

Evaluating Dynamic Delivery of Audio+Visual Message Notifications in XR

Hyunsung Cho*
Carnegie Mellon University

Drew Edgar†
University of Glasgow

David Lindlbauer‡
Carnegie Mellon University

Joseph O'Hagan§
University of Glasgow

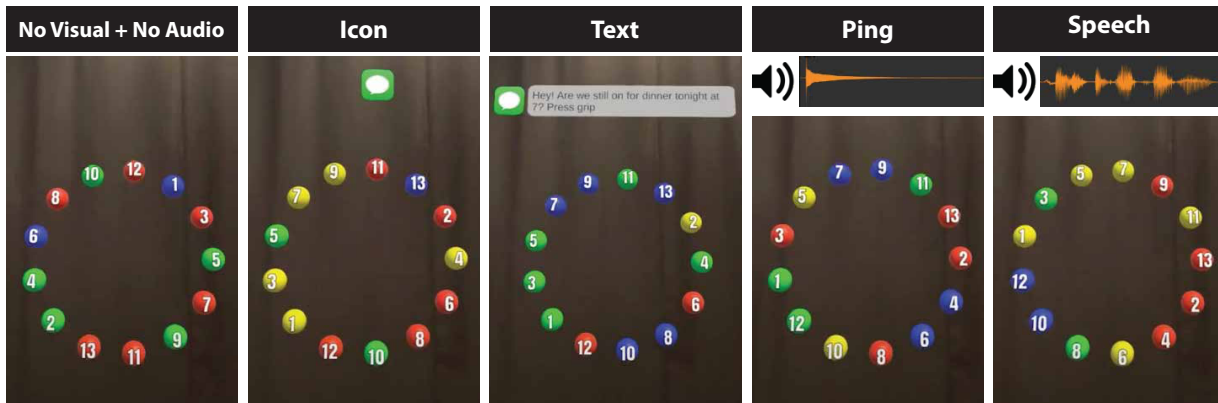


Figure 1: User evaluation setup showing Fitts' target selection task with our visual and audio notification components (detailed in Section 3.1), implemented in our Augmented Reality prototype using Unity passthrough.

ABSTRACT

The spatial flexibility of Extended Reality (XR) allows for personalized, context-aware organization of applications aligned with a user's tasks and priorities. Notifications play a crucial role here, e.g., informing users of received messages they might otherwise miss. However, questions remain around how attention-grabbing they should be, how much information they should present, and how the presentation should adapt to the message's context and content. While prior studies examined facets of message notification design, the impact of *multimodal* notifications and how they could be used holistically to support message awareness has not yet been explored. We address this by evaluating nine audio-visual notifications, investigating usability, interruptibility, preferences, and their use to inform of received messages. Our results show differing effects of multimodal notification designs and that individuals want notification modality and design to vary based on delivered message content. These results offer new insights into developing context-aware multimodal interaction strategies for spatial notifications and XR messaging.

Index Terms: Human-centered computing — Human computer interaction (HCI) — Empirical studies in HCI; Human-centered computing — Human computer interaction (HCI) — Interaction paradigms — Mixed / augmented reality

1 INTRODUCTION

Extended Reality (XR) allows for personalized, context-aware organization of applications, interfaces, and interactions aligned with a user's tasks and priorities [9]. Applications and virtual elements can be adapted based on users' current context [31], such as user's state and external stimuli [5, 43]. Notifications play a crucial role

in spatial and immersive XR environments. They inform users of timely information, events, or received messages that they might otherwise miss (e.g., interpersonal messaging [52], presence of bystanders [41] or other real-world events, feedback for gaze, gestural, or speech commands [7, 29], navigation [27], and guidance to XR objects [8]).

Similar to general XR content, XR notifications must be informative while being only as disruptive as necessary to users [41, 43]. To achieve this, prior work has investigated a range of different design parameters of XR notifications, including their positioning [48], aesthetics [25, 51], and informativeness [24]. However, dynamic XR environments further require adaptable, non-intrusive notifications that can seamlessly integrate into various contexts, which demands understanding how the impact of notification design might change relative to varying personal and contextual factors. As prior work focuses mainly on visual XR notifications, we also lack knowledge of the usability and preferences of multimodal XR notifications, despite the importance of auditory or haptic signal in timely information delivery.

