.

7.5.7 Attentional Costs of Smartglasses

The interviews made it clear that participants found the smartglasses notifications distracting. Participants found that getting notifications on the smartglasses was more intrusive than receiving them on their phones and that they were difficult to ignore perceptually as well as ignore its content. P1 stated that "When I was ready I could pick my phone back up, but with the glasses, I felt like I didn't have a choice". In some instances, participants also reported that the smartglasses impeded their vision of the real world, primarily the display and the black frame. Some participants also reported that this meant they were heavily distracted from complex tasks, causing them not to wear it during those tasks.

7.6 Discussion

One of the key findings from our study was that notification management in smartglasses must balance usefulness with potential distraction. Participants generally placed their phones aside when they wanted to concentrate, indicating that any form of notification can be disruptive. However, this trade-off was considered acceptable when notifications were perceived as time-sensitive or relevant. Smartglasses were particularly useful when the phone was unreachable, either due to a conscious decision to focus or because situational constraints prevented phone use. Participants also reported feeling reassured that they were not missing important updates.

To address this, smartglasses notification systems should implement a more nuanced filtering mechanism beyond binary toggles for each app. Android OS already supports notification channels [103], allowing users to control notification categories at a granular level. Extending this functionality to smartglasses would allow users to receive only high-priority notifications, such as filtering app notifications to include direct messages but not group messages. Based on our findings, we propose the following priority model for smartglasses notifications:

- **Urgent Priority**: Full-text pop-up (e.g., critical alerts)
- **High Priority**: Icon and text preview (e.g., direct messages)
- **Medium Priority**: Icon and pull notification (e.g., reminders)
- **Low Priority**: Only on phone or no smartglasses indicator (e.g., promotional content)

This should be investigated in the future in a controlled experiment.Faulhaber et al. [72] have already looked into priority for AR notifications, and found that participants had challenges to distinguish between medium and high-priority notifications.

Importantly, priority settings should be chosen carefully and be easily customizable based on user preference. Additionally, an intelligent filtering and summarization mechanism that categorizes notifications based on urgency and context—similar to existing implementations in Samsung's One UI 7 [104] or Apple's iOS 18 [105]—could improve user experience by reducing overload while maintaining awareness of essential information.

Our findings also show that smartglasses notifications influence phone usage. Participants indicated that while smartglasses allowed them to check notifications discreetly, interactions beyond passive viewing still required the phone. This suggests that smartglasses notifications should help users decide when retrieving their phone is necessary rather than completely replacing phone interactions.

Interacting with smartglasses notifications also posed a challenge, as our participants expressed a strong preference for more discreet and hands-free interaction methods, such as voice input, to reply to notifications. Participants reported that without the ability to reply, they would often still have to reach for their phones, which reduced the benefit of the smartglasses. These findings align with section 6.2, so we suggest a layered approach to interaction:

Primary Input: Touch-based interactions for reliability and when discretion is unimportant.

Secondary Input: Unobtrusive alternatives like gaze-based interaction, voice commands, or multimodal combinations (e.g., gaze + pinch gestures).

Another notable finding was that participants who received fewer high-priority notifications often felt that the benefits of smartglasses notifications did not justify the frustrations. This suggests that smartglasses notifications are most valuable for users with frequent critical updates rather than general-purpose notifications. The IRC model [17] could be applied here, ensuring that only notifications with an interruption level of 1 (i.e., critical alerts) are enabled by default on smartglasses, reducing unnecessary distractions.

Despite their functional benefits, smartglasses notifications face significant challenges in social acceptability. A primary concern was the perception that smartglasses users might be recording others, even when the device lacked a camera or microphone. Participants frequently encountered skepticism from bystanders, highlighting a broader public distrust of wearable AR technology. However, most of the insecurity came not from other people but the internal fear that there *could* be a public backlash to the glasses. Participants often said they stopped wearing the glasses out of fear that other people might think they were being recorded, confirming the survey results of Rauschnabel et al. [88] in a real usage scenario.

To address this, designers should incorporate clear visual indicators, such as LED status lights, signaling when the smartglasses are recording audio and/or video. Additionally, physical shutters for cameras—similar to those found on some laptops—could help mitigate privacy concerns and improve public acceptance. This should, however, go beyond one manufacturer. It is unreasonable to expect every person to know precisely which smartglasses have recording capabilities, much less how that particular pair of glasses expresses that they are currently recording. This is highlighted in our findings, where onlookers were worried about their privacy, even though the smartglasses doesn't have cameras. Another self-conscious aspect of wearing the smartglasses was the looks of the device. Participants expressed that their not liking the way they looked on their faces was a big factor when deciding whether to wear them in public or not. This highlights the need for manufacturers to offer several frame designs for their product and to get inspiration from common and popular frame designs.

Beyond social concerns, physical comfort also emerged as a significant barrier to adoption. Participants, especially those not regularly wearing prescription glasses, reported discomfort when wearing smartglasses for extended periods. This suggests that smartwatches may be an easier entry point for users seeking wearable notification systems, as they do not require continuous wear on the face. The downside is that those notifications might be miss-able more often than smartglasses.

Other ergonomic challenges included eyestrain, display contrast, and frame design. Future smartglasses should prioritize lightweight, adaptable designs that cater to a diverse range of users. Features such as adjustable brightness and electro-chromatic dimming lenses could improve visibility in varying lighting conditions, enhancing overall usability. While some of these issues may be specific to the hardware used in our study, they nonetheless provide valuable insights for future smartglasses development. It should also be considered that users might wear the glasses while looking at a monitor for large parts of the day, so ensuring clarity of both the smartglasses display or the display behind it is critical.

7.6.1 Limitations

One limitation of this experiment was that we told participants to wear the glasses as much as possible during the experiment. This could have caused participants to wear them longer than they would have otherwise, potentially adding to their frustrations.

Another limitation was the participants we used in the experiment. Only two of the participants (25%) wear vision-correcting glasses normally but corrected their vision with contacts during the experiment. According to the Vision Council [106], 63.7% of the US adult population wears prescription glasses, making our participant pool slightly skewed. Reports about the discomfort of wearing glasses might be exaggerated compared to the general population.

7.6.2 Future Work

In the future, we would like to implement and study different types of displays for notifications based on priority and context. Our results show that priority and timeliness are key determining factors for how intrusive AR notifications can and should be.

7.7 Conclusion

We investigated the use of smartglasses for notification delivery. For this, we recruited 8 participants who would wear a pair of smartglasses for a period of five days, during which it would display their phone notifications. Participants were asked to create journal notes and send them to us, and were interviewed after the experiment to inquire about their experiences. Our findings highlight both the advantages and challenges of using smart glasses for managing notifications, and shine a light on crucial design considerations for future AR systems. While participants valued the ability to access notifications hands-free, reducing distractions, and improving notification awareness, key barriers such as interaction constraints, attentional costs, and social discomfort remain significant hurdles to adoption.

Chapter 8

Aiding task resumption using AR cues

My work has so far been focused on the UI implications and problems around AR notifications. But as we have laid out in section 2.2 and as we have seen in our experiment "in-the-wild" (see chapter 7), notifications of any sort can be perceived as highly interruptive during tasks and can negatively impact performance. The use of AR could be able to help reduce these interruptions and decrease the time needed to resume the interrupted task. If the AR system understands the user's context and wants to deliver a notification, it could prepare the user for an incoming interruption and then help the user resume the task with minimal impact. The following paper [5] investigates how AR cues can facilitate task resumption after interruptions, and shows that using AR cues can reduce the resumption lag after short and long interruptions [107].

[5]This work was published as Bahnsen, Kilian L, Lucas Tiemann, Lucas Plabst, and Tobias Grundgeiger. "Augmented Reality Cues Facilitate Task Resumption after Interruptions in Computer-Based and Physical Tasks." In Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–16. Honolulu, HI USA: ACM, 2024. HTTPS://DOI.ORG/10.1145/3613904.3642666.

Augmented Reality Cues Facilitate Task Resumption After Interruptions in Computer-Based and Tangible Tasks

KILIAN L. BAHNSEN, Julius-Maximilians Universität Würzburg, Germany
LUCAS TIEMANN, Julius-Maximilians Universität Würzburg, Germany
LUCAS PLABST, Julius-Maximilians Universität Würzburg, Germany
TOBIAS GRUNDGEIGER, Julius-Maximilians Universität Würzburg, Germany

Many work domains include numerous interruptions, which can contribute to errors. We investigated the potential of augmented reality (AR) cues to facilitate primary task resumption after interruptions of varying lengths. Experiment 1 (N = 83) involved a computer-based primary task with a red AR arrow at the to-be-resumed task step which was placed via a gesture by the participants or automatically. Compared to no cue, both cues significantly reduced the resumption lag (i.e., the time between the end of the interruption and the resumption of the primary task) following long but not short interruptions. Experiment 2 (N = 38) involved a tangible sorting task, utilizing only the automatic cue. The AR cue facilitated task resumption compared to not cue after both short and long interruptions. We demonstrated the potential of AR cues in mitigating the negative effects of interruptions and make suggestions for integrating AR technologies for task resumption.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; **Mixed / augmented reality**.

Additional Key Words and Phrases: Task Resumption, Interruption, Augmented Reality, Resumption Lag, Human Error

ACM Reference Format:
Kilian L. Bahnsen, Lucas Tiemann, Lucas Plabst, and Tobias Grundgeiger. 2024. Augmented Reality Cues Facilitate Task Resumption After Interruptions in Computer-Based and Tangible Tasks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 25 pages. https://doi.org/10.1145/3613904.3642666

1 INTRODUCTION

Interruptions are frequent in many work domains, such as healthcare [15, 24, 32, 39], aviation [63], software development [1] or office work [19]. Interruptions decrease efficiency [3, 89], lead to higher error rates [23, 47, 83, 94] and decrease satisfaction with one's own work performance [6]. Although efforts have been made to avoid disruption altogether [16], this approach requires deep workflow restructuring and is not applicable to all tasks. For example, time-critical tasks that must be addressed immediately may occur and thus require the interruption of the primary task. A different approach to interruption management is to attempt to mitigate the negative effect of the interruptions. For this purpose, various types of visual cues have been explored, demonstrating varying degrees of efficacy. For example, carrying a syringe enhanced task resumption in nurses [31]. Conversely, cues, such as the simple act of placing the mouse cursor on the correct task point for task resumption did not improve task resumption [90].

We investigated whether cues on a head-mounted augmented reality (AR) display can mitigate the negative effect of interruptions. The application of cues via AR holds the promise of being task-independent, possible for task were users are mobile and customizable, but, to the best of our knowledge, the efficacy of AR cues has not yet been investigated.

Manuscript submitted to ACM

Based on the so-called Memory for Goals theory [2], we used the resumption lag (i.e., the time between the end of the interruption and the resumption of the primary task) to measure the disruptive effect of an interruption on the performance of the primary task. In two experiments, we investigated whether AR cues at the point of task resumption could reduce the resumption lag. In Experiment 1, we used a computer-based primary task, which is frequently used in the context of the Memory for Goals theory. We chose such a task to test that AR cues can be studied using the Memory for Goals theory and to investigate what kind of AR cues provide good task resumption support. More specifically, we investigated whether a cue automatically placed by task-tracking software or manually placed through a hand gesture by the user provided better support for task resumption. In Experiment 2, we used a tangible primary task to investigate the effectiveness of the automatically set cues in a task that more closely resembled a tangible working task, such as medication sorting. Considering the relevance of spatial memory for task resumption [76], our goal was to evaluate how cues performed in a task with a more notable spatial dimension than the computer-based task in Experiment 1. We contribute to HCI-interruptions research by providing (1) empirical data that show that AR cues in computer-based and tangible tasks facilitate the resumption of primary tasks after an interruption; (2) empirical data that show that manually set cues provide no further advantage over automatically set cues; and (3) a new task paradigm for studying interruptions of tangible primary tasks that allows for easy adaptation of task complexity.

2 RELATED WORK

2.1 HCI Research on Interruptions

In the past two decades, interruptions have been extensively investigated in the field of HCI [64, 65]. Salvucci et al. [82] introduced a Multitasking Continuum, encompassing concurrent multitasking and sequential multitasking. Concurrent multitasking includes more task switches and shorter actions on the different tasks and frequently requires self-initiated switching (e.g., driving and talking). Sequential multitasking is characterized by fewer task switches and longer periods of attention devoted to the single tasks (e.g., cooking and reading a book). Interruptions are considered sequential multitasking, and numerous studies have highlighted the disruptive effects of interruptions in laboratory settings and various work domains, including, for example, office work [19, 45], computer programming [102], healthcare [25, 34], aviation [57, 63] or assembly work in manufacturing [50].

Early work on interruption coordination by McFarlane and Latorella [65] differentiated between immediate, negotiated, mediated and scheduled. HCI work attempted to reduce the disruptive effects of interruptions through mediated or scheduled interruptions [44, 87, 102]. For example, Iqbal and Bailey [44] developed and evaluated a notification management system for computer-based work. The system reduced frustration and confirmed the users' preference to schedule interruptions during phases of low workload. Whereas Iqbal and Bailey's system used task data to determine the next fine, medium or coarse task-breakpoint, researchers also used psycho-physiological sensors [102, 103] or the sensors of an off-the-shelf virtual reality device [14] to predict the interruptibility of users. Finally, a combination of task and physical data seems to provide the best interruptibility predictions [103].

A large corpus of HCI work addresses the negotiation of interrupting tasks to understand when interruptions are less disruptive and more acceptable. For example, laboratory studies showed that users prefer to monotask and accept interruptions more often in phases of low workload [43, 80], but that time constraints, distance to the next good breakpoint [10] and task properties [33, 93] also influence the decision to accept an interruption.

Finally, immediate interruptions have been addressed due to the difficulty of resuming the task at hand after handling the interruption [65]. In healthcare, immediate interruptions are the most prevalent interruption coordination method [34, 75] and definitions of interruptions in the healthcare context have been characterized by an unexpected and immediate onset that results in a temporary break in the task at hand [11, 32]. For example, in the emergency department, 75.4% of all interruptions were immediately attended to, whereas 22.2% were coordinated via concurrent multitasking and only 2% were negotiated or rejected.

2.2 Quantifying and Mitigating the Effect of Immediate Interruptions

To quantify the effect of immediate interruptions on primary task resumption, Trafton et al. [91] suggested measuring the so-called resumption lag. The resumption lag is the time from the end of an interruption to the continuation of the primary task. Longer resumption lags indicate a more difficult task resumption compared to shorter resumption lags. The idea that the resumption lag is sensitive to interruptions is built on the activation-based Memory for Goals theory [2]. According to the Memory for Goals theory, the most active goal at any moment directs a person's behavior. Furthermore, the level of activation of a suspended (i.e., interrupted) goal decays over time. The higher the activation level of the suspended goal, the easier it is to resume the suspended task, which is indicated by short resumption lags. The activation level of a goal can be raised by providing environmental cues that are associated with the goal. In this context, a cue can be any perceptual signal that is directly associated with the goal or guides the individual towards something in the environment that is associated with the goal [92].

The Memory for Goals theory and the resumption lag construct have received empirical support in various studies. Studies in laboratory [52, 54, 61] and field settings such as intensive care nursing [31] showed that longer interruptions (i.e., more goal decay of the to-be-resumed primary task) resulted in longer resumption lags. Furthermore, longer interruptions were associated with forgetting to resume the interrupted tasks in emergency departments [24]. Interruptions at moments of low workload resulted in smaller resumption lags [43, 102]. The positive effects of environmental cues on the resumption lag have shown that visibility of the primary task during the interruption reduced the resumption lag compared to occluded primary tasks [40, 74]. Furthermore, providing visual cues at the end of the resumption in the form of mouse cursors at the to-be-resumed location also reduced the resumption lag in some studies [5], while it did not in others [90]. Besides general environmental cues, additional artifacts can be used as cues to facilitate primary task resumption. Such cues might be artifacts that are relevant for reaching the goal (e.g., healthcare workers holding on to a syringe during the interruption [31]), as well as additional cues such as sticky notes [28] or reminder systems designed for specific tasks such as unmanned aerial vehicle supervision [84].

Recommendations to diminish the negative effects of interruptions can be summarized in two general strategies [65, 75]. The first recommendation is cognitive rehearsal, which increases the activation of the to-be-resumed task's memory representation. Although effective in laboratory settings [4, 66], cognitive rehearsal has not yet been observed in field settings [31]. Furthermore, the nature of immediate interruptions may restrict users from conducting cognitive rehearsal before attending to the interruption or during the interruption.

The second recommendation is to maintain the activation of the to-be-resumed task's memory representation high via environmental cues. This may include visual access to the primary task during the interruption [74] or post-interruption cues, such as artifacts in the environment [31], a mouse cursor [5], marked text passages [17, 18, 90] or open software in the foreground [45]. Placing environmental cues in form of mouse cursors or marked text passages in desktop-based tasks at the end of the interruption has been shown to be effective in assisting task resumption in some [13, 90] but not all cases [17, 18, 90]. Salient cues that provided spatial information about the to-be-resumed tasks seems to be the most effective method [13, 90].

However, using physical artifacts or computer screen-based cues has drawbacks. These cues require that the user has to either stay at the location of the primary task, maintaining the primary task in their field of view for visual perception, or carry the cue along during the interruption. Furthermore, such cues may be only used for the specific task or sub-task that they have been design for. AR cues do not face these disadvantages, as they can be displayed irrespective of location, are task independent, and can be operated hands-free. These represent distinct advantage, particularly for environments where interruptions often require the user to change the physical location or work on different task that may include a computer screen or are tangible tasks. For example, in intensive care units [31] or emergency departments, nurse or doctors need to be mobile, need technology that can be used hands-free and satisfies a bare below the elbows policy and have computer-based tasks such as charting but also manual tasks such as setting up infusions.

2.3 Augmented Reality and Interruptions

In recent years, there has been a surge in research dedicated to utilizing the seamless integration of digital information via AR in fields like education [8, 21], architecture [59] and remote assistance [36, 69]. The inherent adaptability of head-mounted AR displays yields potential for application in interruption-rich environments, particularly in domains such as healthcare [30, 86], aviation [97] manufacturing [72, 100, 101] and assembly tasks [41, 42, 55, 79]. Researchers started to investigate AR systems for patient monitoring in intensive care units [49, 78]. These developments enable real-time data access and visualization, providing healthcare professionals with location-independent and hands-free information. However, providing information on head-mounted displays also may be a further source of interruptions. For example, supervising anesthesiologist who used a head-mounted display with information of six ongoing operations felt an obligation to consider vital sign changes in the operating rooms and interpret the situation. This resulted in looking at the head-mounted display with a "frozen" head posture and 10-25 second-pauses of the ongoing task [53].

In manual assembly tasks, instructions presented via AR reduced cognitive load and received higher usability ratings compared to print instructions [55]. Similarly, for warehouse order picking, workload and performance showed benefits for head-mounted display compared to all other methods [35]. Because interruptions are more disruptive during periods of higher cognitive load [43], the potential improvement in task resumption by transferring tasks onto the AR glasses warrants further investigation. Studies have demonstrated the effectiveness of AR-based task guidance in supporting car repair, particularly for workers facing novel tasks [41, 42] and performance was better for central presentation compared to peripheral presentation on the head-mounted display [99]. In general, task guidance via AR offers potential benefits for reorienting workers after interruptions, but does not offer explicit support for task resumption by leveraging AR cues. Unlike previous research, such as that by Rukubayihunga et al. [79] on assembly step recognition in AR, which lays the groundwork for automatic deployment of cues, our study focuses on utilizing manually or automatically set cues in a more versatile setting.

Finally, recent research has also addressed how one should interrupt head-mounted display users in virtual reality [14, 26, 70]. For example, researchers have collected sensor data of VR devices to predict interruptability to schedule interruptions [14]. Despite this interest in interruptions, the current research is rather limited regarding task resumption support via AR cues. More related to immediate interruptions and the need for task resumption, a study investigated the effect of interruptions on opaque or transparent smart glasses on the resumption lag [52]. Based on the Memory for Goals theory, researchers expected that transparent smart glasses would provide environmental cues of the to-be-resumed task while attending to the interruption on the smart glass. However, compared to opaque smart glasses or a tablet control condition, transparent smart glasses did not reduce the resumption lag. It may be that the focus on the smart glass in order to work on the interrupting task resulted in inattentional blindness for the environmental cues [51], and as a result, the cuing was not effective. Consistent with the findings of Trafton et al. [90], who showed that subtle cues such as placing the mouse cursor at the to-be-resumed task step did not improve task resumption, the mere visibility of the environment might have been too subtle a cue to facilitate task resumption.

Following Trafton et al. [90], the design of our AR cue as a salient red arrow aims to resolve this issue by providing a blatant cue. In addition to noticing the cue, task resumption is also a spatial memory problem [21, 77]. Ratwani and Trafton [77] demonstrated the effect of intact or compromised mental spatial representations on task resumption and suggested that memory for spatial location may guide task resumption. Different to, for example, verbal information or semantic information about the to-be-resumed task step, a red AR arrow can be placed at the exact location and support spatial memory during task resumption. Finally, in relation to above described artifact cuing were cues are most of the time only working for a specific task, we also expect the red AR arrow to be a generic cue design that is task-independent and can work for computer-based and tangible tasks. Wolff et al. [97] showed that such salient cues can improve task resumption by circling the to-be-resumed task step in a scanflow task (i.e., participants had to confirm the color of various shapes in a 48-shape array). While their approach involved a Wizard of Oz paradigm and mock-up AR cues instead of circles on a monitor, making it less versatile compared to our paradigm, they still demonstrated that AR cues can reduce the number of errors and mental load resulting from interruptions. Applying Wolff et al. [97] findings to an actual AR setup would enable us to leverage the task- and location-independent characteristics of AR cues and empirically test whether the benefits translate to actual AR applications for computer-based and tangible tasks.

Despite a broad interest in interruptions in recent AR and virtual reality research and the possible benefits of AR cues in form of location and task independent cueing of to-be-resumed tasks, theory-based empirical research on the effectiveness of AR cues in task resumption is missing. Because the current empirical evidence for the effectiveness of environmental cuing in relation to opaque or transparent smart glasses [52] and subtle cues [90] is not fully conclusive and the only AR study included mock-up AR and Wizard of Oz [97], we decided to conduct experiments in which we compared a no cue condition with the different AR cue conditions. Our study's contributions lie in investigating the effectiveness of manually or automatically set AR cues at the point of task resumption in both computer-based and tangible tasks.

