

# 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.

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

\*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

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 gen-

eral 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 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

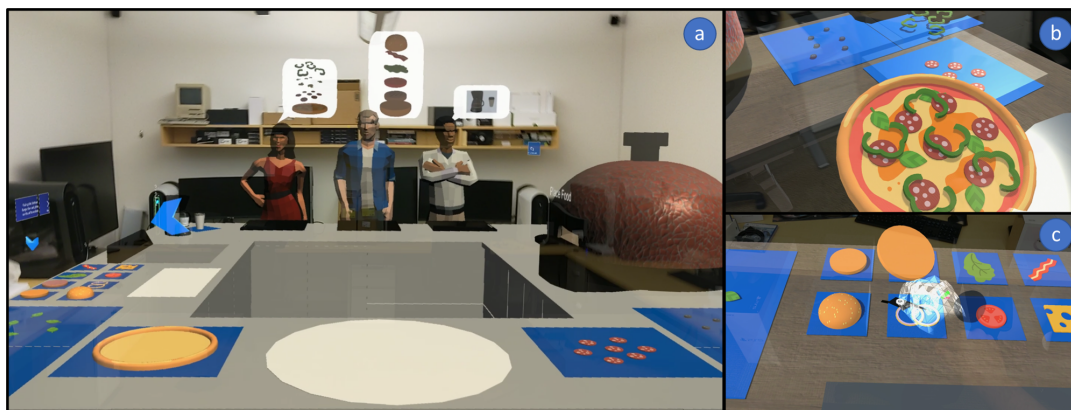


Figure 2: (a) Complete Setup of ARTisan Bistro; (b) Perfectly Cooked Pizza Taken out of the Oven; (c) Making Burger, Picking up Bottom Bun by Hand [30]

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



Figure 5: Notification on Dock

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.

## 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 cus-

tomers (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 ( $M = 7.92$ ,  $SD = 1.61$ ).

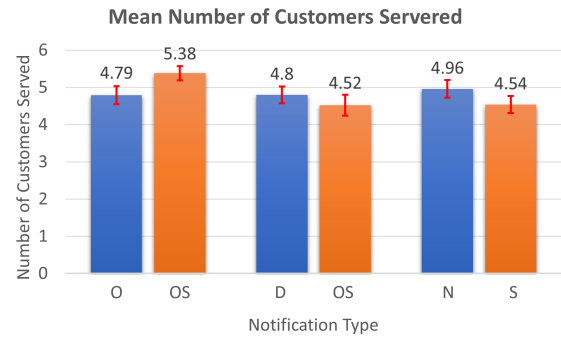


Figure 6: Performance of participants in different notification conditions based on how many customers they served.

Table 2: Mean performance of participants in different notification conditions

Notification Techniques	Mean	STD	SE
<i>O</i>	4.79	1.24	0.24
<i>OS</i>	5.38	0.96	0.19
<i>D</i>	4.8	1.14	0.22
<i>DS</i>	4.52	1.44	0.28
<i>N</i>	4.96	1.19	0.23
<i>S</i>	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 condi-

tions 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 ( $M = 4.1, SD = 2.71$ ).

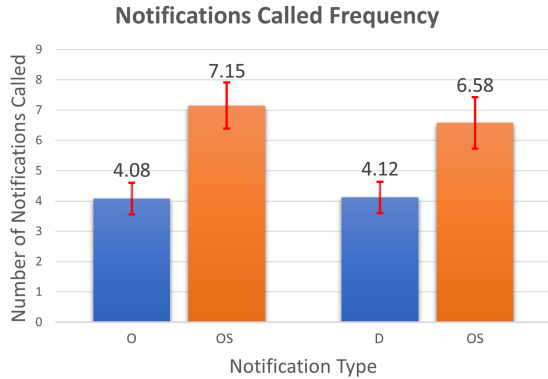


Figure 7: Average number of notifications called in different notification conditions

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

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

### 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 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.

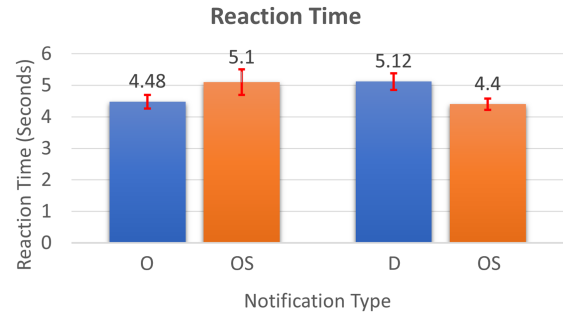


Figure 8: Average reaction time to notice the notification in different notification conditions in seconds

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

Notification Techniques	Mean	STD	SE
<i>O</i>	4.48	1.1	0.21
<i>OS</i>	5.1	2.07	0.41
<i>D</i>	5.12	1.34	0.26
<i>DS</i>	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>. 4
- [2] Apple. Apple notifications. <https://support.apple.com/en-us/108781>. 4
- [3] M. Cidota, S. Lukosch, D. Dacu, 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. 2
- [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, 5
- [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 6
- [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 6
- [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. 1
- [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 6
- [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, 2
- [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, 5
- [14] M. G. Kenward and J. H. Roger. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, pp. 983–997, 1997. 6
- [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, 8
- [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, 5
- [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, 5
- [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 7
- [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/NuilLab/Artisan-Bistro>. 3
- [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, 7
- [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
- [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. 2
- [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. 1, 5
- [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 6
- [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. 2, 3, 4