In this work, we contribute to this understanding by investigating how different XR audiovisual notification designs affect usability and preferences in XR environments, and how these change in varying contextual factors of the message content such as urgency and task engagement. We conducted a user study (N=18) which evaluated the design of nine multimodal (audio+visual) notifications for informing an AR user of a delivered message. Notifications were delivered with different visual conditions (*No Visual*, *Icon*, and *Text*), paired with different auditory presentation styles (*No Audio*, *Ping*, and *Speech*) as illustrated in Figure 1. Participants were asked to complete a Fitts' reciprocal selection primary task with a speak-aloud component while reacting to notifications as secondary task.

We investigated the influence of notification modality on aspects such as task performance, usability, interruptibility, impact on message acknowledgement, and subjective preferences. Our results show differing effects between our nine notification designs including differences in their noticeability and their informativeness among other things. Qualitative feedback also indicates that users want the modality and design of their message notifications to vary based on the received message's content. For example, users want

*e-mail: hyunsung@cs.cmu.edu

†e-mail: 2647430e@student.gla.ac.uk

‡e-mail: davidlindlbauer@cmu.edu

§e-mail: Joseph.OHagan@glasgow.ac.uk

more disruptive notifications for messages of high urgency (e.g. a time sensitive message) but often want less disruptive, “missable”, notifications for low urgency messages. Our findings inform the development of context-aware multimodal interaction strategies for spatial notifications and XR messaging.

2 RELATED WORK

2.1 XR Messaging and Message Awareness

Mobile instant messaging applications allow users to send text, photo, video, and voice messages to others. These applications have become essential to modern smartphones, enabling users to communicate and stay connected with others [21, 56]. Users are known to respond, on average, within minutes of new messages being delivered to them [47]. As XR technologies evolve into ubiquitous, wearable devices [40, 44, 39], mobile instant messaging is expected to become a core feature of XR platforms as well [28, 51].

Unlike smartphones, which have fixed positions for notifications (such as a status bar) [53], XR’s spatial flexibility offers no predetermined location for message windows. XR users prefer to position their message windows based on personal preferences and the current task [9], prompting research into designing effective notifications in these dynamic environments [48, 52]. Researchers have started exploring various notification design approaches, including heads-up displays and floating pop-ups, to inform users of incoming messages, as discussed in Section 2.2.

When to notify users of a new message is of similar importance as *where* to present it. On smartphones, users are typically alerted immediately when a message arrives [53], but prior studies have shown that users may prefer to delay notifications based on message content, only receiving alerts for urgent or important messages [13, 14, 45]. While dynamic delivery systems are rare in smartphones, some systems, like Apple’s CarPlay, have started adapting notifications contextually [1], transforming a text-based messaging application into an audio-only interface. Given XR’s inherently dynamic nature [9, 41, 43], future XR messaging systems could leverage this flexibility by contextually adjusting both the timing and presentation of notifications based on user context and awareness. While previous studies have shown varying the timing of VR message notifications can enhance user experience [5], little work has explored how the notification presentation itself might vary to better support users’ message awareness needs in XR.

2.2 XR Message Delivery Notifications

Extensive research has examined different aspects of XR notifications design, particularly the placement of visual notifications to ensure that they are seen by users [20, 22, 26, 48, 51, 52]. These works have led to design guidelines suggesting that low-urgency notifications should be placed in-situ within the environment, while higher-urgency notifications should float in front of the user for greater visibility [48, 51].

Other studies have focused on the modality of notifications. Ghosh et al. [17] found haptic notifications were the least effective among visual, aural, and haptic designs, and often mistaken for application feedback, a result echoed by other researchers [16, 49]. The aesthetics of notification animations [23] and the relative effectiveness of icons versus text pop-ups [24] have also been explored, showing that the effectiveness of icons depends on user familiarity, encoding density, and the visual complexity of the surrounding environment. These studies offer recommendations for optimizing notification design to preserve immersion, improve noticeability, and enhance user experience [32, 34, 46].