3 EXPERIMENT 1: COMPUTER-BASED PRIMARY TASK

Experiment 1 aimed to examine the potential of AR cues in mitigating the negative effects of interruptions on the resumption lag. Additionally, we sought to determine whether the mode of cue presentation, specifically whether the cue is automatically placed by task-tracking software or manually placed by the user, impacts the effectiveness of the cues. The factor cue was manipulated between subjects and included the three groups: automatic cue (i.e., automatically set cues), manual cue (i.e., cues set via a hand gesture at the start of the interruption) and no cues (i.e., a control group with no cues at all). To examine the potential influence of interruption length on the effectiveness of cues, the study varied the duration of interruptions between 15 and 45 seconds within-subject. Due to the fact that, to the best of our knowledge, this is the first experiment to address the effect of AR cues on the resumption lag, we used a computer-based primary task with a structure that is regularly used to investigate the resumption lag [52, 76]. In this task, participants had to transfer patient information from tables in the center of the screen into their corresponding input fields above and below the tables. In line with resumption lag research [52, 76], there is only a single task to-be-resumed and the focus is on task resumption times. Based on previous research on the resumption lag [67, 90], we hypothesized that a shorter resumption lag would occur with (1) an AR cue (automatic and manual) compared to no cue and (2) short interruptions compared to long interruptions. Additionally, based on the Memory for Goals theory, because the availability of a cue at the end of the interruption should render the activation loss during the interruption irrelevant, we expected that (3) longer interruptions would have a smaller effect when an AR cue (automatic and manual) was presented. Finally, setting the cue manually increases the interruption lag and includes a physical gesture. As longer interruption lags [40] and tangible tasks [25] may increase memory encoding of the to-be-resumed task, we further hypothesized that (4) a manual cue reduces the resumption lag more than an automatic cue. To assess the subjective experience of using a manual cue or an automatic cue, we used the NASA Task Load Index (NASA TLX) [38] to investigate workload differences between the cue types, which is a commonly used metric in interruption research [e.g. 43, 48, 58, 62, 83] and AR research [e.g. 35, 42]. The research question, empirical hypothesis, research design, sample size, and analysis plan for Experiment 1 have been preregistered (https://doi.org/10.17605/OSF.IO/GBX2V).

3.1 Method

3.1.1 Participants. A total of 87 undergraduate university students were recruited via the departments participant recruitment and study management system. Students participated in exchange for course credits or € 12.50. Four participants were excluded (two data sets were lost due to server issues, one participant did not follow the instructions and for another the gesture for setting the manual cue did not function correctly). The final data set consisted of 28 participants in the automatic cue group (22 female, average age 26.6 years), 27 participants in the manual cue group (17 female, average age 23.2 years) and 28 participants in the no cue group (20 female, average age 24.2 years). All participants had correct or corrected-to-normal vision. This research was approved by the institutional review board, and informed written consent was obtained from each participant.

3.1.2 Design. We used a mixed 3 (cue) x 2 (interruption length) design. Participants were randomly assigned to three groups: manual cue, automatic cue and no cue. The within-subject factor interruption length varied between 15 and 45 seconds. We conducted a prospective power analysis to determine the required number of participants for this study. Based on a previous unpublished study, we estimated the effect size of the cue × interruption length to be f = .175 (power = .80, alpha = .05, two-sided test). The power analysis indicated a sample size of 3 × 28 participants.

3.1.3 Procedure and Material. Participants gave written consent and were seated in front of a computer screen, with the tablet on a second table behind them. First, participants had one or, if necessary, several training trials in which we explained the study procedure and let the participants practice the primary task and the interrupting task. The participants had to work on a computer-based task in which they must transfer patient data (Figure 1, number 1) into different input fields. The patient data and the input fields were grouped into six different modules, which had to be completed one after the other. A participant had to start each module by clicking on the "Edit" button (Figure 1, number 2) and finish by clicking on the "OK" button (Figure 1, number 3) of the respective module when they have entered the patient data. Previous data got cleared after clicking on "OK" to not give the participant any cue about the current progress, thereby forcing them to remember their current progress. Clicking on the "OK" button may also have triggered an interruption, indicated by an acoustic signal and the mouse rendering immobile at the center of the screen.

The interruption consists of either three or nine simple arithmetic problems with results ranging from 0 to 30 (Figure 3) which were presented on a tablet that was located on a table behind the participant. Each problem was presented for 5 s. The problems, their answers and their order were the same for all participants. When the interruption ended, an audio cue sounded, and the tablet turned black. The participant had to turn around again and resume the primary task by activating the next module. For the resumption lag, we measured the time between the end of the interruption

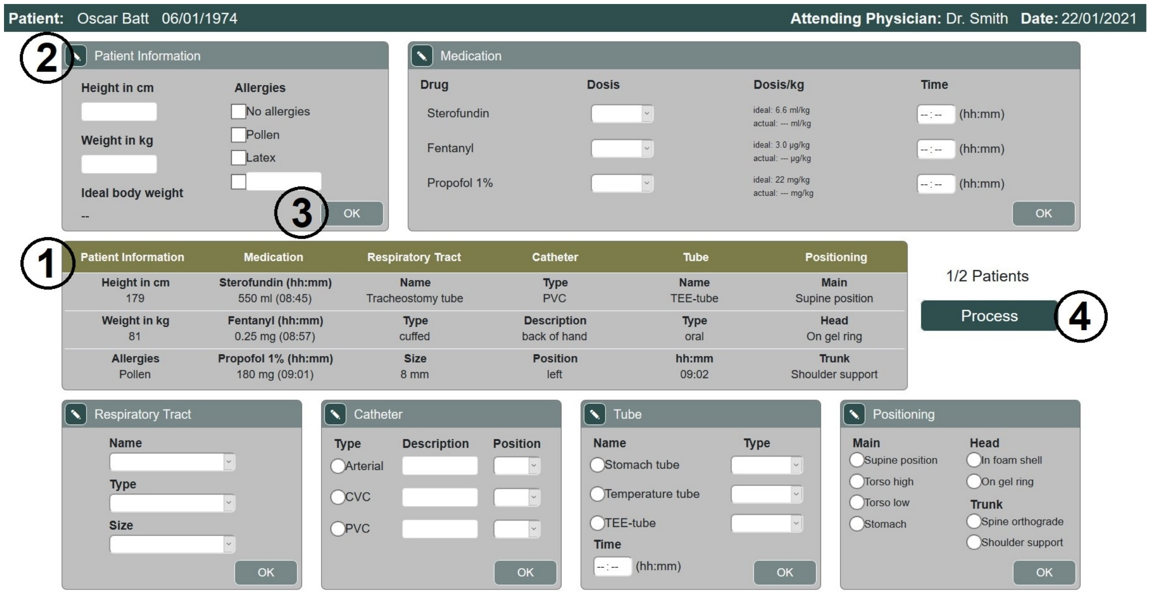


Fig. 1. Screenshot of the primary task interface. Central to the interface is a table containing six columns of different categories regarding the monitoring of an anesthesia. The participants needed to transfer the information from each column into the corresponding surrounding modules. The primary task was developed by Kruse et al. [52].

and the first correct click in the primary task (= pressing the next "Edit" button). Once the participant had finished all modules (i.e., a trial), they needed to progress by clicking on the "Process" button on the right-hand side of the screen (Figure 1, number 4). The timeline of the interruptions is shown in Figure 2.

In the automatic cue and manual cue groups, participants wore Microsoft HoloLens. In the manual cue group, participants had to place the cue themselves at the beginning of the interruption by using an air tap gesture (tapping the index finger onto the thumb and raising it again while pointing onto the primary task). Participants in the manual cue group also practiced the air tap gesture for setting the cue. At the end of the interruption, a three-dimensional, red arrow was superimposed via the HoloLens as a cue in both cue groups (Figure 3). The arrow pointed to the "Edit" button of the module that the participant needed to resume or to the "Process" button. Since participants in the no cue group did not benefit from the HoloLens and the display might be distracting when not in use, they did not wear the HoloLens.

Following this, the main part of the experiment commenced. Participants were tasked with inputting patient data for a total of eight unique cases, with interruptions introduced in four of these cases. Within each of the interrupted patient data sets, three interruptions of varying lengths were incorporated. In total, participants were interrupted six times for 15 s and six times for 45 s. The time of the interruptions was defined in four configurations, which were equal across the cue groups and were assigned randomly. The configurations allowed no more than two interruptions in succession within a patient data set, no more than three consecutive data sets with interruptions and balanced the order of the interruption lengths within the data sets.

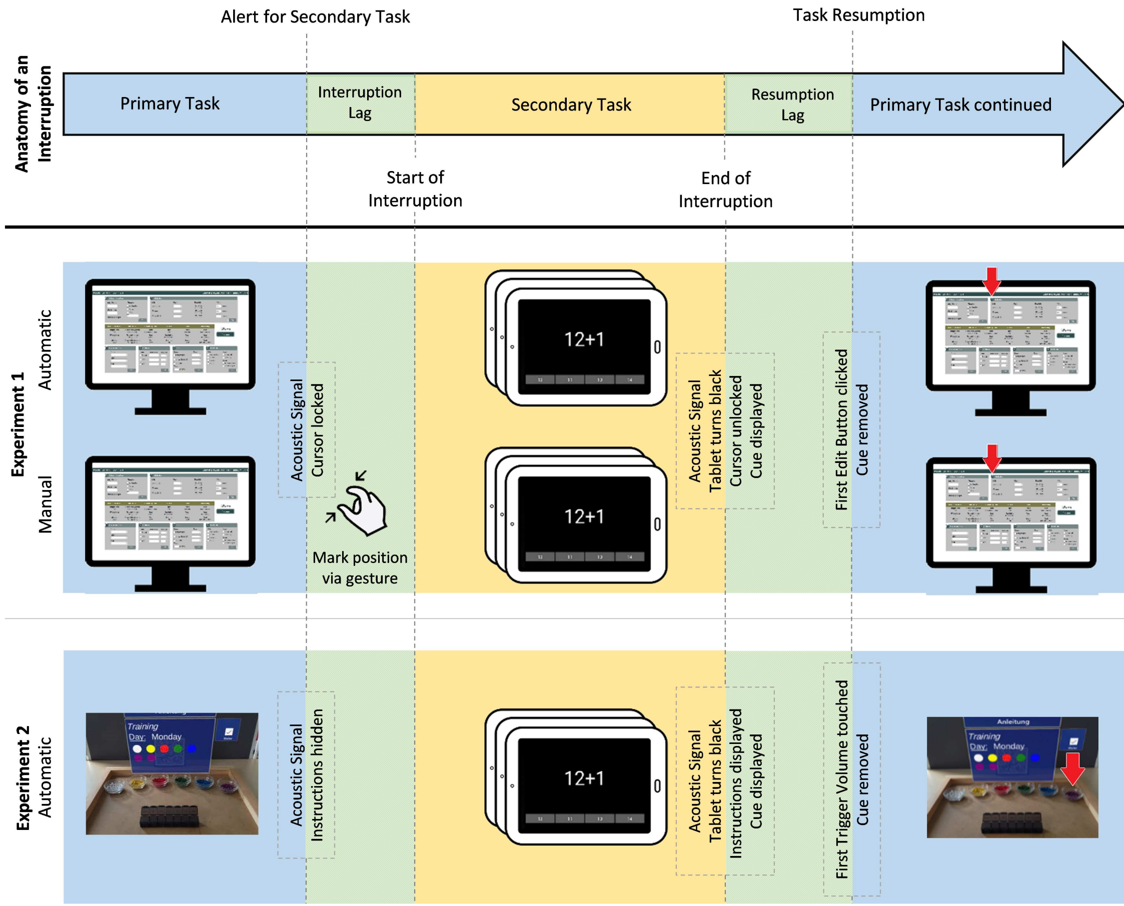


Fig. 2. Structure of interruption (top row) and timeline of cue conditions in Experiment 1 and 2. Acoustic signals indicated the start and the end of the interruptions along with task specific means to prevent a continuation of the primary task. In the manual cue condition of Experiment 1, the interruptions started after the participants marked a position on the primary task via a hand gesture. The resumption lag was measured from the end of the interruption, i.e., the tablet turning black after the last arithmetic task, to the first interaction with the respective primary task. The red AR cue was only present in the cue condition and the arrow was visible during the resumption lag. Note that we enhanced the outline of the AR cues for better visibility in the figure.

Lastly, when participants completed the main part, they were asked to complete a post-survey. Participants completed the NASA TLX [38], provided demographic data and were asked to describe if they used any particular strategy during the experiment.

3.1.4 Analysis. During data preprocessing, interruption lags shorter than 1 second were highlighted and manually investigated. The analysis of these interruption lags indicated that two resumption lags of two participants were invalid because a bug in an early version of the experiment software sometimes automatically set the cue, preventing participants from manually setting it with the gesture. We therefore removed these two resumption lags. Additionally, 31 (3.1 %) resumption lags were removed due to resumption errors, i.e. when the first click after the interruption was

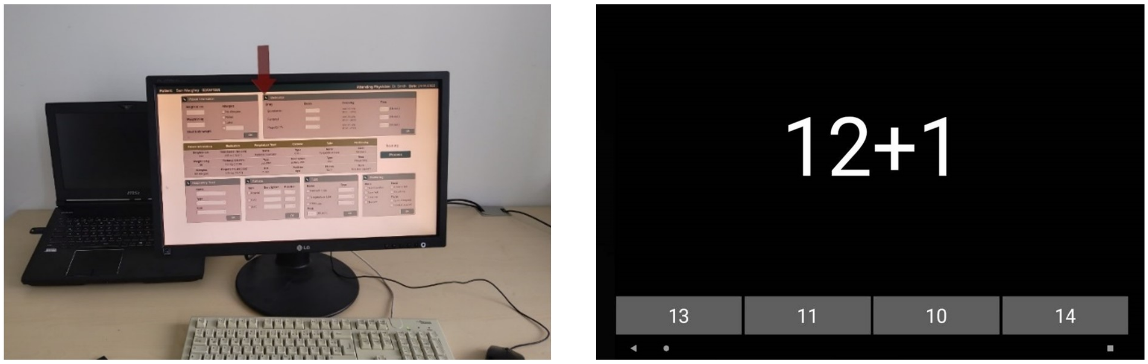


Fig. 3. The primary task (left) with the superimposed cue from the participants' perspective. The cue was pointing to the edit button of the next module and disappeared as soon as the participant clicked on the edit button. The interrupting arithmetic problems task (right) was presented in the center of the tablet, which was placed on a table behind the participants. Four solutions to the problem were presented at the bottom, of which only one is correct.

not on the edit button of the correct module. A total of 963 (96.7 %) interruptions were included in the resumption lag analysis.

We performed a 3 (cue) × 2 (interruption length) repeated measures ANOVA. In three of the six experimental conditions, the values of the resumption lag were not normally distributed (Shapiro- Wilk; all p-values < .05). Because reaction times are most of the time positively skewed [95] we planned to transform the data using the common logarithm, but the transformation did not improve all distributions. As we could assume homogeneity of covariances (Box-Test, p = .055) and homogeneity of the error variances (Levene-Test, all p-values > .05), we assumed the ANOVA to be reliable without transformation of the data [9, 85]. To test our hypothesis that AR cues (automatic or manual) would reduce the resumption lag, we conducted two Helmert contrasts. For our three cue groups, the first Helmert contrast compared the effect of any cue (automatic or manual) vs. no cue. The second Helmert contrast compared the effectiveness of automatic cues vs. manual cues.

3.2 Results

As a general manipulation check to see whether the interruptions were disruptive, we compared the resumption lag with the interaction interval (i.e., the time that is used for the same task step without an interruption). A Wilcoxon signed-rank test showed that the time between the end of an interruption and the next click on the correct module (resumption lag M = 4072 ms, SD = 1221 ms) was significantly longer than the time between a click on "OK" and the next correct click on the next module when no interruption occurred (interaction interval M = 1669 ms, SD = 660 ms), (Z = 3483.000, p < .001).

The main effect of cueing was significant, $F(2,80)$ = 4.249, p = .018, *partial* η^2 = .058 (Figure 4 and Table 1). Supporting our hypothesis, the first Helmert contrast of any cue (automatic or manual) vs. no cue indicated a significantly shorter resumption lag when a cue was present, $t(80)$ = 2.893, p = .005. However, contrary to our expectations, the second Helmert contrast of automatic cue vs. manual cue showed no advantage of the manual cue over the automatic cue, $t(80)$ = 0.973, p = .333.

The main effect of interruption length was significant $F(1,80)$ = 16.727, p < .001, *partial* η^2 = .173. As expected, longer interruptions resulted in longer resumption lags. However, there was a significant cue × interruption length interaction,

Table 1. Descriptive results of Experiment 1. Values indicate the M (SD) of the resumption lag in ms and the M (SD) of the resumption errors in percentages.

	15 s Interruption		45 s Interruption	
	Resumption Lag	Resumption Error	Resumption Lag	Resumption Error
No Cue	3976 (996)	3.57 (6.96)	4894 (1356)	7.74 (10.62)
Automatic Cue	3518 (942)	1.19 (4.37)	4173 (1519)	3.57 (6.96)
Manual Cue	3660 (1161)	0.62 (3.21)	3832 (1137)	1.85 (5.34)

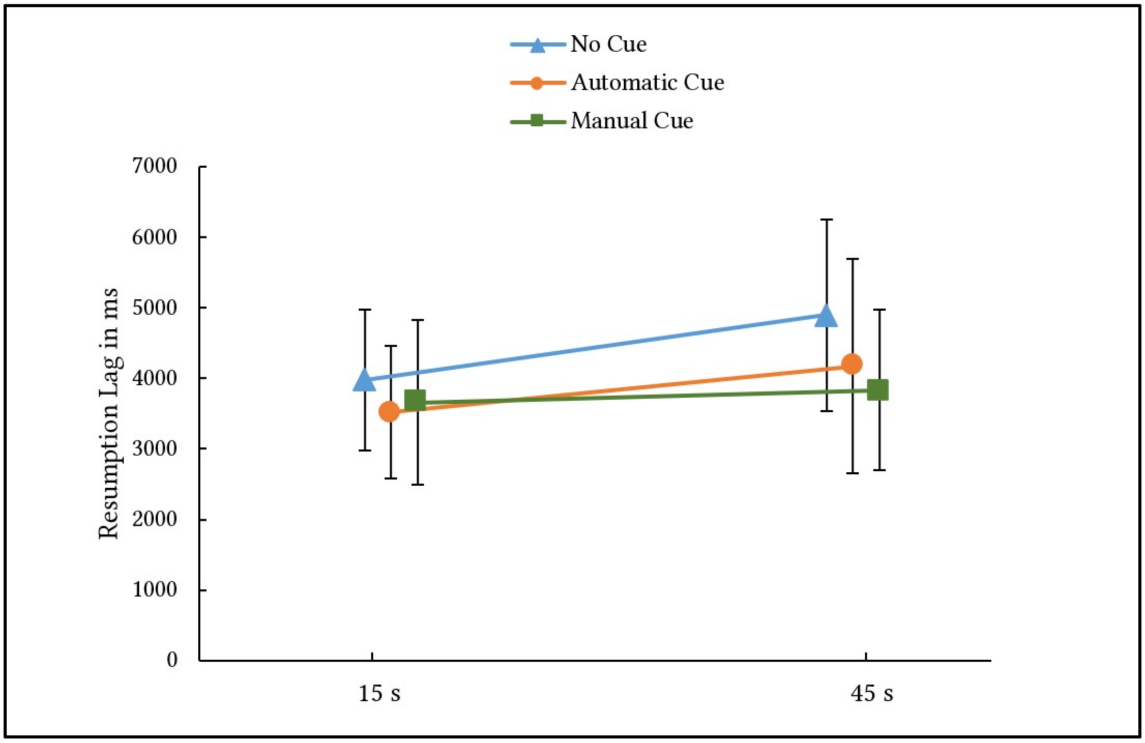


Fig. 4. Results of the mixed ANOVA of Experiment 1. Both manual and automatic cues negate the negative effects of longer interruptions compared to no cue. Error bars represent the standard errors, since the standard error was too small to be displayed clearly.

$F(2,80) = 3.135$, $p = .049$, *partial* $\eta^2 = .073$. Consistent with our hypothesis, a Bonferroni-adjusted post-hoc analysis showed that, when no cue was provided, the resumption lag was significantly longer after the 45s interruption than after the 15s interruption ($p < .001$), whereas there was no difference between the 45s and the 15s interruptions when a cue was provided (automatic cue: $p = 1.000$; manual cue $p = 1.000$).

Resumption errors showed the same result pattern as the resumption lag. The cue had a significant main effect, $F(2,80) = 6.159$, $p = .003$, *partial* $\eta^2 = .072$. The first Helmert contrast indicated that participants exhibited fewer resumption errors when presented with an AR cue compared to no cue, $t(80) = 3.408$, $p < .001$. $d = .536$. The second Helmert contrast indicated no difference in the number of resumption errors between the manual and automatic cue conditions, $t(80) = 0.874$, $p = .385$. Following short interruptions, fewer resumption errors occurred, $F(1,80) = 6.555$, $p = .012$, *partial* $\eta^2 = .034$. There was no significant interaction, $F(2,80) = 0,707$, $p = .496$.

Considering the subjective mental workload, we calculated the NASA TLX raw score by averaging the six sub-scales [37]. A one-way ANOVA showed no difference between the cue groups,
$F(1,80) = 1.008$, $p = .370$, *partial* $\eta^2 = .025$. More specifically, there was no significant difference between the manual and the automatic cue regarding the NASA TLX raw, $t(80) = 0.770$, $p = 1.000$, $d = 0.222$.

In order to explore the temporal dynamics of reorientation towards the primary task when a cue was presented, we compared the first interaction interval following task resumption, the so-called edit lag [71], between the groups. This analysis aimed to determine whether the participants simply clicked where the cue indicated but needed more time to reorient and retrieve task information after making the first click. When a cue was presented ($M = 1673$ ms, $SD = 632$ ms), the following interaction interval was not prolonged compared to when no cue was presented ($M = 1659$ ms, $SD = 725$), $t(80) = 0.105$, $p = .917$.

3.3 Discussion

The objectives of Experiment 1 were to assess the potential of AR cues in facilitating task resumption following interruptions and to examine the influence of interruption length on the effect of AR cues and cue-setting methods. We focus our discussion on the important results for Experiment 2 and provide a broader examination of the results in the general discussion section. The findings only partially support the expected benefits of task resumption facilitated by AR cues. After long interruptions, both the manually and automatically set cues demonstrated superiority over the absence of cues. However, when considering short interruptions, no significant differences were observed between the cue conditions. This lack of differentiation after the 15-second interruptions may be attributed to a floor effect. While we could demonstrate that the interruptions were disruptive, the primary task was still fairly simple, rendering it challenging to detect a facilitation of the task's resumption. Task resumption in the tangible task in Experiment 2 maybe cognitively more demanding and thus might provide more insights on AR cue task resumption facilitation for shorter interruptions.

Importantly, the resumption lag did not significantly increase between 15-second and 45-second interruptions when any type of cue was provided. These findings align with the Memory for Goals theory [2], as the priming constraint suggests that a cue can elevate the activation of the goal, regardless of the duration of the interruption.

In contrast to our expectations, the manual cue provided no additional facilitation of task resumption compared to the automatic cue. The large effect of the presence of any AR cue might lead to a floor effect, precluding any further reduction of the resumption lag through the manual setting. Based on this result, we focused on automatic cues in Experiment 2 only.

