# Exploring Unimodal Notification Interaction and Display Methods in Augmented Reality

Lucas Plabst
HCI Group, University of Würzburg
Würzburg, Germany
lucas.plabst@uni-wuerzburg.de

Aditya Raikwar
Computer Science & NUILAB,
Colorado State University
Fort Collins, CO, USA
aditya.raikwar@gmail.com

Sebastian Oberdörfer
HCI Group, University of Würzburg
Würzburg, Germany
sebastian.oberdoerfer@uni-wuerzburg.de

Francisco Ortega
Computer Science & NUILAB,
Colorado State University
Fort Collins, CO, USA
fortega@colostate.edu

Florian Niebling
Advanced Media Institute, TH Köln
Köln, Germany
florian.niebling@th-koeln.de

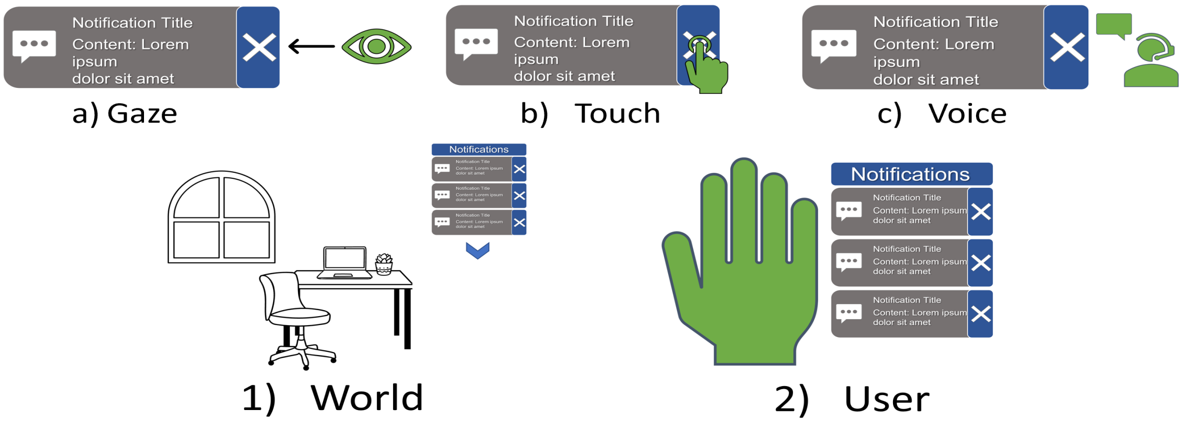

**Figure 1: Techniques for interaction notification: a) Gaze, b) Touch, c) Voice Display methods for notification lists: 1) World 2) User**

## ABSTRACT

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.

## KEYWORDS

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 09–11, 2023, Christchurch, New Zealand*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3611659.3615683

---

*VRST 2023, October 09–11, 2023, Christchurch, New Zealand*

ACM ISBN 979-8-4007-0328-7/23/10…$15.00
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 [12]. 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 [38].

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 [21]. 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 [50]. Multimodal presentation of notifications in AR has been explored by Lazaro et al. [30], with the premise that notifications in AR systems must be designed so theycapture 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 [50].

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 [23]. Research suggests that individuals receive an average of around 80 notifications per day, with some receiving even higher volumes, reaching up to 200 [1]. Smartphones dominate internet access, contributing to over 60% of internet traffic [37]. 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. [46] 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 [28].

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 [44]. Grouping notifications into smaller batches and delivering them multiple times throughout the day has been proposed as a strategy to manage their impact [13]. 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 [42].

## 2.2 Notifications in 3D/AR

AR has been defined as the supplementation of a real-world environment with digital content [7]. 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. [9], 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. [50] 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. [31] 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. [15] 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. [36] 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 [14]; 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 [41] and have shown to be quick for users to learn and engage with [51].

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 [6]. 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. [49] 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 [22] and has been studied by various researchers. Blattgerste et. al. [10] 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. [43] 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. [27] 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. [35] 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. [33] 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. [32], 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 [47]. 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. [26] 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 [48].