However, prior work has primarily focused on evaluating individual notification designs without considering how multiple designs could be used together to meet a user’s varying message awareness needs. Working toward this goal, Zenner et al. proposed a framework that adjusts notification priority (low, medium, high)

based on the sender’s importance, modifying the disruptiveness of a visual pop-up [62]. While their work offers a valuable starting point, the overall system was not evaluated, nor did it explore how the modality or design of notifications could change based on message priority. Our study extends this research by evaluating how users interact with different multimodal notification designs and investigating how they choose from a range of notification designs to be informed of messages with differing levels of urgency. Our work provides a more comprehensive understanding of how XR notifications can be adapted to user preferences and dynamic task environments.

3 MESSAGE DELIVERY SYSTEM & NOTIFICATIONS

Our message delivery system was setup as follows. As illustrated in Figure 2, a floating message window was presented on participants’ left side. The window was populated with messages and positioned to be out of view while participants performed the experimental task. We explore nine audio+visual multimodal notifications to signal a new message being delivered. When a new message was delivered, it was added to the message window. The design of the notifications and the message delivery system is detailed below.

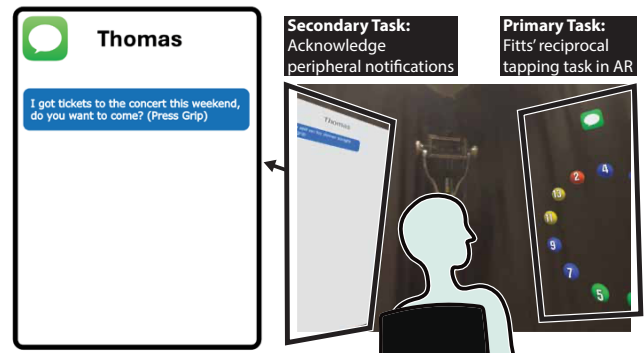


Figure 2: Our message window design styled after existing mobile instant messaging applications, positioned at a 90° angle to the left of the primary experimental task.

3.1 Audio+Visual Notification Conditions

We chose three audio (*No Audio*, *Ping*, *Speech*) and three visual approaches (*No Visual*, *Icon*, *Text*) to notify a user of a new message’s delivery. Our designs varied by the extent and detail to which they inform of the received message’s content. Our designs covered existing notifications as seen in consumer XR devices (e.g. [2, 36]) and designs explored in prior research (e.g. [25, 48]). We investigated the combined use of our visual and audio approaches, resulting in nine conditions summarised in Table 1.

Our three audio approaches were: **No Audio**: no auditory notification of message delivery. **Ping**: a one-off, notification sound effect played. The iPhone ping [4] was used as the *Ping* sound. **Speech**: audio of the message text being read. Google’s female voice for text-to-speech [19] was used for *Speech*.

Our three visual approaches were: **No Visual**: no visual notification of message delivery. **Icon**: a graphic of a messaging app icon appeared. The iMessage icon [3] was used as the *Icon*. **Text**: a pop-up containing the full message text.

The positions of *Icon* and *Text*, as shown in Figure 1, was arranged to not overlap/interfere with the experimental task but be easily seen by our participants. This was based on the recommendation of prior works identifying this position as highly noticeable to users and well suited to notifications [51, 52]. We excluded haptic feedback as Ghosh et al. found this to be less effective for XR notifications when compared to visual and audio feedback [17].

<i>Condition</i>	<i>Description</i>	<i>Modality</i>	<i>Information Conveyed</i>
<i>No Visual-No Audio</i>	No notification signalling new message delivery.	None	None
<i>No Visual-Ping</i>	A one-off sound signals new message delivery. No visuals.	Audio	Existence of new message
<i>No Visual-Speech</i>	The message text is read aloud to the user. No visuals.	Audio	Existence and full content of new message
<i>Icon-No Audio</i>	An icon signals new message delivery. No audio.	Visual	Existence of new message
<i>Icon-Ping</i>	An icon and a one-off sound audio signals new message delivery.	Visual, Audio	Existence of new message
<i>Icon-Speech</i>	An icon signals new message delivery. The message text is read aloud to the user.	Visual, Audio	Existence and full content of new message
<i>Text-No Audio</i>	A pop-up appears with the full message text. No audio.	Visual	Existence and full content of new message
<i>Text-Ping</i>	A pop-up appears with the full message text. A one-off sound signals new message delivery.	Visual, Audio	Existence and full content of new message
<i>Text-Speech</i>	An pop-up appears with the full message text. The message text is read aloud to the user.	Visual, Audio	Existence and full content of new message

Table 1: A summary of the nine conditions investigated in our study. Notifications varied in the amount of information they conveyed, e.g. *Icon-Ping* informs a user only of a new message’s existence while *Text-Ping* informs of a new message’s existence and its full content.