4 EXPERIMENT 2: TANGIBLE PRIMARY TASK

In Experiment 2, we used a tangible primary task instead of a computer-based task to further explore the generalizability and applicability of the observed effects in Experiment 1. Few studies investigate the resumption lag with tangible tasks [25, 31, 62] and, to the best of our knowledge, no studies investigate the effects of AR cues on tangible task resumption. We constructed a tangible primary task that resembles the act of sorting medication into a 7-day pill organizer. Medication preparation is a cognitively demanding [98] and error-prone task [46]. Interruptions have been identified as a contributing factor to medication preparation errors [23, 46].

Based on the observation that there were no discernible objective or subjective differences between manually and automatically setting the cue in Experiment 1, we chose to exclusively test the automatic cue in Experiment 2 due to its user convenience. Furthermore, we used the red arrow as cue again. The cue was effective in Experiment 1 and we deemed its salience and spatial memory support [13, 77] even more important for the tangible task because of more background movement, differently colored backgrounds, and a presumably even larger demand on spatial memory due to the larger and three-dimensional working area compared to the computer-based task in Experiment 1. In Experiment 2, both the factor cue (automatic cue vs. no cue) and the factor interruption length (15 s vs. 45 s) were manipulated within-subject. Drawing upon the Memory for Goals theory and the results of Experiment 1, we expected that (1) the presence of an AR cue would result in a shorter resumption lag compared to no cue, and (2) long interruptions would cause longer resumption lags compared to short interruptions. Furthermore, we anticipated that (3) the negative impact of longer interruptions would be mitigated when an AR cue was presented. As Experiment 1, Experiment 2 has been preregistered (https://doi.org/10.17605/OSF.IO/KY9HZ).

4.1 Method

4.1.1 Participants. A total of 38 undergraduate students participated in exchange for course credits or € 10. Participants were recruited via the same system as in Experiment 1 but participants from Experiment 1 were excluded from signing up for Experiment 2. No participant was excluded, and the final data set included 28 female and 10 male participants with an average age of 22.1 years (*SD* = 3.0). All participants had correct or corrected-to-normal vision. This research was approved by the institutional review board and informed written consent was obtained from each participant.

4.1.2 Design. We used a 2 (cue) x 2 (interruption length) within-design. The factor cue included the conditions automatic cue vs. no cue, and the factor interruption length included 15 s interruptions vs. 45 s interruptions. We conducted a prospective power analysis to determine the required number of participants. Based on the slightly below medium effect size of the main effect of the cue in Experiment 1, we estimated the effect size for cue to be f = .200 (power = .80, alpha = .05, two-sided test). The power analysis indicated a sample size of 36 participants. To account for potential data loss, we included a slightly larger participant sample.

4.1.3 Procedure and Material. Experiment 2 adopted the same general procedure of training, execution and post-survey as Experiment 1 with a different primary task. For the primary task, participants engaged in a pill sorting task, for which they were required to transfer colored beads, symbolizing medical pills, from six distinct bowls into a 7-day pill box. The instructions, including which pills were supposed to be placed in each compartment, were presented in AR above the bowls to guide the participants throughout the task. Participants were instructed to individually pick up each pill using their dominant hand and sequentially place it into the pill box in the order provided by the instructions. Upon completing the pills for a given day, participants advanced the instructions to the subsequent day by interacting with an AR-button located alongside the instructions. Following the placement of pills for Sunday, the last day of the week, participants replaced the current pill box with an empty one. The setup and AR view are illustrated in Figures 5 and 6 below.

Each day of the pill sorting task involved the placement of six to eight pills and featured either no interruptions or one or two interruptions, which resulted in a total of seven interruptions for each pill box. An interruption was triggered when a pill was placed in the pill box, indicated by both an acoustic signal and the disappearance of the instructions until the end of the interruption. The interruption task was the same as in Experiment 1. In the cue condition, a cue in the form of a red arrow was presented when the interruption ended, directing participants' attention to the specific

Fig. 5. The setup for Experiment 2 as seen from above. The tablet with the interruption task (left) was placed outside of the field of view during the primary task (right) but was easily reachable by turning around on the office chair.

Fig. 6. The left figure shows the AR shapes that were used to track which pill was picked up and placed into the pill box to measure resumption performance metrics. The right figure shows the primary task from the participants' point of view. The instructions were superimposed by the HoloLens above the bowls with the pills.

bowl containing the next required pill color. The cue disappeared once the participant picked up the pill. We ensured that the color of the pill immediately following the interruption did not occur more than once during that day, so that the cue was unambiguous. Although participants were informed that a cue could occur, no cue was presented during the training week. To counterbalance the cue manipulation, the cue was introduced in the second and third weeks or

Table 2. Descriptive results of Experiment 2. Values indicate the M (SD) of the resumption lag in ms and the M (SD) of the resumption errors in percentages.

	15 s Interruption		45 s Interruption	
	Resumption Lag	Resumption Error	Resumption Lag	Resumption Error
No Cue	4173 (829)	0.34 (0.48)	5054 (924)	0.76 (0.59)
Automatic Cue	2986 (501)	0.13 (0.34)	3037 (523)	0.11 (0.31)

in the fourth and fifth weeks of the experiment. A total of five pill boxes were used, with the initial box serving as a training session.

To measure the interaction interval, the resumption lag and possible resumption errors, we used the HoloLens hand tracking feature. Each of the six bowls and the pill box were placed in AR shapes that were invisible to participants but enabled us to track when a pill was taken out of a bowl and dropped in the pill box. This tracking also enabled us to provide feedback about the pill color that was picked up by showing a small round shape in the respective color at the fingertips of the participant. Furthermore, we could provide feedback if the wrong pill was about to be picked up by providing a short audio sound. The technical details of the tracking system are provided in the supplementary materials.

4.1.4 Analysis. For the resumption lag analysis, only correct resumptions without resumption errors can be analyzed. Resumption errors occurred, when a participant reached into the pill box or any bowl, but the correct next one. A total of 1013 (95.2 %) interruptions were included in the resumption lag analysis. We performed a 2 (cue) × 2 (interruption length) repeated measures ANOVA to test the hypotheses. In all four experimental conditions, the resumption lag values were not normally distributed (Shapiro- Wilk; all p-values < .05). Analogous to the first experiment, we assumed that the ANOVA is robust to the violation of the normal distribution. Therefore, we report the results of the analysis of the untransformed data.

4.2 Results

A paired samples t-test showed that the time between the end of an interruption and the participant reaching into the next correct bowl (resumption lag M = 3804 ms, SD = 550 ms) was significantly longer than the time between sorting a pill and reaching into the next bowl when no interruption occurred (interruption interval, M = 1255 ms, SD = 304 ms), ($t(37) = 27.862$, $p < .001$, $d = 4.520$).

The main effect cue was significant, $F(1,37) = 150.766$, $p < .001$, *partial* $\eta^2 = .803$ (Figure 7 and Table 2). As expected, the resumption lags with no cues were longer than with cues. The main effect of interruption length was significant, $F(1,37) = 76.689$, $p < .001$, *partial* $\eta^2 = .675$. As expected, long interruptions resulted in longer resumption lags than short interruptions. Finally, consistent with our hypothesis, there was a significant cue × interruption length interaction, $F(1,37) = 60.104$, $p < .001$, *partial* $\eta^2 = .619$. A Bonferroni-adjusted post-hoc analysis revealed that, when no cue was provided, the resumption lag was significantly longer after the 45 s interruption than after the 15s interruption ($p < .001$), whereas there was no difference between the 15 s and the 45 s condition when a cue was provided ($p = 1.000$).

As in Experiment 1, we compared the edit lag between the experimental conditions. For this comparison, we considered only the time between placing one pill into the pill box and picking up the next pill for the interaction intervals, excluding the time taken between picking up a pill and placing it, as placing a pill into the pill box does not require the participant to reorient themselves on the primary task. A repeated measures ANOVA showed no difference

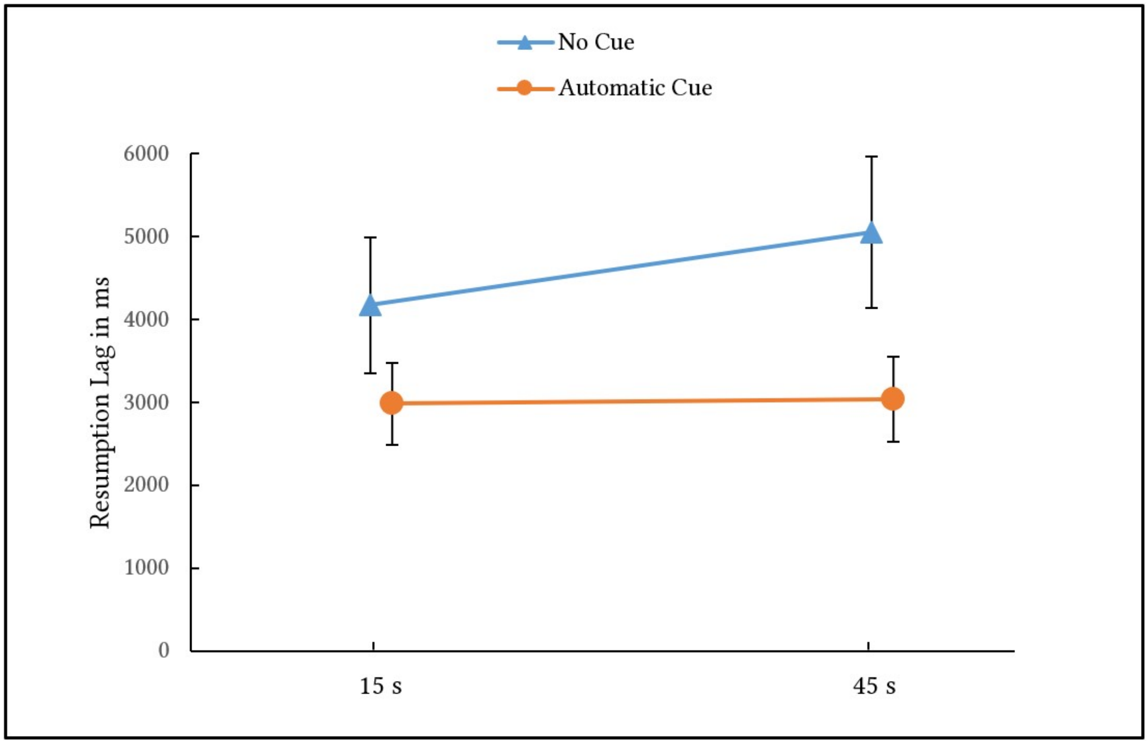


Fig. 7. Comparison of resumption lags between the cue type and interruption length conditions. The error bars represent the standard deviation, since the standard error was too small to be displayed clearly.

in the edit lag between the cueing conditions, $F(1,37) = 0.165$, $p = .687$, the interruption length, $F(1,37) = 3.424$, $p = 0.072$, and no interaction, $F(1,37) = 0.821$, $p = .371$.

The comparison of the resumption errors between the experimental conditions revealed significant main effects and interaction. Participants exhibited fewer resumption errors when presented with an AR cue compared to no cue, $F(1,37) = 34.946$, $p < .001$, *partial* $\eta^2 = .203$. Following short interruptions, fewer resumption errors occurred, $F(1,37) = 5.908$, $p = .020$, *partial* $\eta^2 = .042$. The observed significant interaction, $F(1,37) = 8.963$, $p = .005$, *partial* $\eta^2 = .054$, was further explored using a Bonferroni-adjusted post-hoc analysis. This analysis revealed that in the absence of a cue, resumption errors were higher after a 45 s interruption than a 15 s interruption ($p = .002$), whereas with an AR cue, no such difference was observed ($p = 1.000$).

4.3 Discussion

The primary objectives of Experiment 2 were to assess the potential of AR cues in facilitating task resumption following interruptions and evaluate the mediating effect of interruption length on a tangible primary task. The results support the expected benefits of task resumption facilitated by AR cues. AR cues not only facilitated task resumption after both short and long interruptions, but also negated the adverse effects of increased interruption duration. All comparisons showed a large effect size for the AR cue condition. The resumption errors showed the same pattern of results as the resumption lag. The latter result appears to be due to the fact that almost no errors were made when the cue was presented. Similar to Experiment 1, we were able to demonstrate that the interruptions in this experiment were disruptive. However, the comparison between the resumption lag and the interaction interval warrants careful consideration due to the physical movement required for participants to transition between the interrupting task and the primary task. Due to the relatively small movement involved and given that the resumption lag exceeded the interaction interval by more than threefold, we still deem it plausible to characterize the interruptions in this experiment as disruptive. Additionally, we were able to demonstrate that the reorientation process occurs within the resumption lag, since the interaction directly after the task resumption was not prolonged. Consequently, our chosen primary task proves suitable for further research regarding interruptions and the implementation of visual cues to support task resumption.

5 GENERAL DISCUSSION

Through the two experiments, we successfully demonstrated the facilitative effects of AR cues on task resumption times and errors and these benefits did not translate to compromised performance in subsequent tasks steps (i.e., edit lag). Furthermore, these effects were observed in a computer-based and a tangible task. Finally, we observed that manually set cues by the user were not superior in relation to resumption times and errors than an automatically set cue by the system.

Different to a previous head-mounted display study in which the mere presence of surrounding environmental cues due to the opaque or transparent smart glass design did not support task resumption times [52] or subtle cues on computer screens [17, 18, 90] our red AR arrow cue showed a positive effect on task resumption times. This result is in line with previous research on computer-based tasks that showed that salient non-AR cues that provide spatial information improved task resumption time [13, 90]. Going beyond previous research on cues to support task resumption after interruptions, we showed that the AR cues not only improved task resumption time but also reduced task resumption errors. This is noteworthy because in context of the Memory for Goals research, the focus is on task resumption times and effects on the resumption lag, while effects on resumption errors are seldomly reported [e.g. 13, 17, 18, 52, 90]. In addition, we showed the beneficial effects on resumption times errors as well as the unaffected edit lag in a computer-based task as well as in a tangible task. Previous research on tangible tasks were quantitative observational field studies in the hospital with no experimental control [25, 34] or only investigated the resumption lag [62]. Finally, while Wolff et al. [97] already indicated the potential use of mock-up AR cues in a Wizard of Oz approach, our experimental setup demonstrated the effect of AR cues with real-time augmentations.

One unexpected finding was related to the interruption length effect. Specifically, for the computer-based task, the AR cue only exhibited significant benefits after longer, more disruptive interruptions. In contrast, the AR cue for the tangible primary task exhibited very large effects on both short and long interruptions. The disparity related to interruption length may be attributed to differences in primary task difficulty. In Experiment 1, the modules had distinctive titles and were arranged in two rows on the screen, while in Experiment 2, the pills differed solely in color and were presented in a continuous row. This distinction may have rendered the to-be-resumed task in Experiment 1 easier to remember, because spatial cues are known to enhance task resumption [76]. Consequently, the lower difficulty of task resumption in Experiment 1 could have resulted in a floor effect in the resumption lag following short interruptions, thereby explaining the lack of additional advantage provided by the cue.

Alternatively, the disparity in results between the primary tasks could potentially be attributed to the presence of distinct resumption strategies. In the case of the computer-based task, participants' failure to recall the correct module after the interruption led to guessing, resulting in either a relatively swift task resumption or a resumption error. In contrast, the pill sorting task allowed participants to adopt a counting strategy by tallying the previously sorted pills in the pill box. However, in the no cue group, only four participants reported utilizing this counting approach during the experiment. Interestingly, 17 participants reported that they actively looked at and remembered the next required pill before placing the current one in the pill box.

Furthermore, our study reveals that the rehearsal acquired from manually placing the AR cue did not yield any significant advantages for task resumption. This was surprising, as we expected the increased time for encoding when placing the cue (i.e., interruption lag) to reduce the resumption lag [40] and frustration [44]. The lack of a positive effect might stem from the unfamiliar hand gesture that users needed to perform correctly in order to place the cue. Performing the gesture might have caused excessive extraneous cognitive load, leaving not enough resources for the encoding of the contextual cues [88]. Alternatively, this lack of difference could be explained by soft constraints [29]. According to the Memory for Goals theory, the process of manually setting the cue should raise the activation level of the next goal and thereby facilitate task resumption. If participants perceive following the AR cue as less demanding than retrieving the goal from memory, they will opt for this choice, rendering the initial difference in goal activation irrelevant. A final explanation may stem from the repetitive nature of the hand gesture for placing the AR cue. Wilson and Emmorey [96] demonstrated that repetitive hand movements decreased working memory performance. In our task, this negative impact on working memory might have offset any potential positive effects on additional encoding of the primary task goal. However, users in our study only used the gesture a total of 14 times compared to one more than one hand signs per seconds in Wilson and Emmorey's study [96] which makes this explanation less likely.

Based on our performance and subjective mental workload results, an automatic cue would be preferable, as it seems to offer greater convenience for users with the same performance benefits and workload demands. However, it is crucial to consider the potential challenges linked to implementing an automatic cue system in actual work or everyday environments, such as technical complexities and compatibility issues. These challenges might necessitate the consideration of a manually set cue as a viable alternative in certain practical contexts. Future research could explore the possibility of combining both cue types to optimize task resumption in complex work scenarios. Finally, in both experiments, we found no prolonged time to execute the first action after task resumption (i.e., edit lag [71, 102] with an AR cue compared to no cue. Within the constraints of null hypothesis testing, this result indicates that the participants reoriented and restabilized the task context during the resumption lag rather than passively following the AR cue and simply executing the indicated action. This is importation because if reorientation had merely been delayed, the advantage of a visual cue in facilitating task resumption would have been limited. In terms of the Memory for Goals theory [2], the AR cue appears to have facilitated task resumption. The resumed task step, in turn, aided in cueing the next task step through associated links.

Future research will need to investigate whether simple AR cues also work for more complex tasks such as resuming an assembly task in manufacturing [50]. For more complex tasks, enhancing cues with additional information such as providing multiple next steps instead of only indicating the immediate next step might be necessary. Furthermore, in our case, supplementing the AR cue with pertinent details, such as the name of the medication being administered or the name of the patient, may prove beneficial for safety. Exploring the existing domain-specific task resumption strategies can guide the decision on what information to incorporate.

We now turn to the contribution of our results in relation to the larger HCI research on interruption and interruption coordination [65, 82] and head-mounted display research in general. Our benefitable findings of AR cues on resumption times and errors and specifically the edit lag demonstrate that the cue-based approach is an effective strategy to compromise the detrimental effects of immediate interruptions and might be a viable alternative or addition to mediated or scheduled interruptions. That is, in our study, participants were effective in dealing with the interruption and the AR cue minimized detrimental effects of the interruption. The benefits of AR cues may be particularly pronounced for environments in which users are mobile and have many different primary tasks whereas the benefits might be less pronounced for computer-based office work.

Moreover, AR cues may also affect how users address interruptions with a negotiable starting point [64]. For example, Bogunovich and Salvucci [10] showed that workload and time constraints affect negotiation strategies and participants delayed to attend interruptions during times of higher workload. The availability of AR cues may lead to users changing their strategy regarding the timing of interruptions. Users might become more willing to choose points of high mental load when a cue facilitates task resumption. Future research will need to address the how AR cues affect interruption negotiation strategies.

Furthermore, in our study, we focused on sequential multitasking, which has characterized by fewer task switches and longer periods of attention devoted to, most of the time, two tasks (i.e., primary and interrupting task). Concurrent multitasking includes more task switches and shorter actions on the different tasks. The latter situations are addressed in Salvucci and Taatgen [81] Threaded Cognition theory. The general idea of AR cues may also work in environment with multiple ongoing or paused tasks; however, such environments will pose technical challenges in relation to a tracking system for automatic cues and design challenges in relation to designing AR cues for multiple tasks.

Additionally, AR cues offer a potential alternative for managing interruptions in guided assembly tasks. Rather than relying on a smart interruption system that schedules interruptions at predefined task breakpoints [50], the task guidance system could enhance task resumption through the integration of an AR cue. Future research will need to compare the effectiveness of both approaches.

Beyond this, the use of digital cues may also be a viable solution for interruptions in virtual reality. In VR settings, bystanders often hesitate to interrupt users, perceiving interruptions of a VR user as especially disruptive and the user as unwilling to be disturbed [70]. The knowledge that task resumption is facilitated in VR could potentially lower the inhibition to interrupt VR users, facilitating interactions between VR users and bystanders.

Finally, our study showed no disadvantage of placing an AR cue via a hand gesture in regards to mental load or task resumption. Since this approach requires only gesture recognition instead of permanent task tracking, it is a more practical solution for many work environments. Work environments, that require the user to be mobile, e.g. intensive care units [31], profit from the task-independent nature of the manually set cue. The manual setup also allows the use of less powerful AR glasses that would not be able to perform task tracking.

5.1 Limitation and Strength

Our study has limitations. First, even though we used a tangible and more complex task in Experiment 2 and in most of the resumption lag research [e.g. 12, 92], the task in the present study does not have the complexity inherit to those in actual work environments. However, by applying the Memory for Goals theory [2], our research benefits from a well-established cognitive framework that enhances the credibility of our findings. This theory-driven approach boosts the generalizability of our results, particularly when considering the impact of interruptions in more complex, dynamic environments. Studies in the emergency department and intensive care unit confirmed the effect of interruption length [24, 34] and the general concept of environmental cueing [31]. Moreover, the inclusion of a tangible task in Experiment 2 represents a significant strength of our study. While prior research on resumption lags has primarily focused on computer-based tasks [e.g. 52, 73], the tangible task carried out in Experiment 2 may be more representative of tasks in interruption-rich domains such as healthcare [32] and aviation [63].

Second, the interruptions in both experiments were intentionally structured to allow for the indication of the next correct step through a simple arrow cue. However, actual work tasks may involve subtasks or items that may occur more than once within a task (e.g., two pills of the same color in our pill sorting tasks). These aspects must be considered when transferring our findings to practical work environments by, for example, using differently designed cues to inform of varying colors and shapes or providing additional information that can effectively support those carrying out recurring tasks.

Third, the cues in the experimental setup were 100% correct and therefore fully reliable, which might be impossible to achieve in field settings. Falsely recognized interruptions or a small percentage of incorrectly placed cues could reduce the users' trust in the cues and limit their usefulness [60]. The manually set cues could circumvent this limitation until reliable ways for detecting interruptions and placing cues become available. Indeed, with the limitations of null hypothesis testing, our results show that manual cues had no disadvantages in terms of perceived workload or task resumption performance.