## 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.

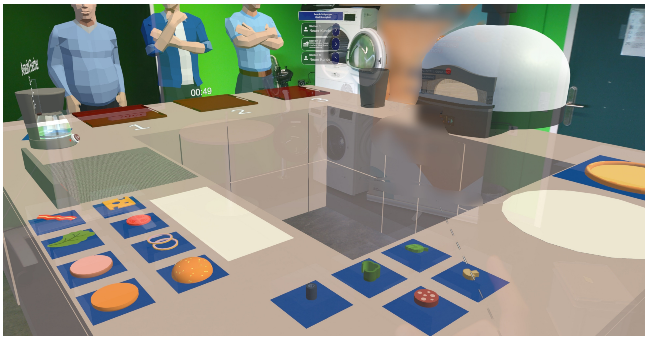

**Figure 2: The virtual cooking environment used in the experiment.**

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

**Figure 3: Notification in the experiment.**

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 [30].

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. [25] and the design recommendations by

Microsoft [39]. 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. [50] 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. [11] 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.

When using HMDs, special attention needs to be paid to the vergence-accommodation-conflict [20]. The Hololens 2 display is fixed at an optical distance of approximately two meters and Microsoft recommends not placing any information closer than 40cm [40]. 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 [2, 3], 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 [45]. 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 [5, 29].

*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 [4, 16]. 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 [9] 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 4) 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.

**Figure 4: The list of notifications fixed to the left hand.**

The **world list** 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.

**Figure 5: The list of notifications located in the environment.**

## 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 [52]. 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 [17] was deployed, along with a NASA Task-Load-Index questionnaire (NASA-TLX) [19] used for assessing task load. Hart et. al.[18] 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.

## 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 [8]. 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 [24] 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).

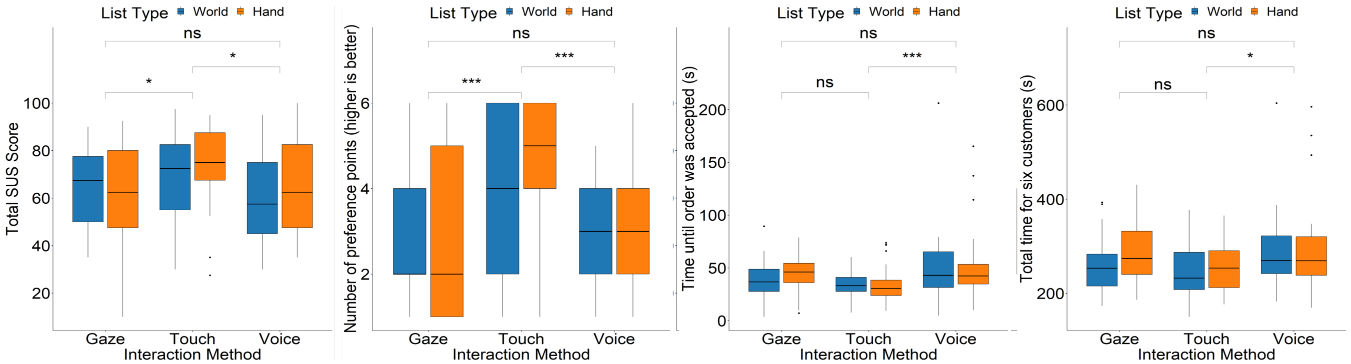

**Figure 6: Results for the SUS-Score, Preference ranking, Time until orders were accepted and Total Time.
***: p < 0.001; **: p < 0.01; *: p < 0.05; ns: p > 0.05**

**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 |

*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 reassignment 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. [34], 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. [26] 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] 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.

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

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

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

[5] 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

[6] 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

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

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

[9] 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

[10] 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

[11] 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

[12] 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.

[13] 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

[14] 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.

[15] 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.

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

[17] 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.

[18] 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.

[19] 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.

[20] 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

[21] 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

[22] 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

[23] 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

[24] 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

[25] 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.

[26] 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

[27] 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

[28] 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

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

[30] 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

[31] 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.

[32] 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

[33] 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

[34] 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

[35] 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

[36] 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

[37] 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.

[38] 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.

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

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

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

[42] 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.

[43] 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

[44] 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

[45] 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

[46] 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

[47] 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

[48] 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

[49] 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

[50] 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

[51] 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

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

[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