3.2 Message Window

The message window design, illustrated in Figure 2, is modeled after existing mobile instant messaging interfaces commonly found in consumer XR platforms, such as Apple visionOS [2]. The message window was attributed to a fictional close friend named “Thomas,” and participants were instructed to act as if “Thomas” were indeed a close friend. The *Icon* was used in place of a profile picture for “Thomas” in both the message window and *Text*. Similar to standard mobile instant messaging interfaces, new messages appeared sequentially, one after another, within the message window.

The message window was positioned to the participants’ left, at a 90° angle from the experimental task. This position was informed by insights from Cho et al. [9] on the preferred positioning of XR messaging widgets relative to primary tasks—specifically, out of direct line of sight but easily viewable with a head turn. This arrangement ensured that participants had to actively engage with notifications, as they could not view both the experimental task and the message window simultaneously due to their spatial separation.

3.3 Messages’ Content and Their Delivery

When evaluating our notifications, each condition lasted 90 seconds. During this time, participants received four messages delivered according to a specific timing protocol. The 90-second condition was divided into four 22.5 second time slots. Each time slot was randomly allocated a message text to be delivered. A minimum gap of 15 seconds was maintained between message deliveries, meaning each message was delivered at a random time within the first 7.5 seconds of its time slot. For the first time slot, no message was delivered within the initial 5 seconds, allowing participants time to begin the task before receiving their first notification. The length of the condition and frequency of messages were determined based on previous studies evaluating XR notifications, e.g. [17, 43, 51].

The messages consisted of short, self-contained texts such as: “Hey! Are we still on for dinner tonight at 7?”, “Just saw the funniest video, I’ll send it to you later!”, “Don’t forget, we’re meeting at the park tomorrow morning for a jog!”, and “I got tickets to the concert this weekend, do you want to come?”. Each message ended with an acknowledgment instruction for participants (discussed in Section 4.2).

4 STUDY DESIGN

The notification type was the independent variable, with the nine combinations of visual (*No Visual*, *Icon*, and *Text*) and audio (*No Audio*, *Ping*, and *Speech*) being the study conditions (Table 1).

4.1 Primary Task: Fitts’ Reciprocal Tapping Task

We used a standard Fitts’ reciprocal tapping task [54] with a speak-aloud component as a primary task. Our tapping task had 13 circular targets arranged in a circle, labeled “1” to “13” (Figure 1), with the labels randomly assigned for every condition. Participants had to select targets from opposing sides of the circular arrangement as they moved round the circle in a clockwise direction in a standard Fitts’ selection task (ISO 9241-9). Participants were instructed which target to select through an audio cue stating the number of the next target to press. This audio cue triggered after each target was pressed regardless of whether the correct target was selected for the current target. The audio cue used Google’s male voice for text-to-speech [19]. Users selected each target by pointing a raycast line from the right hand controller and pressing the ‘A’ face button.

Each target was colored with one of four colors (red, green, yellow, or blue), randomized for every condition with at least one instance of each color. To incorporate a speech component, participants were instructed to say aloud the color of each target as they selected it. We added auditory engagement by requiring participants to listen to the next target number and speak aloud its color, simulating a scenario where both auditory and visual channels are occupied with the primary task to assess the combined interruptiveness. In our case, the speech component was “discourse” [12], simulating a cooperative, one-way conversation to deliver information from speaker to listener. We acknowledge the limitation of focusing only on this type of speech and suggest future work should examine notification delivery in other speech contexts, such as “dialogues”, “diatribes”, and “debates” [12].

4.2 Secondary Task: Acknowledging Peripheral Notifications

Participants were instructed to acknowledge new delivered messages by pressing a button (either the “trigger”, “grip”, “X face button”, or “Y face button”) on the left hand controller of the Meta Quest 3. The button was randomized for every delivered message, with the specific instruction appended to the end of the delivered message text. This approach required participants to actively engage with each message’s content to correctly acknowledge it. For notifications without full text, participants turned to the message window on their left periphery (Figure 2) to read this message, as they would in general in-situ notifications [52]. The messages conveyed the same levels of information across conditions.