Fourth, interruptions in both experiments were positioned solely at discrete sub-task boundaries such as completing a module (Experiment 1) or dropping a pill in the box (Experiment 2). This may limit the transferability to everyday tasks in which interruptions may occur at any time including more cognitively demanding mid-task interruptions [7, 66]. However, research showed that users frequently avoid such mid-task interruptions in field studies [31, 102] or if the study setting in the laboratory allows such discretionary behavior [7, 93]. Furthermore, in many tangible everyday tasks, users use artifacts that are associated with a task goal. Research showed that such artifacts make task resumption easy [31]. In Experiment 2, for example, being interrupted when holding a pill in the hand and keeping the pill in the hand during the interruption might have acted as such an artifact. Finally, based on the Memory for Goals theory [91], our approach of cueing the location of the to-be-resumed task-step should also work for mid-task interruptions. However, such cues would require a more sophisticated tracking system and some tasks may lack clear sub-tasks boundaries at all [22, 56]. Future research is needed to investigate the effects of AR cues on the resumption of mid-task interruptions and more continuous tasks such as driving.

5.2 Design Implications

This study provides several design recommendations. First, AR cues should be displayed automatically as automatic cues help task resumption as well as manual cues while saving the user time. Furthermore, automatic cues are likely more convenient. Manually set cues might be a viable option when interruption detection or correct placement of the AR cue is not possible. Indeed, we found no drawbacks to manual cues in relation to task resumption performance (i.e., resumption lag, resumption error, subjective workload); manual cues reduce the computational effort while providing increased flexibility.

Second, simple cues such as salient red arrows are not only suitable for finding the to-be-resumed location [13, 77] but also seem to be enough to recall and establish task context. Based on our results, the cue can be used for computer-based tasks and tangible tasks. For more complex tasks, the cue may also be enhanced with additional information. However, it has been argued that additional information or more complex cues might distract and diminish performance on the interrupting task if they are in the field of view of the user [68]. When moving towards more specific tasks contexts and actual users such as nurses sorting medication in elderly homes in our case, cue design may become more specific, and designers can carefully evaluate the benefits of more complex cues. When working with actual users and tasks, we would also recommend using qualitative measures to assess the experience of users and their feedback on the appropriateness of the technology and implementation for the intended use.

Third, the incorporation of highly visible AR cues may serve as a preventive measure against users losing track of the to-be-resumed task entirely. It is possible to design AR cues in a manner that remains perceptible even when the user leaves the room during an interruption. This could be particularly helpful for addressing interruptions that occur in healthcare settings, as they often require staff to change location. Fong and Ratwani [25] showed that such interruptions can result in unfinished tasks. In addition to being advantageous for interruptions that require the user to leave the location of the primary task, these visible AR cues may benefit those who experience nested interruptions (i.e., interrupting tasks that are interrupted again). For unfinished tasks and nested interruptions, the red arrow may need to be replaced or enhanced by icons, that allow users to differentiate between the interrupted tasks.

Fourth, the second experiment demonstrated a new paradigm for task resumption research. The task paradigm enables the tracking of task resumption metrics errors, resumption lag and edit lag reliably. Subsequent experiments with this paradigm have the flexibility to adapt the difficulty levels of both the primary task and the interruption task to accommodate specific research needs. Additionally, experimental designs with multiple secondary tasks or nested interruptions could easily be implemented.

Fifth, due to the simplicity of the AR cue, it could be easily combined with AR task guidance systems. Task guidance via AR facilitates repair work [41, 42]. Since the red arrow is only presented temporarily, it could blend in well with the interface of a task guidance system. Safety-critical environments, such as aviation, where checklists are commonly employed to manage interruptions [20], could potentially benefit from the combination of AR task guidance and AR resumption cues.

6 CONCLUSION

Interruptions are unavoidable in many work domains and in particular immediate interruptions can result in unfinished tasks and contribute to errors [65]. The presented research shows that AR cues can reduce resumption times and resumption errors and reinstate the task context when compared to the absence of cues. AR cues seem to be a viable solution to compromise the negative effects of immediate interruptions, especially in domains where tangible tasks or staff need to change location, such as healthcare [32], aviation [63] or manufacturing [50]. Moreover, AR offers a more streamlined alternative to conventional task resumption methods, like relying on sticky notes [28] carrying artifacts such as a syringe [31] or building task-specific systems [84] or reminders.

Beyond work settings, translating our findings to virtual reality appears promising, as interruptions within virtual reality environments can significantly detract from the immersive experience, similar to the impact of interruptions in real-world settings [27]. Exiting the virtual environment to attend to a real-world interruption may yield a level of disruption similar to that induced by interruptions that require the user to leave the physical workspace. The potential transferability of our insights underscores the relevance of our research in addressing the challenges posed by interruptions across virtual, augmented and actual reality. Finally, as head-mounted AR displays, such as the upcoming Apple Vision Pro, become more ubiquitous, user interactions with interruptions are likely to become increasingly frequent and potentially disruptive. Given its simplicity and versatility, the manual setting of cues offers a viable solution that could be integrated seamlessly into a multitude of everyday tasks, such as cooking, to improve user comfort and reduce the risk of forgetting.

ACKNOWLEDGMENTS

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – (grant number 425868361) and is part of Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments. We also thank Julius Arzberger for his help in conducting the experiments.

REFERENCES

[1] Zahra Shakeri Hossein Abad, Oliver Karras, Kurt Schneider, Ken Barker, and Mike Bauer. 2018. Task Interruption in Software Development Projects: What Makes some Interruptions More Disruptive than Others?. In *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018* (Christchurch, New Zealand) *(EASE '18)*. Association for Computing Machinery, New York, NY, USA, 122–132. https://doi.org/10.1145/3210459.3210471

[2] Erik M. Altmann and J. Gregory Trafton. 2002. Memory for goals: an activation-based model. *Cognitive Science* 26 (2002), 39–83. https://doi.org/10.1207/s15516709cog2601_2

[3] Erik M. Altmann and J. Gregory Trafton. 2004. Task Interruption: Resumption Lag and the Role of Cues. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (2004). https://escholarship.org/uc/item/18b4r661.

[4] Erik M. Altmann and J. Gregory Trafton. 2015. Brief Lags in Interrupted Sequential Performance: Evaluating a Model and Model Evaluation Method. *International Journal of Human-Computer Studies* 79 (2015), 51–65. https://doi.org/10.1016/j.ijhcs.2014.12.007

[5] Jonathan Back, Ann Blandford, and Paul Curzon. 2007. Slip errors and cue salience. In *Proceedings of the 14th European Conference on Cognitive Ergonomics: Invent! Explore!* (London, United Kingdom) *(ECCE '07)*. Association for Computing Machinery, New York, NY, USA, 221–224. https://doi.org/10.1145/1362550.1362595

[6] Anja Baethge and Thomas Rigotti. 2013. Interruptions to workflow: Their relationship with irritation and satisfaction with performance, and the mediating roles of time pressure and mental demands. *Work & Stress* 27 (2013), 43 – 63. https://doi.org/10.1080/02678373.2013.761783

[7] Brian P. Bailey and Shamsi T. Iqbal. 2008. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction* 14 (2008), 21:1–21:28. https://doi.org/10.1145/1314683.1314689

[8] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2001. The MagicBook: a transitional AR interface. *Computers & Graphics* 25 (2001), 745–753. https://doi.org/10.1016/S0097-8493%2801%2900117-0

[9] María J. Blanca, Jaume Arnau, F. Javier García-Castro, Rafael Alarcón, and Roser Bono. 2023. Non-normal Data in Repeated Measures ANOVA: Impact on Type I Error and Power. *Psicothema* 35 1 (2023), 21–29. https://doi.org/10.7334/psicothema2022.292

[10] Peter Bogunovich and Dario Salvucci. 2011. The effects of time constraints on user behavior for deferrable interruptions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (, Vancouver, BC, Canada,) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 3123–3126. https://doi.org/10.1145/1978942.1979404

[11] Juliana J. Brixey, David J. Robinson, Craig W. Johnson, Todd R. Johnson, James P. Turley, and Jiajie Zhang. 2007. A Concept Analysis of the Phenomenon Interruption. *Advances in Nursing Science* 30 (2007), 26–42. https://doi.org/10.1097/00012272-200701000-00012

[12] Duncan P. Brumby, Christian P. Janssen, and Gloria Mark. 2019. *How Do Interruptions Affect Productivity?* Apress, Berkeley, CA, 85–107. https://doi.org/10.1007/978-1-4842-4221-6_9

[13] James E. Cane, Fabrice Cauchard, and Ulrich W. Weger. 2012. The time-course of recovery from interruption during reading: Eye movement evidence for the role of interruption lag and spatial memory. *The Quarterly Journal of Experimental Psychology* 65, 7 (2012), 1397–1413. https://doi.org/10.1080/17470218.2012.656666

[14] Kuan-Wen Chen, Yung-Ju Chang, and Liwei Chan. 2022. Predicting Opportune Moments to Deliver Notifications in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (, New Orleans, LA, USA,) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 186, 18 pages. https://doi.org/10.1145/3491102.3517529

[15] Carey D. Chisholm, Agnieszka Dornfeld, David R. Nelson, and William H. Cordell. 2001. Work interrupted: a comparison of workplace interruptions in emergency departments and primary care offices. *Annals of Emergency Medicine* 38 2 (2001), 146–51. https://doi.org/10.1067/mem.2001.115440

[16] Lacey Colligan, Stephanie A. Guerlain, Susan E. Steck, and Tracey R. Hoke. 2012. Designing for distractions: a human factors approach to decreasing interruptions at a centralised medication station. *BMJ Quality & Safety* 21 (2012), 939 – 947. https://doi.org/10.1136/bmjqs-2011-000289

[17] Edward Cutrell, Mary Czerwinski, and Eric Horvitz. 2001. Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. In *IFIP TC13 International Conference on Human-Computer Interaction* (2001). https://doi.org/10.1016/j.apergo.2014.02.005

[18] Mary Czerwinski, Edward Cutrell, and Eric Horvitz. 2000. Instant Messaging: Effects of Relevance and Timing. In *People and Computers XIV: Proceedings of HCI. Vol. 2. British Computer Society, 2000* (2000). 71–76.

[19] Mary Czerwinski, Eric Horvitz, and Susan Wilhite. 2004. A diary study of task switching and interruptions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. Association for Computing Machinery, New York, NY, USA, 175–182. https://doi.org/10.1145/985692.985715

[20] Key Dismukes and Jessica Lang Nowinski. 2006. *Prospective Memory, Concurrent Task Management, and Pilot Error.* Oxford University Press, New York, 225. https://doi.org/10.1093/acprof:oso/9780195305722.003.0016

[21] Andreas Dünser and Eva Hornecker. 2007. Lessons from an AR book study. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction* (Baton Rouge, Louisiana) *(TEI '07)*. Association for Computing Machinery, New York, NY, USA, 179–182. https://doi.org/10.1145/1226969.1227006

[22] Justin Edwards, Christian Janssen, Sandy Gould, and Benjamin R. Cowan. 2021. Eliciting Spoken Interruptions to Inform Proactive Speech Agent Design. In *Proceedings of the 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 23, 12 pages. https://doi.org/10.1145/3469595.3469618

[23] Elizabeth Allan Flynn, Kenneth N. Barker, J. Tyron Gibson, Robert E. Pearson, Bruce A. Berger, and Lesley Anne Smith. 1999. Impact of Interruptions and Distractions on Dispensing Errors in an Ambulatory Care Pharmacy. *American Journal of Health-System Pharmacy : AJHP : Official Journal of the American Society of Health-System Pharmacists* 56 13 (1999), 1319–25. https://doi.org/10.1093/ajhp/56.13.1319

[24] Allan Fong, A. Zachary Hettinger, and Raj M. Ratwani. 2017. A Predictive Model of Emergency Physician Task Resumption Following Interruptions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2405–2410. https://doi.org/10.1145/3025453.3025700

[25] Allan Fong and Raj M. Ratwani. 2018. Understanding Emergency Medicine Physicians Multitasking Behaviors Around Interruptions. *Academic Emergency Medicine* 25 (2018), 1164–1168. https://doi.org/10.1111/acem.13496

[26] Ceenu George, Philipp Janssen, David Heuss, and Florian Alt. 2019. Should I Interrupt or Not? Understanding Interruptions in Head-Mounted Display Settings. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 497–510. https://doi.org/10.1145/3322276.3322363

[27] Sarthak Ghosh, Lauren Winston, Nishant Panchal, Philippe Kimura-Thollander, Jeff Hotnog, Douglas Cheong, Gabriel Reyes, and Gregory D. Abowd. 2018. NotifiVR: Exploring Interruptions and Notifications in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 24 (2018), 1447–1456. https://doi.org/10.1109/TVCG.2018.2793698

[28] Victor M. González and Gloria Mark. 2004. "Constant, constant, multi-tasking craziness": managing multiple working spheres. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. Association for Computing Machinery, New York, NY, USA, 113–120. https://doi.org/10.1145/985692.985707

[29] Wayne D. Gray and Wai-Tat Fu. 2004. Soft constraints in interactive behavior: the case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science* 28 (2004), 359–382. https://doi.org/10.1016/j.cogsci.2003.12.001

[30] Tobias Grundgeiger, Alea Münz, Paul Schlosser, and Oliver Happel. 2023. Supervising Multiple Operating Rooms Using a Head-Worn display: A Longitudinal Evaluation of the Experience of Supervising Anesthesiologists and Their Co-Workers. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 345, 18 pages. https://doi.org/10.1145/3544548.3581180

[31] Tobias Grundgeiger, Penelope Sanderson, Hamish G. MacDougall, and Balasubramanian Venkatesh. 2010. Interruption management in the intensive care unit: Predicting resumption times and assessing distributed support. *Journal of Experimental Psychology: Applied* 16, 4 (2010), 317–334. https://doi.org/10.1037/a0021912

[32] Tobias Grundgeiger and Penelope M. Sanderson. 2009. Interruptions in healthcare: Theoretical views. *International Journal of Medical Informatics* 78 5 (2009), 293–307. https://doi.org/10.1016/j.ijmedinf.2008.10.001

[33] Tobias Grundgeiger, Penelope M. Sanderson, C. Beltran Orihuela, Andrea Thompson, Hamish G. MacDougall, Leo Nunnink, and Balasubramanian Venkatesh. 2013. Prospective memory in the ICU: the effect of visual cues on task execution in a representative simulation. *Ergonomics* 56 (2013), 579 – 589. https://doi.org/10.1080/00140139.2013.765604

[34] Tobias Grundgeiger, Penelope M. Sanderson, Cristina Beltran Orihuela, Abbie Kate Thompson, Hamish G. MacDougall, Leo Nunnink, and Balasubramanian Venkatesh. 2010. Distractions and Interruptions in the Intensive Care Unit: A Field Observation and a Simulator Experiment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 54 (2010), 835 – 839. https://doi.org/10.1177/1541931210054012

[35] Anhong Guo, Shashank Raghu, Xuwen Xie, Saad Ismail, Xiaohui Luo, Joseph Simoneau, Scott Gilliland, Hannes Baumann, Caleb Southern, and Thad Starner. 2014. A comparison of order picking assisted by head-up display (HUD), cart-mounted display (CMD), light, and paper pick list. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers* (Seattle, Washington) *(ISWC '14)*. Association for Computing Machinery, New York, NY, USA, 71–78. https://doi.org/10.1145/2634317.2634321

[36] Pavel Gurevich, Joel Lanir, Benjamin Cohen, and Ran Stone. 2012. TeleAdvisor: a versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 619–622. https://doi.org/10.1145/2207676.2207763

[37] S. G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50 (2006), 904 – 908. https://doi.org/10.1177/154193120605000909

[38] S. G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology* 52 (1988), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[39] Andrew N. Healey, Christopher P Primus, and Maria Koutantji. 2007. Quantifying distraction and interruption in urological surgery. *Quality and Safety in Health Care* 16 (2007), 135 – 139. https://doi.org/10.1136/qshc.2006.019711

[40] Helen M. Hodgetts and Dylan M. Jones. 2006. Contextual cues aid recovery from interruption: the role of associative activation. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 32 5 (2006), 1120–32. https://doi.org/10.1037/0278-7393.32.5.1120

[41] Clemens Hoffmann, Sebastian Büttner, Michael Prilla, and Kai Wundram. 2020. Impact of augmented reality guidance for car repairs on novice users of AR: a field experiment on familiar and unfamiliar tasks. In *Proceedings of Mensch Und Computer 2020* (Magdeburg, Germany) *(MuC '20)*. Association for Computing Machinery, New York, NY, USA, 279–289. https://doi.org/10.1145/3404983.3405594

[42] Clemens Hoffmann, Michael Prilla, Kai Wundram, and Björn Emmermann. 2021. Augmented Reality Guidance for Car Repairs: Immediate and Long-Term Effects: Augmented Reality Guidance for Car Repairs. In *Proceedings of Mensch Und Computer 2021* (Ingolstadt, Germany) *(MuC '21)*. Association for Computing Machinery, New York, NY, USA, 110–119. https://doi.org/10.1145/3473856.3473892

[43] Shamsi T. Iqbal and Brian P. Bailey. 2005. Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1489–1492. https://doi.org/10.1145/1056808.1056948

[44] Shamsi T. Iqbal and Brian P. Bailey. 2008. Effects of intelligent notification management on users and their tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 93–102. https://doi.org/10.1145/1357054.1357070

[45] Shamsi T. Iqbal and Eric Horvitz. 2007. Disruption and recovery of computing tasks: field study, analysis, and directions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 677–686. https://doi.org/10.1145/1240624.1240730

[46] K. Lynette James, David E. Barlow, Rowena McArtney, Sarah J. Hiom, Dave Roberts, and Cate Whittlesea. 2009. Incidence, type and causes of dispensing errors: A review of the literature. *International Journal of Pharmacy Practice* 17 (2009), 9–30. https://doi.org/10.1211/ijpp/17.1.0004

[47] Stuart A. Jones, Sandy J. J. Gould, and Anna Louise Cox. 2012. Snookered by an interruption? Use a cue. In *British Computer Society Conference on Human-Computer Interaction* (2012). https://doi.org/10.14236/ewic/HCI2012.33

[48] Dina Kanaan and Nadine Marie Moacdieh. 2021. Eye Tracking to Evaluate the Effects of Interruptions and Workload in a Complex Task. *Human Factors: The Journal of Human Factors and Ergonomics Society* 64 (2021), 1168 – 1180. https://doi.org/10.1177/0018720821990487

[49] Simon Kimmel, Vanessa Cobus, and Wilko Heuten. 2021. opticARe - Augmented Reality Mobile Patient Monitoring in Intensive Care Units. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (Osaka, Japan) *(VRST '21)*. Association for Computing Machinery, New York, NY, USA, Article 11, 11 pages. https://doi.org/10.1145/3489849.3489852

[50] Ari Kolbeinsson, Peter Thorvald, and Jessica Lindblom. 2017. Coordinating the interruption of assembly workers in manufacturing. *Applied Ergonomics* 58 (2017), 361–371. https://doi.org/10.1016/j.apergo.2016.07.015

[51] Stas Krupenia and Penelope M Sanderson. 2006. Does a head-mounted display worsen inattentional blindness?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (2006), Vol. 50. SAGE Publications Sage CA: Los Angeles, CA, 1638–1642.

[52] Mari Kruse, Vera Volk, Nathalie Merdan, Penelope Sanderson, and Tobias Grundgeiger. 2020. Comparison between head-mounted displays regarding the resumption of a disrupted work task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (2020), Vol. 64. Human Factors and Ergonomics Society, Chicago, 364–365. https://doi.org/10.1177/1071181320641082

[53] Jeremias Kuge, Tobias Grundgeiger, Paul Schlosser, Penelope Sanderson, and Oliver Happel. 2021. Design and Evaluation of a Head-Worn Display Application for Multi-Patient Monitoring. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 879–890. https://doi.org/10.1145/3461778.3462011

[54] Katherine Labonté and François Vachon. 2021. Resuming a Dynamic Task Following Increasingly Long Interruptions: The Role of Working Memory and Reconstruction. *Frontiers in Psychology* 12 (2021). https://doi.org/10.3389/fpsyg.2021.659451

[55] Eva Lampen, Jonas Teuber, Felix Gaisbauer, Thomas Bär, Thies Pfeiffer, and Sven Wachsmuth. 2019. Combining Simulation and Augmented Reality Methods for Enhanced Worker Assistance in Manual Assembly. *Procedia CIRP* 81 (2019), 588–593. https://doi.org/10.1016/j.procir.2019.03.160

[56] Jakob Landesberger, Ute Ehrlich, and Wolfgang Minker. 2020. Do the Urgent Things first! - Detecting Urgency in Spoken Utterances based on Acoustic Features. In *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) *(UMAP '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 53–58. https://doi.org/10.1145/3386392.3397598

[57] Kara A. Latorella. 1998. Effects of Modality on Interrupted Flight Deck Performance: Implications for Data Link. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 42 (1998), 87 – 91. https://doi.org/10.1177/154193129804200120

[58] Byung Lan Lee, Kwanghun Chung, and Sung-Hee Kim. 2018. Interruption Cost Evaluation by Cognitive Workload and Task Performance in Interruption Coordination Modes for Human–Computer Interaction Tasks. *Applied Sciences* 8, 10 (2018). https://doi.org/10.3390/app8101780

[59] Gun A. Lee, Andreas Dünser, Seungwon Kim, and Mark Billinghurst. 2012. CityViewAR: A mobile outdoor AR application for city visualization. *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)* (2012), 57–64. https://doi.org/10.1109/ISMAR-AMH.2012.6483989

[60] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of Human Factors and Ergonomics Society* 46 (2004), 50 – 80. https://doi.org/10.1518/hfes.46.1.50_30392

[61] Simon Yau Wai Li, Anna Louise Cox, Ann Blandford, Paul A. Cairns, Richard M. Young, and Aliza Abeles. 2006. Further investigations into post-completion error : the effects of interruption position and duration. In *Proceedings of the Annual Meeting of the Cognitive Science Society, 28* (2006). https://escholarship.org/uc/item/9zk2187c.

[62] Natalie R. Lodinger and Patricia R. Delucia. 2018. Angle of Camera View Influences Resumption Lag in a Visual-Motor Task. *Human Factors: The Journal of Human Factors and Ergonomics Society* 61 (2018), 793 – 804. https://doi.org/10.1177/0018720818794526

[63] Loukia D. Loukopoulos and Immanuel Barshi. 2001. Cockpit Interruptions and Distractions: A Line Observation Study. In *Proceedings of the 11th International Symposium on Aviation Psychology*.