In conditions involving a visual element, i.e. *Icon* and *Text*, acknowledging a message dismissed the notification and removed it from the scene. If a participant did not acknowledge a message

before the next one was delivered, any existing visual notification was replaced by the notification for the new message. For *Icon*, this meant dismissing the icon and spawning a new instance. For *Text*, this dismissed the current pop-up message and spawned a new instance with the new message text.

4.3 Measures

For each participant, we recorded notification events, including when notifications were created, acknowledged by the participant, or timed out if not acknowledged. We also logged each button press during the target selection task, noting whether the press hit a target, whether the correct target was selected, and the internal IDs, number, and color of the relevant targets.

4.4 Per Condition Questionnaire

To evaluate usability, we designed a questionnaire, based on the similar evaluations used in prior works, e.g., [43, 51, 52]. All questions were on a 5-point Likert scale.

We asked: “*How easy was it to notice the notification?*” and “*If you noticed the notification, how easy was it to understand what it meant?*” (1 = Extremely Difficult, 5 = Extremely Easy), “*If you noticed the notification, what level of urgency did the notification convey?*” (1 = Not At All Urgent, 5 = Extremely Urgent), “*How much of a hindrance was the notification to your experience?*” (1 = Not At All Hindering, 5 = Extremely Hindering).

Participants were also asked to what extent they agreed (1 = Strongly Disagree, 5 = Strongly Agree) that the notification: “*felt comfortable to use when alone*”, “*felt comfortable to use in a room with other people*”, “*disrupted my experience*”, “*provides a feeling of not missing out*”, “*provides the information I want*”, “*was useful*”, and “*felt natural*”.

4.5 Rank Ordering Questions

After evaluating each condition, participants were asked to rank each condition in terms of their general preference (best to worst), frustration (most to least frustrating), and noticeability (most to least noticeable). Participants were instructed to think aloud while doing this and justify their first and last-ranked choices.

4.6 Assessing Message Awareness Needs

While our other measures captured participants usability responses and preference towards our conditions, they did not address how, when, and why a user might prefer one method of message delivery over another given the many options available. Therefore, we concluded our study with a think-aloud task to explore these factors.

Participants were instructed to imagine they were at home using XR for work when a close friend sent them six messages. For each message, we captured which condition the participant preferred for receiving the message’s delivery notification. Participants were told to think aloud and justify their choices, and it was acceptable to choose the same notification for every message. Messages were presented in random order. The six messages included three time-sensitive ones of higher urgency than the others.

The three high-urgency messages were: **A&E (accident and emergency)**: “*I’ve had to go to A&E, can you come & pick me up please?*”, **House Keys**: “*I’m locked out of the house again - do you remember where the spare key is kept?*”, and **Can’t Meet**: “*I won’t be able to make it for 10AM tomorrow, can we do 2PM instead? Sorry!*”.

The three less urgent messages were: **Llama Photo**: “*omg look who I saw at university!! <plus a photo of a llama>*”, **Happy Birthday**: “*Happy Birthday! Hope you have a great day today :)*”, and **Compliment**: “*Great work today. You were a pleasure as always!*”.

We focused on message urgency, an influential factor identified in non-XR message notification systems [35, 55], to examine its impact on preferences towards multimodal XR notifications. This approach keeps other contexts around message delivery (e.g., sender identity, receiver activity) consistent while addressing urgency’s broad applicability to various XR notifications, such as bystander presence, navigation, and guidance. Future work should explore other contextual factors to better understand their influence on notification preferences.

4.7 Procedure

Upon arrival, the purpose of the experiment was explained, and the participants gave their informed consent and completed a demographic questionnaire. The experimental task was described, including the various message delivery notifications that participants would experience during the task, the message board, and the message acknowledgment process. After this, participants were instructed where to sit and how to wear and adjust the headset.

Participants then went through a training, during which they practiced the experimental task and experienced each type of notification. First, participants were shown the task environment and the position of the experimental task and the message board. They then practiced the task by selecting 3 targets with the experimenter’s guidance and then selecting 4 targets independently to ensure understanding. After this, each condition was demonstrated and explained to the participant once. During the training, participants adjusted the headset volume to a “comfortable level where they could clearly hear the experimental task and audio notifications”.