[64] Daniel C. McFarlane. 2002. Comparison of Four Primary Methods for Coordinating the Interruption of People in Human-Computer Interaction. *Human–Computer Interaction* 17, 1 (2002), 63 – 139. https://doi.org/10.1207/S15327051HCI1701_2

[65] Daniel C. McFarlane and Kara A. Latorella. 2002. The Scope and Importance of Human Interruption in Human-Computer Interaction Design. *Human–Computer Interaction* 17 (2002), 1 – 61. https://doi.org/10.1207/S15327051HCI1701_1

[66] Christopher A. Monk, Deborah A. Boehm-Davis, George Mason, and J. Gregory Trafton. 2004. Recovering From Interruptions: Implications for Driver Distraction Research. *Human Factors: The Journal of Human Factors and Ergonomics Society* 46 (2004), 650 – 663. https://doi.org/10.1518/hfes.46.4.650.56816

[67] Christopher A. Monk, J Gregory Trafton, and Deborah A Boehm-Davis. 2008. The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology: Applied* 14, 4 (2008), 299–313. https://www.ncbi.nlm.nih.gov/pubmed/19102614 Monk, Christopher A Trafton, J Gregory Boehm-Davis, Deborah A eng Research Support, U.S. Gov't, Non-P.H.S. 2008/12/24 J Exp Psychol Appl. 2008 Dec;14(4):299-313. doi: 10.1037/a0014402..

[68] Y. Ian Noy. 1989. Intelligent route guidance: will the new horse be as good as the old? *Conference Record of papers presented at the First Vehicle Navigation and Information Systems Conference (VNIS '89)* (1989), 49–55. https://doi.org/10.1109/VNIS.1989.98739

[69] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven K. Feiner, and Barbara Tversky. 2015. Virtual Replicas for Remote Assistance in Virtual and Augmented Reality. *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (2015). https://doi.org/10.1145/2807442.2807497

[70] Joseph O'Hagan, Julie R. Williamson, and Mohamed Khamis. 2020. Bystander interruption of VR users. In *Proceedings of the 9TH ACM International Symposium on Pervasive Displays* (Manchester, United Kingdom) *(PerDis '20)*. Association for Computing Machinery, New York, NY, USA, 19–27. https://doi.org/10.1145/3393712.3395339

[71] Chris Parnin and Spencer Rugaber. 2011. Resumption strategies for interrupted programming tasks. *Software Quality Journal* 19 (2011), 5–34. https://doi.org/10.1007/s11219-010-9104-9

[72] George Plakas, Stavros T. Ponis, K. Agalianos, Eleni Aretoulaki, and Sotiris P. Gayialis. 2020. Augmented Reality in Manufacturing and Logistics: Lessons Learnt from a Real-Life Industrial Application. *Procedia Manufacturing* 51 (2020), 1629–1635. https://doi.org/10.1016/j.promfg.2020.10.227

[73] Raj Ratwani and J Gregory Trafton. 2010. An eye movement analysis of the effect of interruption modality on primary task resumption. *Human Factors* 52, 3 (2010), 370–380. https://doi.org/10.1177/0018720810374195

[74] Raj M. Ratwani, Alyssa E. Andrews, Malcolm McCurry, J. Gregory Trafton, and Matthew S. Peterson. 2007. Using Peripheral Processing and Spatial Memory to Facilitate Task Resumption. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 51 (2007), 244 – 248. https://doi.org/10.1177/154193120705100421

[75] Raj M. Ratwani, Allan Fong, Josh S. Puthumana, and Aaron Zachary Hettinger. 2017. Emergency Physician Use of Cognitive Strategies to Manage Interruptions. *Annals of Emergency Medicine* 70 (2017), 683–687. https://doi.org/10.1016/j.annemergmed.2017.04.036

[76] Raj M. Ratwani, J. Malcolm McCurry, and J. Gregory Trafton. 2008. Predicting postcompletion errors using eye movements. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 539–542. https://doi.org/10.1145/1357054.1357141

[77] Raj M. Ratwani and J. Gregory Trafton. 2008. Spatial memory guides task resumption. *Visual Cognition* 16 (2008), 1001 – 1010. https://doi.org/10.1080/13506280802025791

[78] Daniel Roth, Kevin Yu, Frieder Pankratz, Gleb Gorbachev, Andreas Keller, Marc Lazarovici, Dirk Wilhelm, Simon Weidert, Nassir Navab, and Ulrich Eck. 2021. Real-time Mixed Reality Teleconsultation for Intensive Care Units in Pandemic Situations. *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (2021), 693–694. https://doi.org/10.1109/VRW52623.2021.00229

[79] Alia Rukubayihunga, Jean-Yves Didier, and Samir Otmane. 2016. Towards assembly steps recognition in augmented reality. In *Proceedings of the 2016 Virtual Reality International Conference* (Laval, France) *(VRIC '16)*. Association for Computing Machinery, New York, NY, USA, Article 17, 5 pages. https://doi.org/10.1145/2927929.2927953

[80] Dario D. Salvucci and Peter Bogunovich. 2010. Multitasking and monotasking: the effects of mental workload on deferred task interruptions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 85–88. https://doi.org/10.1145/1753326.1753340

[81] Dario D. Salvucci and Niels Anne Taatgen. 2008. Threaded cognition: an integrated theory of concurrent multitasking. *Psychological Review* 115 1 (2008), 101–30. https://doi.org/10.1037/0033-295X.115.1.101

[82] Dario D. Salvucci, Niels A. Taatgen, and Jelmer P. Borst. 2009. Toward a unified theory of the multitasking continuum: from concurrent performance to task switching, interruption, and resumption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1819–1828. https://doi.org/10.1145/1518701.1518981

[83] Chiara Santomauro, Madeleine Powell, Chelsea Davis, David Liu, Leanne M. Aitken, and Penelope M. Sanderson. 2018. Interruptions to Intensive Care Nurses and Clinical Errors and Procedural Failures: A Controlled Study of Causal Connection. *Journal of Patient Safety* 17 (2018), e1433 – e1440. https://doi.org/10.1097/PTS.0000000000000528

[84] Farzan Sasangohar, Stacey D. Scott, and Mary L. Cummings. 2014. Supervisory-level interruption recovery in time-critical control tasks. *Applied Ergonomics* 45 4 (2014), 1148–56. https://doi.org/https://doi.org/10.1016/j.apergo.2014.02.005

[85] Holger Schielzeth, Niels J. Dingemanse, Shinichi Nakagawa, David F. Westneat, Hassen Allegue, Céline Teplitsky, Denis Réale, Ned A. Dochtermann, László Zsolt Garamszegi, and Yimen Gerardo Araya-Ajoy. 2020. Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution* 11 (2020), 1141 – 1152. https://doi.org/10.1111/2041-210X.13434

[86] Paul Schlosser, Ben Matthews, and Penelope M. Sanderson. 2021. Head-worn displays for healthcare and industry workers: A review of applications and design. *International Journal of Human-Computer Studies* 154 (2021), 102628. https://doi.org/10.1016/j.ijhcs.2021.102628

[87] Ashutosh Shivakumar, Aishwarya Bositty, Nia Peters, and Yong Pei. 2020. Real-Time Interruption Management System for Efficient Distributed Collaboration in Multi-tasking Environments. *Proceedings of the ACM on Human-Computer Interaction* 4 (2020), 1 – 23. https://doi.org/10.1145/3392844

[88] Alexander Skulmowski and Günter Daniel Rey. 2020. Subjective cognitive load surveys lead to divergent results for interactive learning media. *Human Behavior and Emerging Technologies* 2, 2 (2020), 149–157. https://doi.org/10.1002/hbe2.184

[89] Cheri Speier, Joseph S. Valacich, and Iris Vessey. 1999. The Influence of Task Interruption on Individual Decision Making: An Information Overload Perspective. *Decision Sciences* 30 (1999), 337–360. https://doi.org/10.1111/j.1540-5915.1999.tb01613.x

[90] J. Gregory Trafton, Erik M. Altmann, and Derek P. Brock. 2005. Huh, what was I Doing? How People Use Environmental Cues after an Interruption. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 49 (2005), 468 – 472. https://doi.org/10.1177/154193120504900354

[91] J. Gregory Trafton, Erik M. Altmann, Derek P. Brock, and Farilee Mintz. 2003. Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies* 58 (2003), 583–603. https://doi.org/10.1016/S1071-5819(03)00023-5

[92] J. Gregory Trafton and Christopher A. Monk. 2007. Task Interruptions. *Reviews of Human Factors and Ergonomics* 3 (2007), 111–126. https://doi.org/10.1518/155723408X299852

[93] Michael Weng, Stephan Huber, Elizabeth Vilgan, Tobias Grundgeiger, and Penelope M. Sanderson. 2017. Interruptions, visual cues, and the microstructure of interaction: Four laboratory studies. *International Journal of Human-Computer Studies* 103 (2017), 77–94. https://doi.org/10.1016/j.ijhcs.2017.02.002

[94] Johanna I. Westbrook, Magdalena Z. Raban, Scott R. Walter, and Heather E. Douglas. 2018. Task errors by emergency physicians are associated with interruptions, multitasking, fatigue and working memory capacity: a prospective, direct observation study. *BMJ Quality & Safety* 27 (2018), 655 – 663. https://doi.org/10.1136/bmjqs-2017-007333

[95] Robert A. Whelan. 2008. Effective Analysis of Reaction Time Data. *The Psychological Record* 58 (2008), 475–482. https://doi.org/10.1007/BF03395630

[96] Margaret Wilson and Karen Emmorey. 1997. A visuospatial "phonological loop" in working memory: Evidence from American Sign Language. *Memory & Cognition* 25 (1997), 313–320. https://doi.org/10.3758/BF03211287

[97] Marion Wolff, Stéphane Lefort, Patxi Bérard, and Régis Mollard. 2022. Task Interruption Management in Aeronautics: Contribution of an Augmented Reality Device: Task interruption in Aeronautics and Augmented Reality. In *Proceedings of the 17th "Ergonomie et Informatique Avancée" Conference* (Bidart, France) *(ErgoIA '21)*. Association for Computing Machinery, New York, NY, USA, Article 3, 11 pages. https://doi.org/10.1145/3486812.3486833

[98] Andrea L. Zartman, Robin C. Hilsabeck, Charles A Guarnaccia, and Andrew W. Houtz. 2013. The Pillbox Test: an ecological measure of executive functioning and estimate of medication management abilities. *Archives of Clinical Neuropsychology : the Official Journal of the National Academy of Neuropsychologists* 28 4 (2013), 307–19. https://doi.org/10.1093/arclin/act014

[99] Xianjun Sam Zheng, Cedric Foucault, Patrik Matos da Silva, Siddharth Dasari, Tao Yang, and Stuart Goose. 2015. Eye-Wearable Technology for Machine Maintenance: Effects of Display Position and Hands-free Operation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2125–2134. https://doi.org/10.1145/2702123.2702305

[100] Omid Ziaee and Mohsen Hamedi. 2021. Augmented reality applications in manufacturing and its future scope in Industry 4.0. *ArXiv* abs/2112.11190 (2021). https://doi.org/10.48550/arXiv.2112.11190

[101] Tanja Zigart and Sebastian Schlund. 2020. *Evaluation of Augmented Reality Technologies in Manufacturing - A Literature Review*. Vol. 1207. Springer, Cham, 75–82. https://doi.org/10.1007/978-3-030-51369-6_11

[102] Manuela Züger and Thomas Fritz. 2015. Interruptibility of Software Developers and its Prediction Using Psycho-Physiological Sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2981–2990. https://doi.org/10.1145/2702123.2702593

[103] Manuela Züger, Sebastian C. Müller, André N. Meyer, and Thomas Fritz. 2018. Sensing Interruptibility in the Office: A Field Study on the Use of Biometric and Computer Interaction Sensors. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC,Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174165

Received 14 September 2023; revised 12 December 2023; accepted 19 January 2024

Chapter 9

Discussion

Our research revealed many conditional factors like :“**If** the user is stationary. . . ”, “**If** hands are free. . . ,” and so on—which makes it very difficult to reach a consensus on AR notification design. As Grubert et al. [108] stated: “The concept of pervasive augmented reality requires that the AR system can adapt to the current, changing requirements and constraints of the user’s context and thus allows for continuous usage.” While VR applications are generally self-contained, AR systems are designed to work within a more general, real-world context. Consider the supermarket example for notifications [61]: When crafting a VR environment, you have the option of deliberately choosing where notifications will be displayed and how the user should interact with them. The system designer knows what tasks the user might be performing and understands how the environment is structured, which allows for a highly context-optimized design. Of course, when designing an AR system for a very specific use case—such as a system to aid in surgery—the design can be optimized for that environment. However, while VR is restricted to your current tracking space, AR allows you to leave your environment and do something else without taking off the headset. This context switching is critical for AR applications, as these headsets must be able to understand and adapt to the user’s current context. What may have been an ideal system for a scenario in which the user is stationary, and their hands are preoccupied with another task may become highly distracting or non-functional when the user moves into a different environment or engages in a task that allows them to use their hands freely. Our experiment in the wild shows this clearly. The benefits of the AR notifications are heavily influenced by context; the situation the user is currently in, but also the relevance of the notification content to the user. However, not just the benefits are amplified by context, but the downsides as well. Participants felt more interrupted by the notifications in AR, which was sometimes seen as desirable if the notification information was important to them, and other times as undesirable, if the notification was irrelevant. In the case of the smartglasses used, notification information was always displayed in the center of the FOV, which results in the notification overlaying real-world tasks. Using a world-stabilized placement adjacent to a task but not overlaid, like we found in our first experiment (section 5.1), could be a beneficial way to still inform the user without obstructing their view.

Based on our findings, we have compiled a list of design suggestions that should be considered when creating notifications for AR.

9.1 Design Implications

Presentation

Notifications should be generally placed in the bottom center of the user's FOV. This position turned out to be hard to miss, ensuring timely notification delivery, while not being too obtrusive and distracting. Notifications should be displayed like this, if the user's current task is unknown, or when the user is moving around a lot. When transitioning to a stationary task, like for example maintenance of a machine or sitting at a desk, notifications should be presented peripheral to the main focus of the user. Taking the example of working at a desk with a monitor, notifications could be displayed on the edges of the monitor, calling attention to the current task, but not obstructing it. If the current task is fully in AR, and requires heavy use of the hands, displaying notifications on the wrist is also a viable option, however comes at the risk of missed notifications if the hands are not in view. If objects in the environment produce notifications, displaying notifications over the objects also yields good results.

To not clutter the environment, notifications should be displayed for a few seconds and if they are not immediately attended, they should move into a list with other notifications that have not been attended. This list should be easy and quick to access, by for example summoning it by looking at one's palm and attaching it to the hand. Notifications should also be easy to quickly dismiss if they are interruptive at the moment, to be accessed at a later time from this list. To strengthen notification noticeability, a combination of visual and audio should be used, as visual notifications alone are missed at a much higher rate than the combination of visual and audio. Lastly, notifications should be classified according to their importance to their user. Not all notifications a person would receive on their phone should be displayed in AR, or at least in the FOV. Users should be able to choose between different ways of presentation, on a granular level, as not even notifications from a single source have equal importance.

Interaction

For attending notifications, like dismissing them, Touch interaction should be used. Compared to interaction techniques using only a single modality like Eye-Gaze and Dwell, Touch was shown to outperform them, as well as being preferred by the users. Compared to interaction techniques with multiple modalities, Touch still performed at least on par or better than other interaction techniques, so the use of Touch is still encouraged. However, since it did not significantly outperform other techniques, the use of multimodal interaction techniques could be viable in settings where the hands are preoccupied. It is then highly important to prioritize the speed and reliability of the techniques, as most remarks about the interaction techniques besides Touch were focused on perceived reliability. Looking at Gaze + Pinch, Pfeuffer et al. [59] mention this in their work as well, calling it “Early and Late Triggers”. They suggest including generous and predictive timing when using Gaze + Pinch, to ensure that indication and confirmation processing can align. We see from our work that this is not just an issue with Gaze + Pinch, but also with other combinations of modalities. This means that multimodal techniques need to offer some sort of fail-safe that gives the user a bit of room for error.

In any case, interaction should not be limited to a single technique. As described, Touch might not be ideal in scenarios where the hands are preoccupied. Giving the user several ways to interact introduces flexibility to the system and ensures seamless interaction. The techniques should also be configurable, since not everyone prefers the same techniques, or some techniques might not be appropriate depending on the current usage scenario, such as voice commands in a conversational setting. Great attention has to then be put into the specificity of gestures or voice commands, as too low precision could cause frequent accidental triggers, whereas too high precision could cause many false negatives and could affect perceived reliability.

9.2 Limitations

With any research, it is critical to not overemphasize the findings and to acknowledge that there are always limitations in every scientific experiment. While we have designed and conducted these experiments with utmost care, there are still limitations to our work that we will point out. One limitation of this dissertation is the task used for the majority of papers presented. While the virtual cooking environment used was a flexible task, requiring users to be stationary as well as mobile, it cannot be used as a metaphor for every kind of situation someone wearing an AR headset could find themselves in. The experience a user will have in AR can be influenced by many elements in the real-world environment. Different background colors, environmental noise, and the people in the environment; all of it influence the experience, which, of course, can never be replicated in a laboratory study.

Another limitation of this research is the device used in five out of the six experiments, the Microsoft Hololens 2. Its FOV is at a relatively narrow 50°, which can limit the perception of virtual objects and environments. This is especially important with world-fixed content, as locations in the real world may be visible, but outside the render area. The device is also physically bulky and may cause discomfort during extended use. Additionally, interaction methods, such as hand tracking and voice commands, can occasionally be unreliable in varying lighting or noisy environments. However, our highlights include much insight into general problems with AR notification. It is unknown whether improved hardware would influence our results, and if it did, in what strength and direction.

9.3 Future Work

Our presented research highlighted several challenges that must be addressed for AR to be truly beneficial for notification delivery. These challenges inform the following future research directions:

One key finding was that participants benefited most from notifications that were important to them. However, some users expressed a preference for receiving all notifications on the headset.

This suggests that a simple on/off toggle for notifications is insufficient; instead, we need dynamic display mechanisms that differentiate notifications based on their priority. Future research should explore the use of color, animations, and icons to convey priority levels. Additionally, multimodal indicators such as audio cues or subtle haptic feedback (e.g., vibrations) could be investigated as a means of enhancing notification delivery. Participants valued unobtrusive interaction methods that allowed them to read or dismiss notifications without drawing attention from conversation partners. Future studies should explore interaction techniques such as unimodal eye-tracking, microgestures, and hand-worn devices (e.g., a smart band with a pinch gesture). Other potential interaction methods include subtle swipe interactions on the glasses' frame or discreet head gestures. These techniques should be evaluated in terms of usability, efficiency, and social acceptability. A recurring concern was privacy, as the presence of smartglasses made both wearers and bystanders uncomfortable due to potential recording capabilities. Future research should investigate ways for AR glasses to clearly communicate their recording status to onlookers. This should go beyond user manuals and be immediately evident to bystanders. Additionally, research could examine whether the presence of AR glasses—whether recording or not—alters social behavior during conversations. To better understand multimodal interaction, future work could explore how users switch between different interaction techniques depending on the context. A controlled task environment, such as a cooking scenario, could be used to allow participants to interact with notifications using all researched techniques simultaneously. This would help identify which techniques are most effective under different conditions and provide insights into user preferences for adaptive interaction methods. Our current findings may have been constrained by hardware limitations. Replicating our experiments with more advanced AR headsets could improve and validate our results. For example, in our previous study (section 5.1), wrist-based notifications were effective when users' hands were within the FOV but performed poorly when the hands were out of view. A headset with a larger FOV could enhance the functionality of wrist-based notifications. Similarly, the accuracy of pinch gestures and the reliability of eye-tracking could improve with better hardware, warranting further investigation. A way to improve the pinch gesture detection could be to use a wrist worn gesture-tracking system such as DoublePoint[109], which enables the use of a smartwatch to detect pinch gestures. A logical next step would be to develop a system based on our laboratory findings and evaluate it in real-world scenarios. However, current AR head-mounted displays (HMDs) are not yet comfortable or unobtrusive enough for prolonged daily use. As AR hardware evolves—becoming lighter, more compact, and offering better battery life—future research should revisit our findings in broader real-world contexts. Additionally, improved hardware could help address some of the limitations we identified, allowing for a more comprehensive evaluation of AR notification interactions in everyday environments.

These future research directions will help refine AR notification delivery, making it more context-aware, privacy-conscious, and seamlessly integrated into users' daily lives.

Chapter 10

Conclusion

This thesis has examined different approaches to displaying notifications in AR environments and optimizing interactions with them. In the first experiment, we studied the location of notifications in a 3D AR space by using a memory card game with interrupting notifications on an AR headset. Each notification included a prompt that users needed to respond to, and we found that placing notifications in the bottom center of the FOV is ideal for general or mobile tasks, while notifications placed in the world are ideal for stationary tasks. The second experiment also looked at the location of notifications, as well as the modality. Using a simulated cooking environment, we displayed notifications about the task either in a fixed list positioned in the environment or directly located above virtual objects that caused the notification. We also investigated the effects of the modality of notifications by sending either audio notifications, visual notifications, or a combination of both. We found that combined notifications located on the source objects provided the best results. The third experiment investigated interaction with notifications in a simulated cooking environment, as well as the display of multiple notifications. We found that having a notification list attached to the body for quick access was well-liked by users. Additionally, touch interaction was found to be superior to unimodal alternatives such as gaze or voice. In the fourth experiment, we combined the modalities used in the previous experiment to explore the potential benefits of multimodal interaction over unimodal interaction. We found that multimodal techniques performed similarly to touch interaction; however, issues with fusing multiple modalities led to perceived unreliability. In a fifth experiment, we investigated the effects of notifications on smart glasses in the real world. Participants wore smart glasses for five days, during which the devices displayed notifications from their smartphones. We found that participants liked the ability to access their notification hands-free, seeing time-critical notifications instantly and being able to identify important notifications. However, there are still critical barriers for user adoption of this technology, such as social hurdles, attentional costs and interaction limitations. Finally, we investigated means to help user's resume tasks after being interrupted in a sixth experiment. Using AR, we developed a system that could help reduce distractions by presenting a visual cue when an interruption happens, at the current location in a task. We found that the resumption lag after short and long interruptions was significantly longer without the visual cue present. This could help in minimizing the distractions introduced by notifications, as the system could place a cue right when an intrusive notification is delivered. From my work we can derive several design recommendations:

- Notifications in AR should be placed in the bottom center of the FOV except in stationary tasks, where they should be placed adjacent to the task, or task-relevant locations.
- Notifications placed on the wrist should be used when it can be guaranteed that the hands are in the visible area of the user.
- Touch interaction should be used for the notifications, with hands-free fallback techniques.
- Mulitomodal interaction should be prioritized over unimodal techniques, with a focus on reliability of the interaction.
- AR visual cues can help with task resumption and should be deployed to minimize attentional costs of notifications.
- Notifications should include audio and visual components for optimal noticeability.
- Importance and circumstance of notifications influence the willingness of users to be interrupted by them.
- Notifications should be stored in an easily accessible way for later attendance.

Notifications have been—and will continue to be—integral parts of daily computing. Although they are very useful, they can also have detrimental effects on productivity and well-being. With smartphones carried in our pockets for most of the day, they are already far more invasive than notifications on a desktop. In the not-so-distant future, when smartglasses might be widely worn, the invasiveness of notifications could be very high, while their usefulness could be equally significant.

This research will assist in designing these systems to minimize the impact of notifications when these devices become pervasive. Since AR devices interact with the real world, whose properties can change dramatically, they will also need to better understand the current situation and adapt accordingly. Notification delivery is no exception, and there is, unfortunately, no "one-size-fits-all" solution… yet.

Bibliography

[1] M. Czerwinski, E. Cutrell, and E. Horvitz, "Instant Messaging and Interruption: Influence of Task Type on Performance," en,

[2] B. P. Bailey, J. A. Konstan, and J. V. Carlis, "The Effects of Interruptions on Task Performance, Annoyance, and Anxiety in the User Interface," en, p. 9,

[3] D. C. McFarlane, "Comparison of Four Primary Methods for Coordinating the Interruption of People in Human-Computer Interaction," en, *Human–Computer Interaction*, vol. 17, no. 1, pp. 63–139, Mar. 2002, ISSN: 0737-0024, 1532-7051. DOI: 10.1207/S15327051HCI1701_2. [Online]. Available: https://www.tandfonline.com/doi/full/10.1207/S15327051HCI1701_2 (visited on 07/03/2024).

[4] R. T. Azuma, "The road to ubiquitous consumer augmented reality systems," en, *Human Behavior and Emerging Technologies*, vol. 1, no. 1, pp. 26–32, 2019, _eprint: https://onlinelibrary.wiley.co ISSN: 2578-1863. DOI: 10.1002/hbe2.113. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/hbe2.113 (visited on 08/04/2022).

[5] K. Matsuda, *Hyper-Reality*, 2016. [Online]. Available: http://hyper-reality.co/ (visited on 03/18/2025).

[6] P. Milgram and F. Kishino, "A TAXONOMY OF MIXED REALITY VISUAL DISPLAYS," en, *IEICE Transactions on Information Systems*, vol. Vol E77-D, no. No.12, Dec. 1994.

[7] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev, *3D User Interfaces: Theory and Practice, 2nd Edition*. USA: Addison-Wesley Professional, Apr. 2017, ISBN: 978-0-13-403432-4.

[8] S. T. Iqbal and B. P. Bailey, "Oasis: A framework for linking notification delivery to the perceptual structure of goal-directed tasks," en, *ACM Transactions on Computer-Human Interaction*, vol. 17, no. 4, pp. 1–28, Dec. 2010, ISSN: 1073-0516, 1557-7325. DOI: 10.

1145/1879831.1879833. [Online]. Available: https://dl.acm.org/doi/10.1145/1879831.1879833 (visited on 05/31/2022).

[9] M. DeGusta, *Are Smart Phones Spreading Faster than Any Technology in Human History?* en, 2012. [Online]. Available: https://www.technologyreview.com/2012/05/09/186160/are-smart-phones-spreading-faster-than-any-technology-in-human-history/ (visited on 02/02/2022).

[10] U. Acer, A. Mashhadi, C. Forlivesi, and F. Kawsar, *Energy Efficient Scheduling for Mobile Push Notifications*. Aug. 2015, vol. 2, Journal Abbreviation: EAI Endorsed Transactions on Energy Web Publication Title: EAI Endorsed Transactions on Energy Web. DOI: 10.4108/eai.22-7-2015.2260067.

[11] M. Pielot, K. Church, and R. de Oliveira, "An in-situ study of mobile phone notifications," en, in *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, Toronto ON Canada: ACM, Sep. 2014, pp. 233–242, ISBN: 978-1-4503-3004-6. DOI: 10.1145/2628363.2628364. [Online]. Available: https://dl.acm.org/doi/10.1145/2628363.2628364 (visited on 04/24/2023).

[12] U. Lee, J. Lee, M. Ko, *et al.*, "Hooked on smartphones: An exploratory study on smartphone overuse among college students," en, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto Ontario Canada: ACM, Apr. 2014, pp. 2327–2336, ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557366. [Online]. Available: https://dl.acm.org/doi/10.1145/2556288.2557366 (visited on 04/29/2024).

[13] J. Radesky, H. Weeks, A. Schaller, S. Mann, and A. Lenhart, "Constant Companion: A Week in the Life of a Young Person's Smartphone Use," Common Sense, San Francisco, California, USA, Tech. Rep., 2023.

[14] Apple, *Use notifications on your iPhone or iPad*, en, Sep. 2022. [Online]. Available: https://support.apple.com/en-us/HT201925 (visited on 06/15/2023).

[15] alison@uxmag.com, *Designing Notifications for Apps*, en-US, Mar. 2020. [Online]. Available: https://uxmag.com/articles/designing-notifications-for-apps (visited on 06/09/2023).

[16] J. Nielsen, "Enhancing the explanatory power of usability heuristics," en, in *Proceedings of the SIGCHI conference on Human factors in computing systems celebrating interdependence - CHI '94*, Boston, Massachusetts, United States: ACM Press, 1994, pp. 152–158, ISBN: 978-0-89791-650-9. DOI: 10.1145/191666.191729. [Online]. Available: http://portal.acm.org/citation.cfm?doid=191666.191729 (visited on 07/01/2024).

[17] D. S. McCrickard, C. M. Chewar, J. P. Somervell, and A. Ndiwalana, "A model for notification systems evaluation—assessing user goals for multitasking activity," en, *ACM Transactions on Computer-Human Interaction*, vol. 10, no. 4, pp. 312–338, Dec. 2003, ISSN: 1073-0516, 1557-7325. DOI: 10.1145/966930.966933. [Online]. Available: https://dl.acm.org/doi/10.1145/966930.966933 (visited on 02/01/2024).

[18] M. Pielot, A. Vradi, and S. Park, "Dismissed!: A detailed exploration of how mobile phone users handle push notifications," en, in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, Barcelona Spain: ACM, Sep. 2018, pp. 1–11, ISBN: 978-1-4503-5898-9. DOI: 10.1145/3229434.3229445. [Online]. Available: https://dl.acm.org/doi/10.1145/3229434.3229445 (visited on 05/31/2022).

[19] A. Mehrotra, M. Musolesi, R. Hendley, and V. Pejovic, "Designing content-driven intelligent notification mechanisms for mobile applications," en, in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, Osaka, Japan: ACM Press, 2015, pp. 813–824, ISBN: 978-1-4503-3574-4. DOI: 10.1145/2750858.2807544. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2750858.2807544 (visited on 07/20/2022).

[20] H. H. Clark, *Using Language* ('Using' Linguistic Books). Cambridge: Cambridge University Press, 1996, ISBN: 978-0-521-56158-7. DOI: 10.1017/CBO9780511620539. [Online].

Available: https://www.cambridge.org/core/books/using-language/4E7EBC4EC742C26436F6CF187C43
(visited on 07/02/2024).

[21] A. Smith and D. Page, "US Smartphone Use in 2015," en, [Online]. Available: https://www.pewresearch.org/wp-content/uploads/sites/9/2015/03/PI_Smartphones_0401151.pdf.

[22] A. Mehrotra, V. Pejovic, J. Vermeulen, R. Hendley, and M. Musolesi, "My Phone and Me: Understanding People's Receptivity to Mobile Notifications," en, in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose California USA: ACM, May 2016, pp. 1021–1032, ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858566. [Online]. Available: https://dl.acm.org/doi/10.1145/2858036.2858566 (visited on 08/31/2021).

[23] A. Exler, M. Braith, A. Schankin, and M. Beigl, "Preliminary investigations about interruptibility of smartphone users at specific place types," en, in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, Heidelberg Germany: ACM, Sep. 2016, pp. 1590–1595, ISBN: 978-1-4503-4462-3. DOI: 10.1145/2968219.2968554. [Online]. Available: https://dl.acm.org/doi/10.1145/2968219.2968554 (visited on 02/10/2025).

[24] C. Stothart, A. Mitchum, and C. Yehnert, "The attentional cost of receiving a cell phone notification.," en, *Journal of Experimental Psychology: Human Perception and Performance*, vol. 41, no. 4, pp. 893–897, Aug. 2015, ISSN: 1939-1277, 0096-1523. DOI: 10.1037/xhp0000100. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/xhp0000100 (visited on 08/31/2021).

[25] I. H. Robertson, T. Manly, J. Andrade, B. T. Baddeley, and J. Yiend, "'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects," en, *Neuropsychologia*, vol. 35, no. 6, pp. 747–758, May 1997, ISSN: 00283932. DOI: 10.1016/S0028-3932(97)00015-8. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0028393297000158 (visited on 06/20/2024).

[26] K. Kushlev, J. Proulx, and E. W. Dunn, “"Silence Your Phones": Smartphone Notifications Increase Inattention and Hyperactivity Symptoms,” en, in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose California USA: ACM, May 2016, pp. 1011–1020, ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858359. [Online]. Available: https://dl.acm.org/doi/10.1145/2858036.2858359 (visited on 02/02/2022).

[27] D. Warnock, M. R. McGee-Lennon, and S. Brewster, “The impact of unwanted multimodal notifications,” en, in *Proceedings of the 13th international conference on multimodal interfaces - ICMI ’11*, Alicante, Spain: ACM Press, 2011, p. 177, ISBN: 978-1-4503-0641-6. DOI: 10.1145/2070481.2070510. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2070481.2070510 (visited on 05/31/2022).

[28] M. Pielot and L. Rello, “Productive, anxious, lonely: 24 hours without push notifications,” en, in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, Vienna Austria: ACM, Sep. 2017, pp. 1–11, ISBN: 978-1-4503-5075-4. DOI: 10.1145/3098279.3098526. [Online]. Available: https://dl.acm.org/doi/10.1145/3098279.3098526 (visited on 05/31/2022).

[29] S. T. Iqbal and E. Horvitz, “Notifications and awareness: A field study of alert usage and preferences,” en, in *Proceedings of the 2010 ACM conference on Computer supported cooperative work - CSCW ’10*, Savannah, Georgia, USA: ACM Press, 2010, p. 27, ISBN: 978-1-60558-795-0. DOI: 10.1145/1718918.1718926. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1718918.1718926 (visited on 10/06/2021).

[30] E. Horvitz, J. Apacible, and M. Subramani, “Balancing Awareness and Interruption: Investigation of Notification Deferral Policies,” en, in *User Modeling 2005*, D. Hutchison, T. Kanade, J. Kittler, *et al.*, Eds., vol. 3538, Series Title: Lecture Notes in Computer Science, Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 433–437, ISBN: 978-3-540-27885-6 978-3-540-31878-1. DOI: 10.1007/11527886_59. [Online]. Available: http://link.springer.com/10.1007/11527886_59 (visited on 07/04/2024).

[31] J. E. Fischer, C. Greenhalgh, and S. Benford, "Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications," en, in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*, Stockholm, Sweden: ACM Press, 2011, p. 181, ISBN: 978-1-4503-0541-9. DOI: 10.1145/2037373.2037402. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2037373.2037402 (visited on 09/01/2021).

[32] M. Pielot, B. Cardoso, K. Katevas, J. Serrà, A. Matic, and N. Oliver, "Beyond Interruptibility: Predicting Opportune Moments to Engage Mobile Phone Users," en, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–25, Sep. 2017, ISSN: 2474-9567. DOI: 10.1145/3130956. [Online]. Available: https://dl.acm.org/doi/10.1145/3130956 (visited on 04/24/2023).

[33] N. Fitz, K. Kushlev, R. Jagannathan, T. Lewis, D. Paliwal, and D. Ariely, "Batching smartphone notifications can improve well-being," en, *Computers in Human Behavior*, vol. 101, pp. 84–94, Dec. 2019, ISSN: 07475632. DOI: 10.1016/j.chb.2019.07.016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0747563219302596 (visited on 02/02/2022).

[34] C. Park, J. Lim, J. Kim, S.-J. Lee, and D. Lee, "Don't Bother Me. I'm Socializing! A Breakpoint-Based Smartphone Notification System," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '17, New York, NY, USA: Association for Computing Machinery, Feb. 2017, pp. 541–554, ISBN: 978-1-4503-4335-0. DOI: 10.1145/2998181.2998189. [Online]. Available: https://doi.org/10.1145/2998181.2998189 (visited on 02/02/2022).

[35] V. Chotpitayasunondh and K. M. Douglas, "The effects of "phubbing" on social interaction," en, *Journal of Applied Social Psychology*, vol. 48, no. 6, pp. 304–316, Jun. 2018, ISSN: 0021-9029, 1559-1816. DOI: 10.1111/jasp.12506. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/jasp.12506 (visited on 02/11/2024).

[36] J. Orlosky, K. Kiyokawa, and H. Takemura, "Managing Mobile Text in Head Mounted Displays: Studies on Visual Preference and Text Placement," en, vol. 18, no. 2, p. 12, 2014.

[37] S. A. Sobrinho-Junior, A. C. N. de Almeida, A. A. P. Ceabras, C. L. da Silva Carvalho, T. B. Lino, and G. Christofoletti, "Risks of Accidents Caused by the Use of Smartphone by Pedestrians Are Task- and Environment-Dependent," *International Journal of Environmental Research and Public Health*, vol. 19, no. 16, p. 10 320, Aug. 2022, ISSN: 1661-7827. DOI: 10.3390/ijerph191610320. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9408215/ (visited on 07/06/2024).

[38] R. Batchelor, *Library Guides: Future Phonics: History of Phones*, en, 2013. [Online]. Available: https://georgiasouthern.libguides.com/c.php?g=612229&p=4545365 (visited on 03/18/2025).

[39] I. E. Sutherland, "The Ultimate Display," en, *Proceedings of IFIP Congress*, pp. 506–508, 1965.

[40] I. E. Sutherland, "A head-mounted three dimensional display," en, in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I on - AFIPS '68 (Fall, part I)*, San Francisco, California: ACM Press, 1968, p. 757. DOI: 10.1145/1476589.1476686. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1476589.1476686 (visited on 02/10/2025).

[41] T. Caudell and D. Mizell, "Augmented reality: An application of heads-up display technology to manual manufacturing processes," in *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, vol. ii, Jan. 1992, 659–669 vol.2. DOI: 10.1109/HICSS.1992.183317. [Online]. Available: https://ieeexplore.ieee.org/document/183317/?arnumber=183317 (visited on 02/10/2025).

[42] R. T. Azuma, "A Survey of augmented reality," en, *Presence: Teleoperators and Virtual Environments*, pp. 355–385, 1997.

[43] Google, *Tech specs - Google Glass Help*, 2013. [Online]. Available: https://support.google.com/glass/answer/3064128?hl=en (visited on 07/07/2024).

[44] M. Billinghurst and H. Kato, "Collaborative Mixed Reality," en, in *Mixed Reality*, Y. Ohta and H. Tamura, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 261–284, ISBN: 978-3-642-87514-4 978-3-642-87512-0. DOI: 10.1007/978-3-642-87512-0_15. [Online]. Available: http://link.springer.com/10.1007/978-3-642-87512-0_15 (visited on 03/04/2021).

[45] S. Feiner, B. MacIntyre, M. Haupt, and E. Solomon, "Windows on the world: 2D windows for 3D augmented reality," en, in *Proceedings of the 6th annual ACM symposium on User interface software and technology*, Atlanta Georgia USA: ACM, Dec. 1993, pp. 145–155, ISBN: 978-0-89791-628-8. DOI: 10.1145/168642.168657. [Online]. Available: https://dl.acm.org/doi/10.1145/168642.168657 (visited on 07/07/2024).

[46] F. Lu, S. Davari, L. Lisle, Y. Li, and D. A. Bowman, "Glanceable AR: Evaluating Information Access Methods for Head-Worn Augmented Reality," en, in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Atlanta, GA, USA: IEEE, Mar. 2020, pp. 930–939, ISBN: 978-1-7281-5608-8. DOI: 10.1109/VR46266.2020.00113. [Online]. Available: https://ieeexplore.ieee.org/document/9089433/ (visited on 03/04/2021).

[47] S. Davari, F. Lu, and D. A. Bowman, "Validating the Benefits of Glanceable and Context-Aware Augmented Reality for Everyday Information Access Tasks," en, in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Christchurch, New Zealand: IEEE, Mar. 2022, pp. 436–444, ISBN: 978-1-66549-617-9. DOI: 10.1109/VR51125.2022.00063. [Online]. Available: https://ieeexplore.ieee.org/document/9756789/ (visited on 06/20/2024).

[48] C.-C. Lin, "Effects of contrast ratio and text color on visual performance with TFT-LCD," en, *International Journal of Industrial Ergonomics*, vol. 31, no. 2, pp. 65–72, Feb. 2003,

ISSN: 01698141. DOI: 10.1016/S0169-8141(02)00175-0. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169814102001750 (visited on 07/08/2024).

[49] S. Debernardis, M. Fiorentino, M. Gattullo, G. Monno, and A. Uva, "Text Readability in Head-Worn Displays: Color and Style Optimization in Video vs. Optical See-Through Devices," *IEEE transactions on visualization and computer graphics*, vol. 20, pp. 125–39, Jan. 2014. DOI: 10.1109/TVCG.2013.86.

[50] J. L. Gabbard, J. E. Swan, and D. Hix, "The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality," en, *Presence: Teleoperators and Virtual Environments*, vol. 15, no. 1, pp. 16–32, Feb. 2006, ISSN: 1054-7460. DOI: 10.1162/pres.2006.15.1.16. [Online]. Available: https://direct.mit.edu/pvar/article/15/1/16-32/18626 (visited on 10/06/2021).

[51] R. Rzayev, P. W. Woźniak, T. Dingler, and N. Henze, "Reading on Smart Glasses: The Effect of Text Position, Presentation Type and Walking," en, in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC Canada: ACM, Apr. 2018, pp. 1–9, ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173619. [Online]. Available: https://dl.acm.org/doi/10.1145/3173574.3173619 (visited on 08/31/2021).

[52] J. Flittner, J. Luksas, and J. L. Gabbard, "Predicting User Performance in Augmented Reality User Interfaces with Image Analysis Algorithms," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 64, no. 1, pp. 2108–2112, Dec. 2020, Publisher: SAGE Publications Inc, ISSN: 2169-5067. DOI: 10.1177/1071181320641511. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1177/1071181320641511 (visited on 12/02/2021).

[53] G. A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information[1]," en, 1956.

[54] T. Papadopoulos, K. Evangelidis, T. H. Kaskalis, G. Evangelidis, and S. Sylaiou, "Interactions in Augmented and Mixed Reality: An Overview," en, *Applied Sciences*, vol. 11,

no. 18, p. 8752, Sep. 2021, ISSN: 2076-3417. DOI: 10.3390/app11188752. [Online]. Available: https://www.mdpi.com/2076-3417/11/18/8752 (visited on 05/07/2024).

[55] N. Chaconas and T. Hollerer, "An Evaluation of Bimanual Gestures on the Microsoft HoloLens," en, in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Reutlingen: IEEE, Mar. 2018, pp. 1–8, ISBN: 978-1-5386-3365-6. DOI: 10.1109/VR.2018.8446320. [Online]. Available: https://ieeexplore.ieee.org/document/8446320/ (visited on 06/20/2024).

[56] Z. Wang, H. Wang, H. Yu, and F. Lu, "Interaction With Gaze, Gesture, and Speech in a Flexibly Configurable Augmented Reality System," *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 5, pp. 524–534, Oct. 2021, Conference Name: IEEE Transactions on Human-Machine Systems, ISSN: 2168-2305. DOI: 10.1109/THMS.2021.3097973.

[57] M. Lee, M. Billinghurst, W. Baek, R. Green, and W. Woo, "A usability study of multimodal input in an augmented reality environment," en, *Virtual Reality*, vol. 17, no. 4, pp. 293–305, Nov. 2013, ISSN: 1359-4338, 1434-9957. DOI: 10.1007/s10055-013-0230-0. [Online]. Available: http://link.springer.com/10.1007/s10055-013-0230-0 (visited on 06/16/2023).

[58] M. Parisay, C. Poullis, and M. Kersten-Oertel, "EyeTAP: Introducing a multimodal gaze-based technique using voice inputs with a comparative analysis of selection techniques," en, *International Journal of Human-Computer Studies*, vol. 154, p. 102 676, Oct. 2021, ISSN: 10715819. DOI: 10.1016/j.ijhcs.2021.102676. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S107158192100094X (visited on 01/30/2023).

[59] K. Pfeuffer, H. Gellersen, and M. Gonzalez-Franco, "Design Principles and Challenges for Gaze + Pinch Interaction in XR," *IEEE Computer Graphics and Applications*, vol. 44, no. 3, pp. 74–81, May 2024, Conference Name: IEEE Computer Graphics and Applications, ISSN: 1558-1756. DOI: 10.1109/MCG.2024.3382961. [Online]. Available: https://ieeexplore.ieee.org/document/10568496/?arnumber=10568496 (visited on 10/07/2024).

[60] S. Ghosh, L. Winston, N. Panchal, *et al.*, "NotifiVR: Exploring Interruptions and Notifications in Virtual Reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1447–1456, Apr. 2018, Conference Name: IEEE Transactions on Visualization and Computer Graphics, ISSN: 1941-0506. DOI: 10.1109/TVCG.2018.2793698.

[61] A. Zenner, M. Speicher, S. Klingner, D. Degraen, F. Daiber, and A. Krüger, "Immersive Notification Framework: Adaptive & Plausible Notifications in Virtual Reality," en, in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC Canada: ACM, Apr. 2018, pp. 1–6, ISBN: 978-1-4503-5621-3. DOI: 10.1145/3170427.3188505. [Online]. Available: https://dl.acm.org/doi/10.1145/3170427.3188505 (visited on 02/09/2024).

[62] M. Slater, "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments," en, *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3549–3557, Dec. 2009, ISSN: 0962-8436, 1471-2970. DOI: 10.1098/rstb.2009.0138. [Online]. Available: https://royalsocietypublishing.org/doi/10.1098/rstb.2009.0138 (visited on 07/08/2024).

[63] R. Rzayev, S. Mayer, C. Krauter, and N. Henze, "Notification in VR: The Effect of Notification Placement, Task and Environment," en, in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, Barcelona Spain: ACM, Oct. 2019, pp. 199–211, ISBN: 978-1-4503-6688-5. DOI: 10.1145/3311350.3347190. [Online]. Available: https://dl.acm.org/doi/10.1145/3311350.3347190 (visited on 07/04/2021).

[64] R. Rzayev, S. Korbely, M. Maul, A. Schark, V. Schwind, and N. Henze, "Effects of Position and Alignment of Notifications on AR Glasses during Social Interaction," en, in *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, Tallinn Estonia: ACM, Oct. 2020, pp. 1–11, ISBN: 978-1-4503-7579-5. DOI: 10.1145/3419249.3420095. [Online]. Available: https://dl.acm.org/doi/10.1145/3419249.3420095 (visited on 08/30/2021).

[65] Microsoft, *HoloLens 2—Overview, Features, and Specs | Microsoft HoloLens*, en-US, 2021. [Online]. Available: https://www.microsoft.com/en-us/hololens/hardware (visited on 11/20/2024).