After the training, participants were given a break while the experimenter prepared the evaluation phase. Once ready, participants completed each condition sequentially. The condition order was counterbalanced using a balanced Latin square. After all conditions were completed, participants answered the ranking questions and the message awareness needs think-aloud task. Participants think-aloud comments were recorded. The experiment took on average one hour to complete. Participants were compensated for their time with a £10 Amazon voucher.

4.8 Apparatus & Implementation

The study was conducted in a quiet experimental room. The study was built using Unity 2022.3.31f1, Meta XR SDK, and Unity’s XR Interaction Toolkit. Questionnaires were administered using Qualtrics. A Meta Quest 3 headset was used for the study, with the left controller designated for secondary message acknowledgment task and the right controller for the primary experimental task.

4.9 Analysis

We analyzed notification response times using a linear mixed-effects model (GLMM) with a Gamma distribution and log link function, implemented in R (v4.3.0) with the *lme4* and *emmeans* packages. The Gamma distribution was chosen to account for the right-skewed nature of response times, where most responses are short, but a few outliers are substantially longer. The model included the notification type as a fixed effect and the participant ID as a random intercept to account for variability in response times across participants. This setup isolates the effect of notification type on response times while accounting for repeated measures within participants. Pairwise comparisons were conducted using z-tests on the estimated marginal means (EMMs) derived from the GLMM, using the *emmeans* package. Bonferroni correction was applied to adjust for multiple comparisons.

For the Likert-scale questions, we calculated the mean and standard deviation values. We used a Friedman test to identify significant differences between factors and calculated effect sizes using Kendall’s W. We analyzed pairwise comparisons using Wilcoxon Signed Rank tests with Bonferroni-corrected *p*-values. Effect sizes

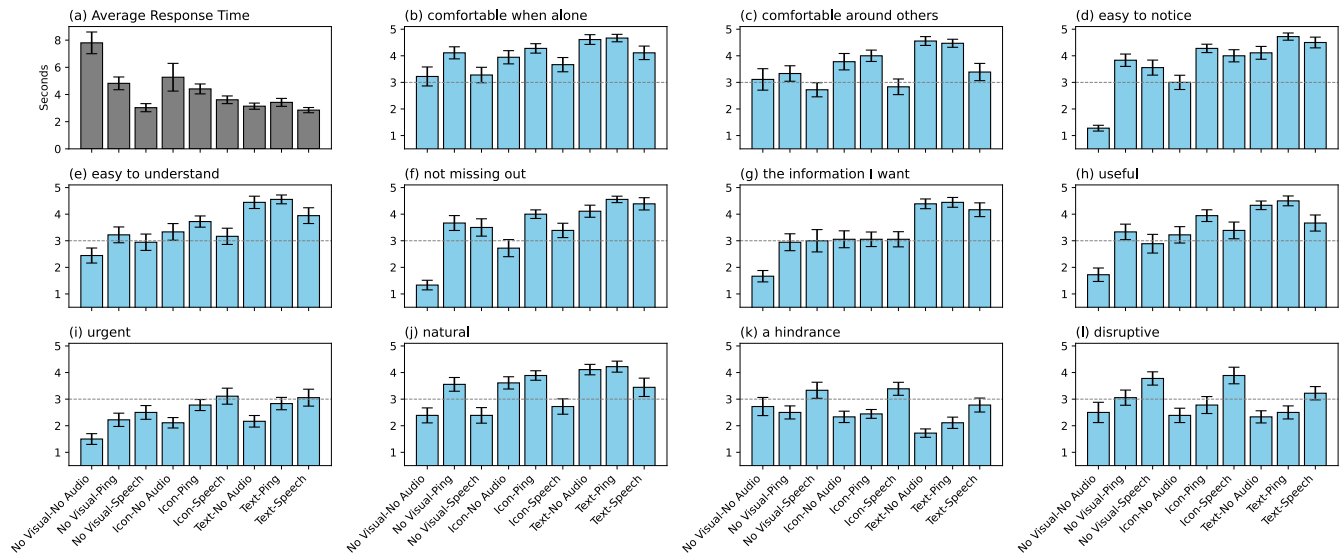


Figure 3: (a) Average response time for each notification condition. The bar chart displays the mean response time (in seconds) for each notification condition, with error bars representing the standard error of the mean. (b)-(l) Mean and standard error for Likert Scale questions.

for statistically significant comparisons using rank-biserial correlation (r). For the preference rankings, the average ranking score for each approach was calculated. Participants' comments justifying their rankings were coded using initial coding [11] where participants' statements were assigned emergent codes over repeated cycles with the codes grouped using a thematic approach. A single coder performed two coding cycles and reviewed the results with another researcher.