[66] H. Lee and W. Woo, "Exploring the Effects of Augmented Reality Notification Type and Placement in AR HMD while Walking," en, in *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, Shanghai, China: IEEE, Mar. 2023, pp. 519–529, ISBN: 9798350348156. DOI: 10.1109/VR55154.2023.00067. [Online]. Available: https://ieeexplore.ieee.org/document/10108490/ (visited on 04/10/2024).

[67] A. Lucero and A. Vetek, "NotifEye: Using interactive glasses to deal with notifications while walking in public," en, in *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*, Funchal Portugal: ACM, Nov. 2014, pp. 1–10, ISBN: 978-1-4503-2945-3. DOI: 10.1145/2663806.2663824. [Online]. Available: https://dl.acm.org/doi/10.1145/2663806.2663824 (visited on 08/31/2021).

[68] Y. Rogers and P. Marshall, *Research in the Wild* (Synthesis Lectures on Human-Centered Informatics), en. Cham: Springer International Publishing, 2017, ISBN: 978-3-031-01092-7 978-3-031-02220-3. DOI: 10.1007/978-3-031-02220-3. [Online]. Available: https://link.springer.com/10.1007/978-3-031-02220-3 (visited on 07/07/2024).

[69] Epson, *Moverio BT-100 Wearable Display | Products | Epson US*, 2013. [Online]. Available: https://epson.com/Certified-ReNew/Wearables/Moverio-BT-100-Wearable-Display/p/V11H423020 (visited on 07/07/2024).

[70] M. J. Lazaro, S. Kim, J. Lee, J. Chun, and M.-H. Yun, "Interaction Modalities for Notification Signals in Augmented Reality," en, in *Proceedings of the 2021 International Conference on Multimodal Interaction*, Montréal QC Canada: ACM, Oct. 2021, pp. 470–477, ISBN: 978-1-4503-8481-0. DOI: 10.1145/3462244.3479898. [Online]. Available: https://dl.acm.org/doi/10.1145/3462244.3479898 (visited on 02/28/2022).

[71] K. Pfeuffer, J. Obernolte, F. Dietz, *et al.*, “PalmGazer: Unimanual Eye-hand Menus in Augmented Reality,” en, in *Proceedings of the 2023 ACM Symposium on Spatial User Interaction*, Sydney NSW Australia: ACM, Oct. 2023, pp. 1–12, ISBN: 9798400702815. DOI: 10.1145/3607822.3614523. [Online]. Available: https://dl.acm.org/doi/10.1145/3607822.3614523 (visited on 10/30/2023).

[72] A. K. Faulhaber, M. Hoppe, and L. Schmidt, “Evaluation of Priority-Dependent Notifications for Smart Glasses Based on Peripheral Visual Cues,” en, *i-com*, vol. 21, no. 2, pp. 239–252, Aug. 2022, ISSN: 2196-6826, 1618-162X. DOI: 10.1515/icom-2022-0022. [Online]. Available: https://www.degruyter.com/document/doi/10.1515/icom-2022-0022/html (visited on 07/07/2024).

[73] Xreal, *Nreal Light: Full Specification*, en, 2019. [Online]. Available: https://vr-compare.com/headset/nreallight (visited on 07/07/2024).

[74] W. K. Kirchner, “Age differences in short-term retention of rapidly changing information.,” en, *Journal of Experimental Psychology*, vol. 55, no. 4, pp. 352–358, 1958, ISSN: 0022-1015. DOI: 10.1037/h0043688. [Online]. Available: https://doi.apa.org/doi/10.1037/h0043688 (visited on 07/07/2024).

[75] N. Janaka, S. Zhao, and S. Sapkota, “Can Icons Outperform Text? Understanding the Role of Pictograms in OHMD Notifications,” en, in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Hamburg Germany: ACM, Apr. 2023, pp. 1–23, ISBN: 978-1-4503-9421-5. DOI: 10.1145/3544548.3580891. [Online]. Available: https://dl.acm.org/doi/10.1145/3544548.3580891 (visited on 12/12/2023).

[76] N. Janaka, S. Zhao, and S. Chan, “NotiFade: Minimizing OHMD Notification Distractions Using Fading,” in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA ’23, New York, NY, USA: Association for Computing Machinery, Apr. 2023, pp. 1–9, ISBN: 978-1-4503-9422-2. DOI: 10.1145/3544549.

3585784. [Online]. Available: https://dl.acm.org/doi/10.1145/3544549.3585784 (visited on 01/12/2024).

[77] T. Kosch, A. Matviienko, F. Müller, *et al.*, "NotiBike: Assessing Target Selection Techniques for Cyclist Notifications in Augmented Reality," en, *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. MHCI, pp. 1–24, Sep. 2022, ISSN: 2573-0142. DOI: 10.1145/3546732. [Online]. Available: https://dl.acm.org/doi/10.1145/3546732 (visited on 02/22/2023).

[78] N. Janaka, J. Gao, L. Zhu, *et al.*, "GlassMessaging: Towards Ubiquitous Messaging Using OHMDs," en, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 3, pp. 1–32, Sep. 2023, ISSN: 2474-9567. DOI: 10.1145/3610931. [Online]. Available: https://dl.acm.org/doi/10.1145/3610931 (visited on 12/12/2023).

[79] L. Plabst, S. Oberdörfer, F. R. Ortega, and F. Niebling, "Push the Red Button: Comparing Notification Placement with Augmented and Non-Augmented Tasks in AR," en, in *Proceedings of the 2022 ACM Symposium on Spatial User Interaction*, Online CA USA: ACM, Dec. 2022, pp. 1–11, ISBN: 978-1-4503-9948-7. DOI: 10.1145/3565970.3567701. [Online]. Available: https://dl.acm.org/doi/10.1145/3565970.3567701 (visited on 04/23/2024).

[80] A. Raikwar, L. Plabst, A. U. Batmaz, F. Niebling, and F. R. Ortega, "Ping! Your Food is Ready: Comparing Different Notification Techniques in 3D AR Cooking Environment," English, IEEE Computer Society, Oct. 2024, pp. 1157–1166, ISBN: 979-8-3315-1647-5. DOI: 10.1109/ISMAR62088.2024.00132. [Online]. Available: https://www.computer.org/csdl/proceedings-article/ismar/2024/164700b157/22f0kYffkUE (visited on 01/13/2025).

[81] A. Raikwar, L. Plabst, and F. R. Ortega, "ARtisan Bistro: A Cooking Task Environment to Conduct Studies in Augmented Reality," en, in *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, Singapore, Singapore: IEEE, Oct. 2022, pp. 909–910, ISBN: 978-1-66545-365-3. DOI: 10.1109/ISMAR-Adjunct57072.

2022.00200. [Online]. Available: https://ieeexplore.ieee.org/document/9974263/ (visited on 04/23/2024).

[82] L. Plabst, A. Raikwar, S. Oberdörfer, F. R. Ortega, and F. Niebling, "Exploring Unimodal Notification Interaction and Display Methods in Augmented Reality," en, in *29th ACM Symposium on Virtual Reality Software and Technology*, Christchurch New Zealand: ACM, Oct. 2023, pp. 1–11, ISBN: 9798400703287. DOI: 10.1145/3611659.3615683. [Online]. Available: https://dl.acm.org/doi/10.1145/3611659.3615683 (visited on 04/23/2024).

[83] J.-C. Martin, "TYCOON: Theoretical Framework and Software Tools for Multimodal Interfaces," en, 1997.

[84] D. Collingridge, *The social control of technology*. St. Martin's Press, 1980, ISBN: 978-0-312-73168-7. [Online]. Available: https://books.google.com/books?id=hCSdAQAACAAJ.

[85] B. Hofmann, D. Haustein, and L. Landeweerd, "Smart-Glasses: Exposing and Elucidating the Ethical Issues," en, *Science and Engineering Ethics*, vol. 23, no. 3, pp. 701–721, Jun. 2017, ISSN: 1353-3452, 1471-5546. DOI: 10.1007/s11948-016-9792-z. [Online]. Available: http://link.springer.com/10.1007/s11948-016-9792-z (visited on 01/28/2025).

[86] R.-B. Meta, *Ray-Ban | Meta AI glasses | Ray-Ban® USA*, en-US, 2024. [Online]. Available: https://www.ray-ban.com/usa/ray-ban-meta-ai-glasses (visited on 03/18/2025).

[87] M. Z. Iqbal and A. G. Campbell, "Adopting smart glasses responsibly: Potential benefits, ethical, and privacy concerns with Ray-Ban stories," en, *AI and Ethics*, vol. 3, no. 1, pp. 325–327, Feb. 2023, ISSN: 2730-5953, 2730-5961. DOI: 10.1007/s43681-022-00155-7. [Online]. Available: https://link.springer.com/10.1007/s43681-022-00155-7 (visited on 01/29/2025).

[88] P. A. Rauschnabel, J. He, and Y. K. Ro, "Antecedents to the adoption of augmented reality smart glasses: A closer look at privacy risks," en, *Journal of Business Research*, vol. 92, pp. 374–384, Nov. 2018, ISSN: 01482963. DOI: 10.1016/j.jbusres.2018.08.008. [Online].

Available: https://linkinghub.elsevier.com/retrieve/pii/S0148296318303849 (visited on 01/29/2025).

[89] R. McNaney, J. Vines, D. Roggen, *et al.*, “Exploring the acceptability of google glass as an everyday assistive device for people with parkinson’s,” en, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto Ontario Canada: ACM, Apr. 2014, pp. 2551–2554, ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557092. [Online]. Available: https://dl.acm.org/doi/10.1145/2556288.2557092 (visited on 01/28/2025).

[90] J. Grubert, D. Hamacher, R. Mecke, *et al.*, “Extended investigations of user-related issues in mobile industrial AR,” en, in *2010 IEEE International Symposium on Mixed and Augmented Reality*, Seoul, Korea (South): IEEE, Oct. 2010, pp. 229–230, ISBN: 978-1-4244-9343-2 978-1-4244-9346-3. DOI: 10.1109/ISMAR.2010.5643581. [Online]. Available: https://ieeexplore.ieee.org/document/5643581/ (visited on 01/29/2025).

[91] F. Steinicke and G. Bruder, “A self-experimentation report about long-term use of fully-immersive technology,” en, in *Proceedings of the 2nd ACM symposium on Spatial user interaction*, Honolulu Hawaii USA: ACM, Oct. 2014, pp. 66–69, ISBN: 978-1-4503-2820-3. DOI: 10.1145/2659766.2659767. [Online]. Available: https://dl.acm.org/doi/10.1145/2659766.2659767 (visited on 01/29/2025).

[92] V. Biener, S. Kalamkar, N. Nouri, *et al.*, “Quantifying the Effects of Working in VR for One Week,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 11, pp. 3810–3820, Nov. 2022, Conference Name: IEEE Transactions on Visualization and Computer Graphics, ISSN: 1941-0506. DOI: 10.1109/TVCG.2022.3203103. [Online]. Available: https://ieeexplore.ieee.org/document/9872089/?arnumber=9872089&tag=1 (visited on 01/29/2025).

[93] *Vuzix Z100 Smart Glasses*, en. [Online]. Available: https://www.vuzix.com/products/z100-smart-glasses (visited on 12/17/2024).

[94] D. Weber, A. Voit, and N. Henze, “Notification Log: An Open-Source Framework for Notification Research on Mobile Devices,” en, in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, Singapore Singapore: ACM, Oct. 2018, pp. 1271–1278, ISBN: 978-1-4503-5966-5. DOI: 10.1145/3267305.3274118. [Online]. Available: https://dl.acm.org/doi/10.1145/3267305.3274118 (visited on 01/22/2025).

[95] V. Braun and V. Clarke, “Using thematic analysis in psychology,” en, *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, Jan. 2006, ISSN: 1478-0887, 1478-0895. DOI: 10.1191/1478088706qp063oa. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa (visited on 03/17/2022).

[96] A. Lucero, “Using Affinity Diagrams to Evaluate Interactive Prototypes,” en, in *Human-Computer Interaction – INTERACT 2015*, J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, and M. Winckler, Eds., vol. 9297, Series Title: Lecture Notes in Computer Science, Cham: Springer International Publishing, 2015, pp. 231–248, ISBN: 978-3-319-22667-5 978-3-319-22668-2. DOI: 10.1007/978-3-319-22668-2_19. [Online]. Available: http://link.springer.com/10.1007/978-3-319-22668-2_19 (visited on 02/11/2025).

[97] T. Iba, A. Yoshikawa, and K. Munakata, “Philosophy and Methodology of Clustering in Pattern Mining: Japanese Anthropologist Jiro Kawakita’s KJ Method,” en,

[98] H. Beyer and K. Holtzblatt, “Contextual design,” en, *Interactions*, vol. 6, no. 1, pp. 32–42, Jan. 1999, ISSN: 1072-5520, 1558-3449. DOI: 10.1145/291224.291229. [Online]. Available: https://dl.acm.org/doi/10.1145/291224.291229 (visited on 04/30/2024).

[99] A. Haskins Lisle, C. Merenda, and J. Gabbard, “Using affinity diagramming to generate a codebook: A case study on young military veterans and community reintegration,” en, *Qualitative Research*, vol. 20, no. 4, pp. 396–413, Aug. 2020, ISSN: 1468-7941, 1741-3109. DOI: 10.1177/1468794119851324. [Online]. Available: https://journals.sagepub.com/doi/10.1177/1468794119851324 (visited on 02/11/2025).

[100] S. Gandhi, P. v. Platen, and A. M. Rush, *Distil-Whisper: Robust Knowledge Distillation via Large-Scale Pseudo Labelling*, arXiv:2311.00430 [cs], Nov. 2023. DOI: 10.48550/arXiv.2311.00430. [Online]. Available: http://arxiv.org/abs/2311.00430 (visited on 02/10/2025).

[101] F. M. Götz, S. Stieger, and U.-D. Reips, "Users of the main smartphone operating systems (iOS, Android) differ only little in personality," en, *PLOS ONE*, vol. 12, no. 5, S. Jiménez-Murcia, Ed., e0176921, May 2017, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0176921. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0176921 (visited on 02/20/2025).

[102] S. Kemp, *The Rise of Smartwatches*, en-GB, Jan. 2023. [Online]. Available: https://datareportal.com/reports/digital-2023-deep-dive-the-rise-of-wearables (visited on 03/18/2025).

[103] Google, *Create and manage notification channels | Views*, en, Feb. 2024. [Online]. Available: https://developer.android.com/develop/ui/views/notifications/channels (visited on 03/18/2025).

[104] K. S. Oberoi, *Samsung's One UI 7 will tame your notification panel with a new filtering system*, en, Section: Operating Systems, Dec. 2024. [Online]. Available: https://www.androidpolice.com/samsung-one-ui-7-new-notification-filter-system/ (visited on 03/18/2025).

[105] Apple, *Summarize notifications and reduce interruptions with Apple Intelligence on iPhone*, en, 2024. [Online]. Available: https://support.apple.com/guide/iphone/summarize-notifications-reduce-interruptions-iph1fbe7d2b9/ios (visited on 03/18/2025).

[106] TVC, *The Vision Council Overview*, 2021. [Online]. Available: https://thevisioncouncil.org/sites/default/files/assets/media/TVC_OrgOverview_sheet_2021.pdf (visited on 03/18/2025).

[107] K. L. Bahnsen, L. Tiemann, L. Plabst, and T. Grundgeiger, "Augmented Reality Cues Facilitate Task Resumption after Interruptions in Computer-Based and Physical Tasks," en,

in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, May 2024, pp. 1–16, ISBN: 9798400703300. DOI: 10.1145/3613904.3642666. [Online]. Available: https://dl.acm.org/doi/10.1145/3613904.3642666 (visited on 07/18/2024).

[108] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht, "Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality," en, *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 6, pp. 1706–1724, Jun. 2017, ISSN: 1077-2626. DOI: 10.1109/TVCG.2016.2543720. [Online]. Available: http://ieeexplore.ieee.org/document/7435333/ (visited on 07/04/2021).

[109] Doublepoint, *Doublepoint - Gesture Recognition for Wearables*, en, 2025. [Online]. Available: https://doublepoint.com (visited on 03/18/2025).

Appendix A

Picture Permissions

A.1 Figure 1.1

1/15/25, 11:38 AM Mail - Plabst,Lucas - Outlook

Outlook

Re: Usage of Images in Doctoral Thesis

From Keiichi Matsuda <info@km.cx>
Date Wed 1/15/2025 11:37 AM
To Plabst,Lucas <Lucas.Plabst@colostate.edu>

**** Caution: EXTERNAL Sender ****

Hi Lucas, thanks for your message and kind words.

Yes of course you are welcome to pull stills from the film. I also have some high res ones here:
https://drive.google.com/drive/folders/1w36xrWGQHBX77nc9-JjM6PkBXCIBlClT?usp=sharing

Best of luck with your work

Keiichi

------ Original Message ------
From "Plabst,Lucas" <Lucas.Plabst@colostate.edu>
To "info@km.cx" <info@km.cx>
Date 15/01/2025 17:42:03
Subject Usage of Images in Doctoral Thesis

> Dear Mr. Matsuda,
>
> My name is Lucas Plabst, and I am a PhD student in Human Computer Interaction at the Colorado State University in Fort Collins USA and at the University of Würzburg in Germany.
> Your film Hyperreality has been incredibly influential on me, and it is a major reason why I chose to dedicate my PhD research to the topic of "Notifications in Augmented Reality".
> I see a future where smartglasses will be ubiquitous, like smartwatches are now, and to avoid having it play out like in your film, I did research on the User Experience and Usability of these notifications in AR.
> Currently I have published three papers on this topic and am working on the fourth.
> I am also finishing my doctoral thesis and wanted to ask for your permission, to use still-frames from your movie in the thesis to highlight the potential problems that could arise from information overload in AR.
> These images would of course be properly credited in the publication of the thesis.
> Thank you for your consideration and thank you for making the art you make.
>
> Best wishes
> Lucas Plabst
> --------
> **Lucas Plabst | PhD Student (Cotutelle)**
> M.Sc. Human Computer Interaction
> NUILab CSU USA &
> HCI Group JMU Germany
> Check out my work!

https://outlook.office.com/mail/inbox/id/AAQkADI4NmYzY2NiLTNkYmMtNGYzMy05NzNlLTA4Yzc0YzM1MWUzMAAQAF3Ow1y6svFKpn%2FKlKc… 1/1

A.2 Figure 4.6

TUESDAY

Lucas Plabst (He/Him) • 12:19 PM

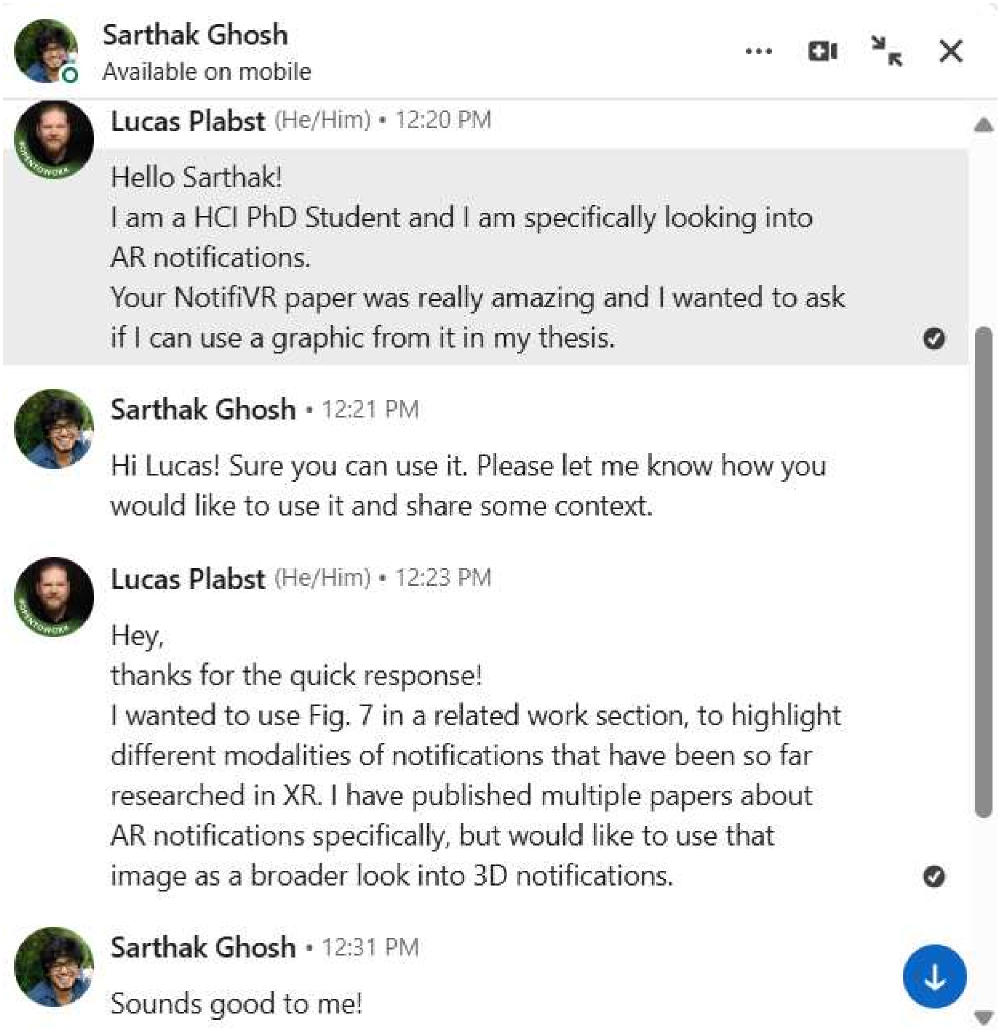
Hello Nuwan!
I am finishing my doctoral thesis about AR notifications, and for my related work part, I wanted to use a figure from your GlassMessaging paper. Is it okay for me to use Fig 1, of course with proper credit? Also, will you be at IEEE VR? I would love to talk to you more about XR notifications, you have done so much great work!

Nuwan Janaka (He/Him) • 9:31 PM

Sure, no problem. Go ahead, thanks

Almost all my work should be under creative Commons (no restrictions).

A.3 Figure 4.1


Sarthak Ghosh
Available on mobile
Lucas Plabst (He/Him) • 12:20 PM
Hello Sarthak!
I am a HCI PhD Student and I am specifically looking into AR notifications.
Your NotifiVR paper was really amazing and I wanted to ask if I can use a graphic from it in my thesis.
Sarthak Ghosh • 12:21 PM
Hi Lucas! Sure you can use it. Please let me know how you would like to use it and share some context.
Lucas Plabst (He/Him) • 12:23 PM
Hey,
thanks for the quick response!
I wanted to use Fig. 7 in a related work section, to highlight different modalities of notifications that have been so far researched in XR. I have published multiple papers about AR notifications specifically, but would like to use that image as a broader look into 3D notifications.
Sarthak Ghosh • 12:31 PM
Sounds good to me!