For the task assessing message awareness needs, we calculated the counts of each notification type selected for each message and the number of unique notifications used by each participant. Participants' comments justifying their choices were analyzed using the same initial coding approach described above.

5 RESULTS: PERFORMANCE, USABILITY, & RANKINGS

5.1 Participant Demographic Data

Participants were recruited via social media and mailing lists. A total of 18 participants (7 female, 11 male) completed the study, ranging in age from 21 to 48 years ($M = 29.06$, $SD = 8.45$). Regarding prior experience with AR, 2 participants indicated "No / I don't know", 3 indicated "Yes, but only smartphone AR", and 13 indicated "Yes, including AR headsets". Participants also rated their prior experience with AR on a 5-point Likert scale (1 = None, 5 = A lot), with an average rating of 2.94 ($SD = 1.26$); 16 participants reported at least "A little (2)" experience with AR. For VR experience, participants also used a 5-point Likert scale (1 = None, 5 = A lot), with an average of 3.33 ($SD = 1.24$); 17 participants reported at least "A little (2)" experience with VR.

5.2 Differences in Notification Response Time

Figure 3(a) illustrates the average response times to the nine notification designs with standard errors. The log-transformed response time used for analysis is illustrated in Figure 7 with a heatmap visualization of pairwise comparison (Figure 4) and full results of the model and pairwise comparison, including coefficients, standard error, and significant levels (Figure 2) in the Appendix.

The GLMM model fit was assessed with $AIC = 2299.7$, $BIC = 2347.8$, $\log\text{-likelihood} = -1138.9$, and marginal $R^2 = 0.18$. The reference condition, *No Visual-No Audio*, resulted in longer response times than all other conditions ($p \leq 0.001$). This suggests that the

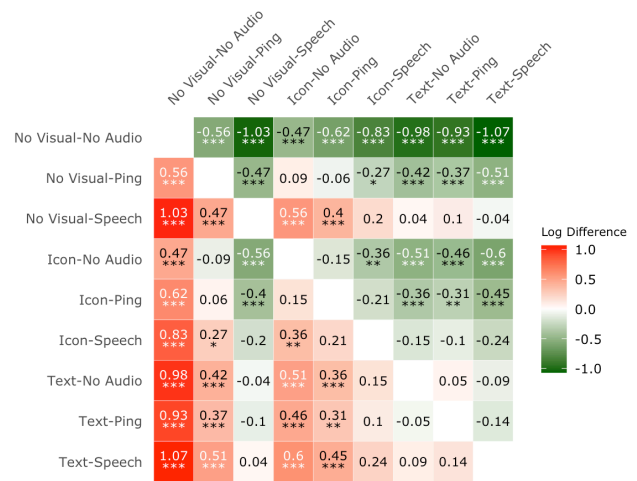


Figure 4: Pairwise comparisons of notification types. The heatmap shows the log differences in response times with estimated values, significance levels ($*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$), and color-coded magnitudes. Red shades indicate longer response times for the column condition compared to the row, while green shades indicate shorter response times.

presence of either visual or auditory cues, or a combination of both, significantly reduces notification response times compared to when no cues are present.

Notifications using detailed visual (*Text*) and/or auditory (*Speech*) information consistently resulted in shorter response times. *Text* conditions significantly reduced response times compared to the corresponding *No Visual* and *Icon* conditions when paired with less detailed audio signals, *No Audio* and *Ping*. For example, participants responded faster to notifications with *Text-No Audio* ($M = 3.14$ s, $SD = 1.91$ s) than *No Visual-No Audio* ($M = 7.80$ s, $SD = 5.76$ s, $p \leq 0.001$) and *Icon-No Audio* ($M = 5.27$ s, $SD = 7.95$ s, $p \leq 0.001$); and faster to *Text-Ping* ($M = 3.42$ s, $SD = 2.48$ s) than *No Visual-Ping* ($M = 4.82$ s, $SD = 3.73$ s, $p \leq 0.001$) and *Icon-Ping* ($M = 4.41$ s, $SD = 2.97$ s, $p \leq 0.01$). However, no signif-