A.4 Figure 4.4

2/25/25, 9:52 AM Mail - Plabst,Lucas - Outlook

Outlook

RE: Permission to use Figure

From 이현진 <clairehj517@kaist.ac.kr>
Date Wed 1/22/2025 6:55 AM
To Plabst,Lucas <Lucas.Plabst@colostate.edu>

**** Caution: EXTERNAL Sender ****

Hello, Lucas!

It's so great to hear from you! I remember meeting you at ISMAR 2022 in Singapore—it was great chatting with you back then, and I'm really glad to reconnect.

Thank you for your kind words about the paper. I'm happy to hear that it's been helpful for your research. Of course, you're welcome to use Figure 1 in your doctoral thesis. I really appreciate you making sure to disclose and properly cite the source.

Your work in this area is incredibly impressive, and I hope we get a chance to collaborate on something in the future. Best of luck with finishing your thesis, and feel free to reach out if there's anything else I can help with!

Best regards,

Hyunjin.

Hyunjin Lee (Ph.D)
Senior Researcher, KAIST KI-ITC Augmented Reality Research Center (ARRC)
291, Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea

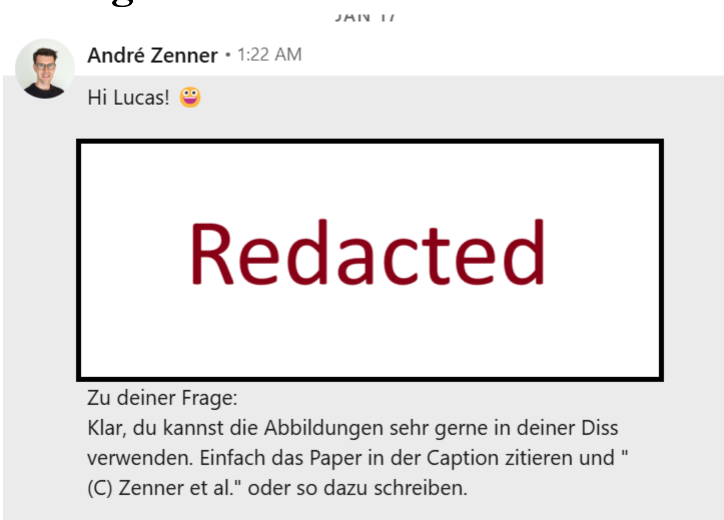
-----Original Message-----
From: "Plabst,Lucas" <lucas.plabst@colostate.edu>
To: "clairehj517@kaist.ac.kr" <clairehj517@kaist.ac.kr>;
Cc:
Sent: 2025-01-22 (수) 04:27:16 (UTC+09:00)
Subject: Permission to use Figure

Hello,
I am a PhD student at Colorado State University, and I am currently finishing my doctoral thesis about XR notifications.
Your paper "Exploring the Effects of Augmented Reality Notification Type and Placement in AR HMD while Walking" was really good work, and I wanted to ask for your permission to use Figure 1 in that paper in the related work section of my thesis.
Of course this would be disclosed and cited in the thesis.
Thank you in advance and best wishes
Lucas

Lucas Plabst | PhD Student (Cotutelle)
M.Sc. Human Computer Interaction

https://outlook.office.com/mail/sentitems/id/AAQkADI4NmYzY2NiLTNkYmMtNGYzMy05NzNlLTA4Yzc0YzM1MWUzMAAQALjSj6rFtiRAjoWApafor… 1/2

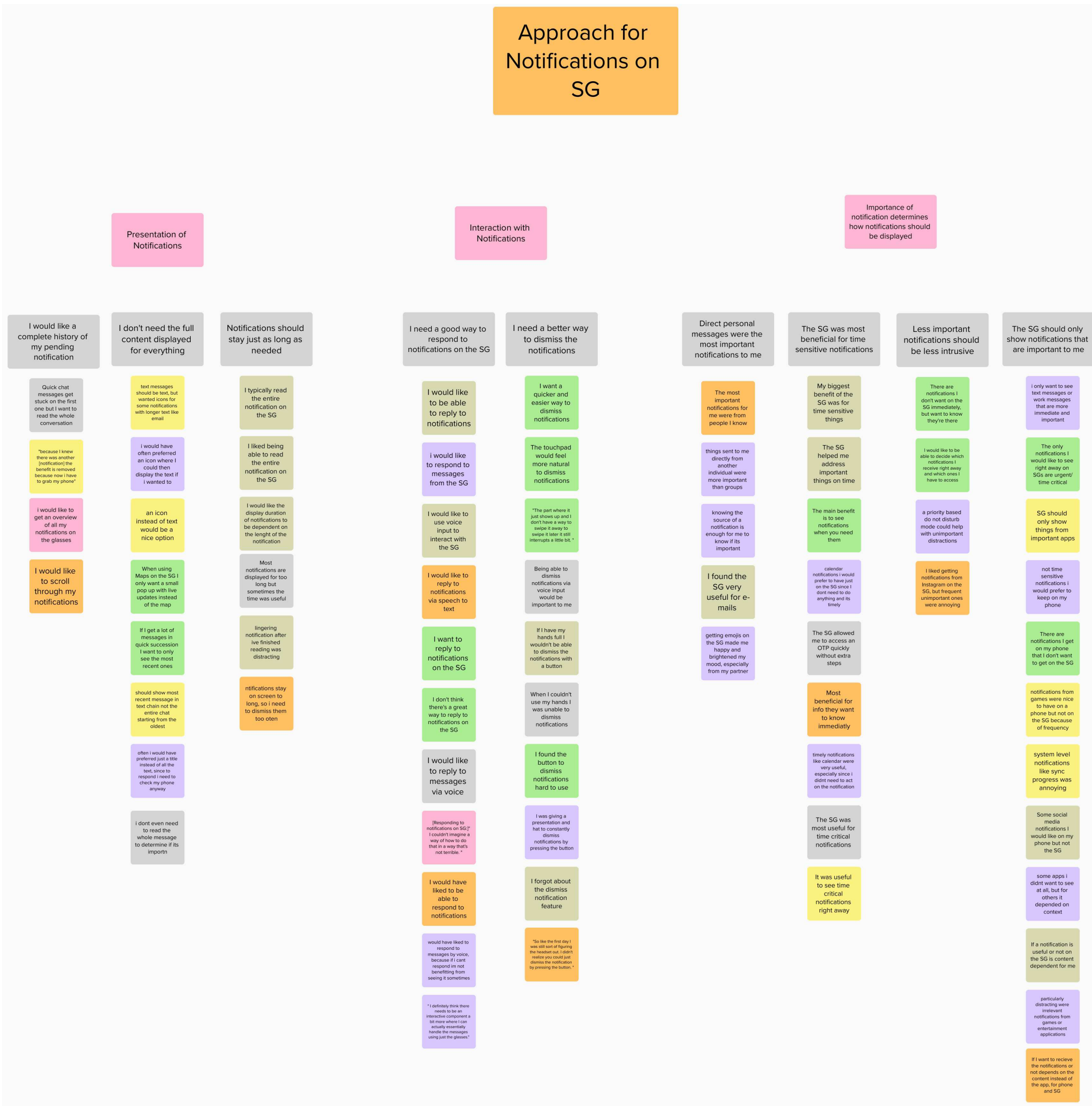
A.5 Figure 4.2


JAN 17
André Zenner • 1:22 AM
Hi Lucas!
Redacted
Zu deiner Frage:
Klar, du kannst die Abbildungen sehr gerne in deiner Diss verwenden. Einfach das Paper in der Caption zitieren und "(C) Zenner et al." oder so dazu schreiben.

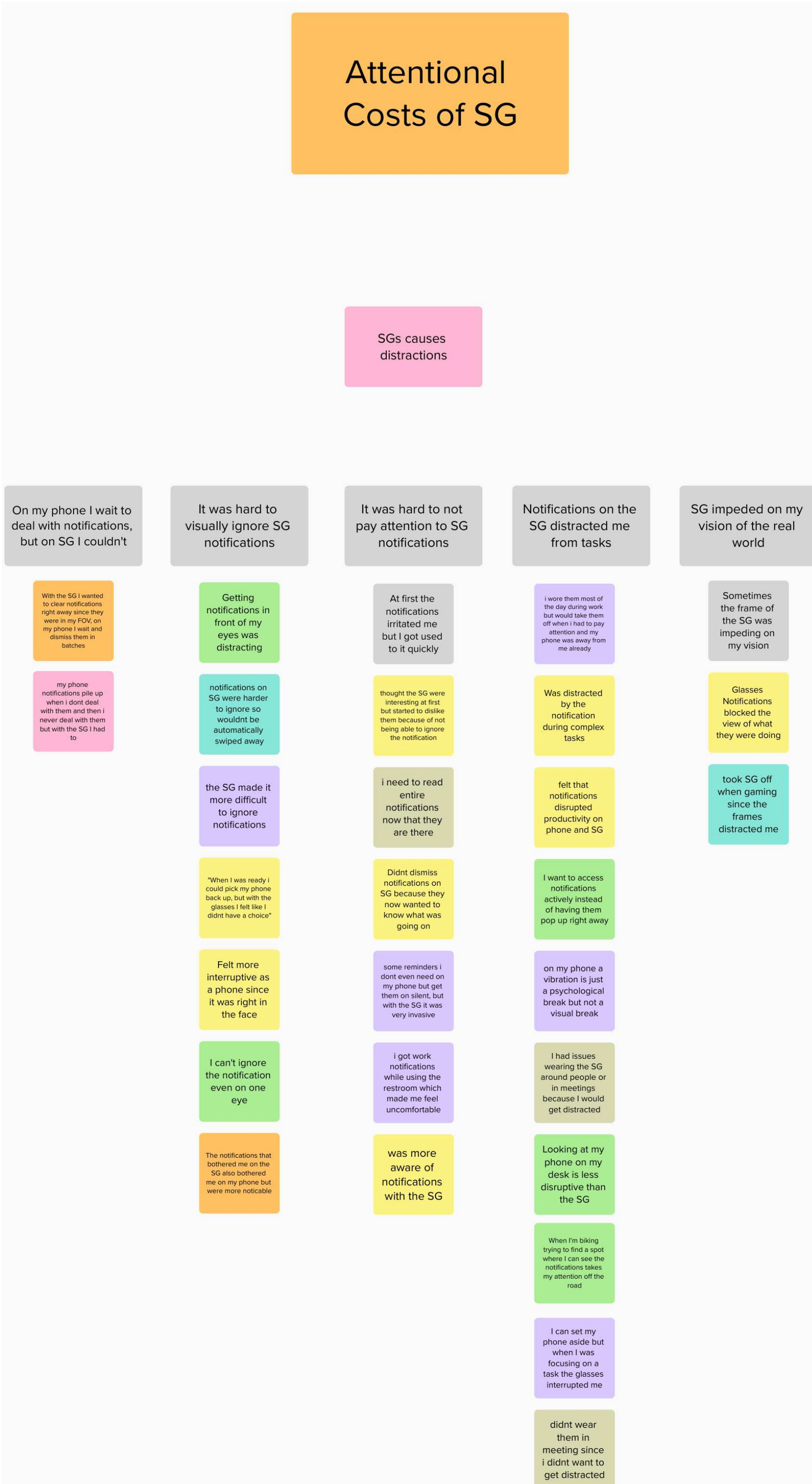
Appendix B

Affinity Diagram for Smartglasses Experiment

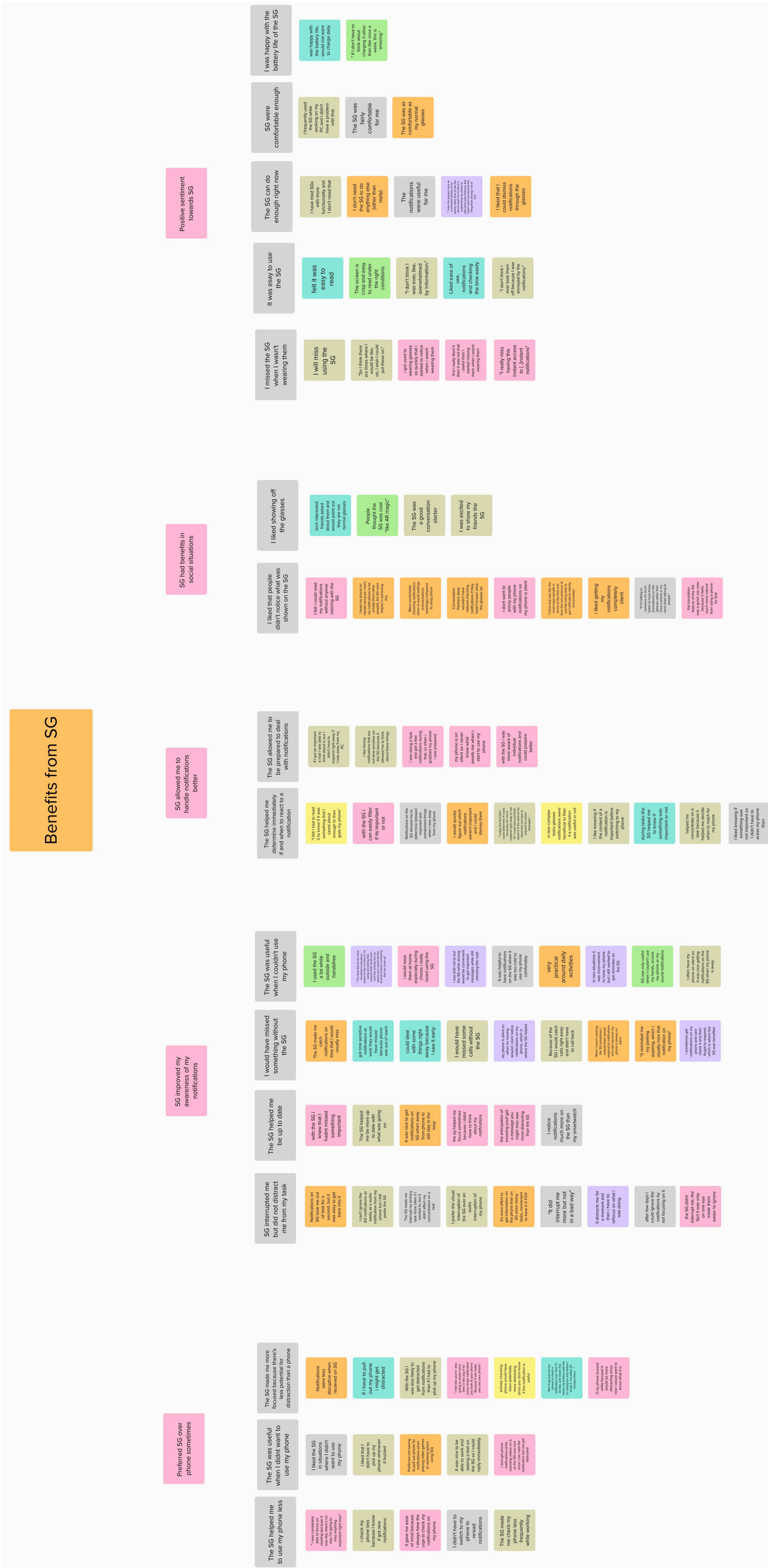
B.1 Approach for Notifications



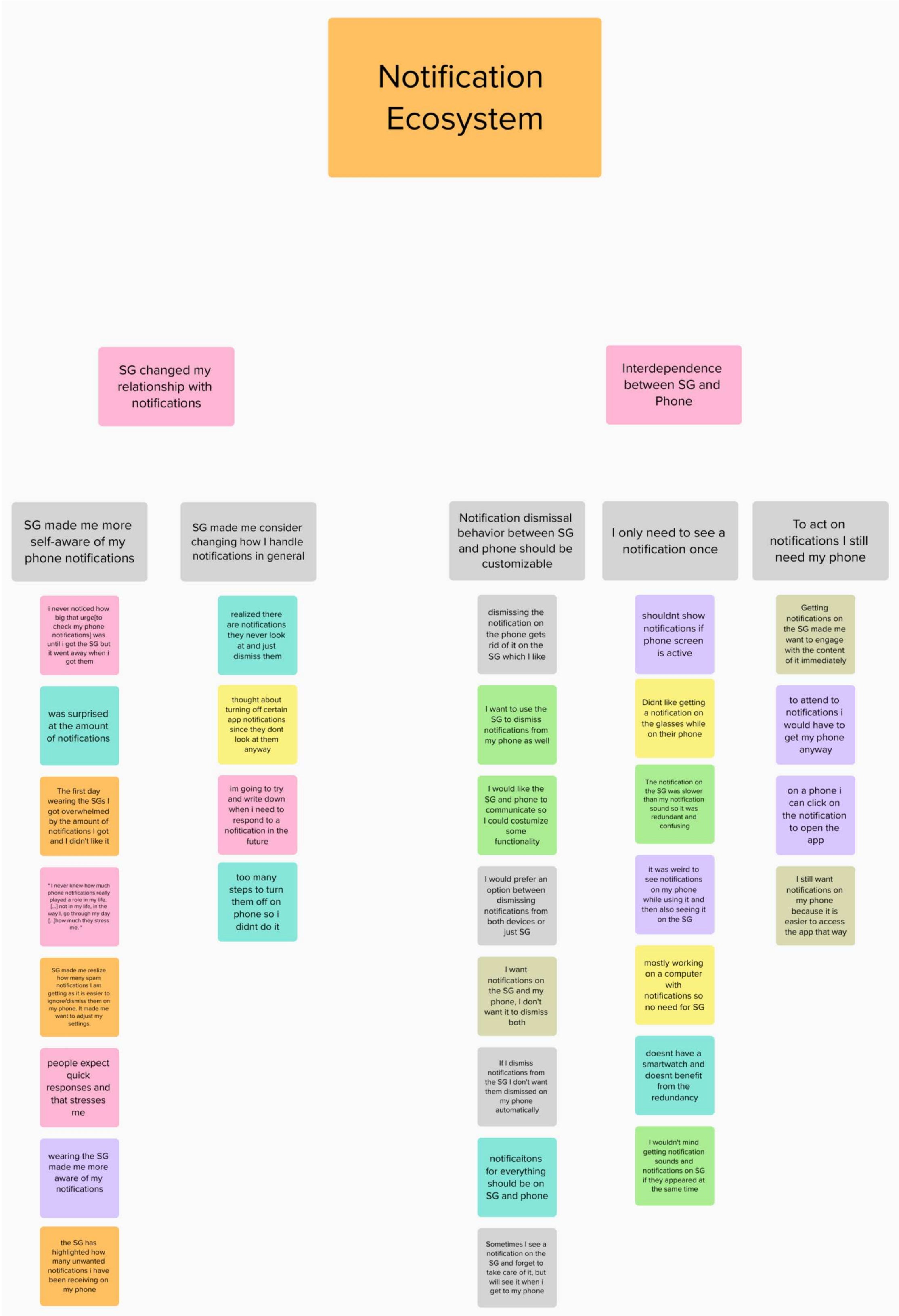
B.2 Attentional Costs



B.3 Benefits of Notifications

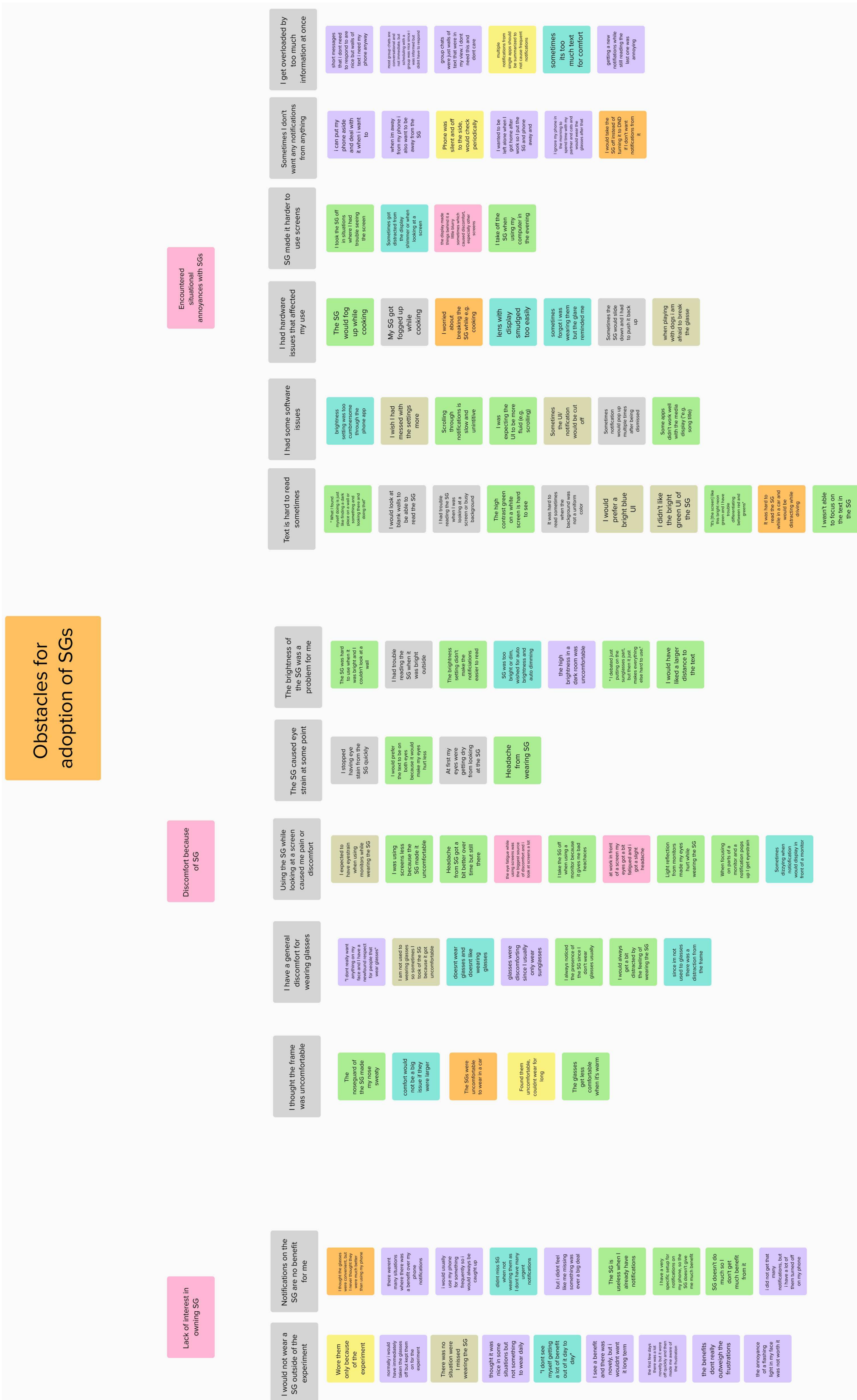


B.4 Notification Ecosystem



B.5 Improvement Opportunities

B.6 Obstacles for Adoption



B.7 Societal Challenges