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	Average Ranking
Text-Ping	8	2	3	3	2	0	0	0	0	7.61
Icon-Ping	3	3	4	5	2	1	0	0	0	6.83
Text-No Audio	1	5	2	3	3	2	1	1	0	6.06
Text-Speech	4	3	1	0	2	6	1	1	0	5.89
Icon-No Audio	1	3	2	3	2	1	4	2	0	5.28
Icon-Speech	1	2	1	3	2	3	5	2	0	4.94
No Visual-Ping	0	0	3	1	4	4	4	2	1	4.39
No Visual-Speech	0	0	2	0	1	1	2	11	1	2.89
No Visual-No Audio	0	0	0	0	0	0	1	0	17	1.11
General Preference										

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	Average Ranking
No Visual-Speech	7	5	0	2	1	1	2	0	0	7.22
No Visual-No Audio	10	2	1	1	1	0	1	0	2	7.17
Icon-Speech	0	4	5	2	2	0	2	3	0	5.61
No Visual-Ping	0	2	3	2	3	5	2	0	1	5.06
Text-Speech	1	2	5	2	2	0	1	0	5	5.00
Icon-No Audio	0	2	3	2	3	2	1	3	2	4.61
Text-No Audio	0	1	1	2	2	5	2	4	1	4.00
Icon-Ping	0	0	0	4	2	2	4	2	4	3.44
Text-Ping	0	0	0	1	2	3	3	6	3	2.89
Frustration										

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	Average Ranking
Text-Speech	12	3	2	1	0	0	0	0	0	8.44
Text-Ping	5	2	5	5	1	0	0	0	0	7.28
Icon-Speech	0	8	3	3	2	0	0	2	0	6.50
Icon-Ping	0	2	5	4	5	1	1	0	0	5.94
No Visual-Speech	0	2	3	1	4	1	4	3	0	4.72
Text-No Audio	0	1	0	2	1	6	7	1	0	4.00
No Visual-Ping	0	0	0	2	2	2	7	3	4	3.72
Icon-No Audio	1	0	0	0	3	3	3	8	0	3.39
No Visual-No Audio	0	0	0	0	0	0	0	0	18	1.00
Noticeability										

Figure 5: The ranking distributions and average ranking scores (out of a possible 9.00) for our conditions. General preference is ordered best to worst. Frustrating is ordered most to least frustrating. Noticeability is ordered most to least noticeable.

icant differences were observed when *Text* was paired with *Speech*: *Text-Speech* ($M = 2.85$ s, $SD = 1.62$ s), *No Visual-Speech* ($M = 3.03$ s, $SD = 2.40$ s), and *Icon-Speech* ($M = 3.61$ s, $SD = 2.30$ s).

Similarly, *Speech* conditions significantly outperformed their corresponding *No Audio* and *Ping* conditions when paired with visual cues that included no or less information. *No Visual-Speech* ($M = 3.03$ s, $SD = 2.40$ s) was faster than *No Visual-No Audio* ($M = 7.80$ s, $SD = 5.76$ s, $p \leq 0.001$) and *No Visual-Ping* ($M = 4.82$ s, $SD = 3.73$ s, $p \leq 0.001$); and *Icon-Speech* ($M = 3.61$ s, $SD = 2.30$ s) was faster than *Icon-No Audio* ($M = 5.27$ s, $SD = 7.95$ s, $p \leq 0.01$) with no statistical significance to *Icon-Ping* ($M = 4.41$ s, $SD = 2.97$ s, $p > 0.05$). However, no significant differences ($p > 0.05$) were observed between *Text-Speech* ($M = 2.85$ s, $SD = 1.62$ s) and *Text-No Audio* ($M = 3.14$ s, $SD = 1.91$ s) and *Text-Ping* ($M = 3.42$ s, $SD = 2.48$ s).

5.3 Perceived Comfort, Informativeness, and Usability

Figure 3(b)-(l) illustrates the mean results of our Likert-scale questions with standard errors. The full results of the Friedman test and post-hoc analysis are provided in our Appendix (Table 5). Effect sizes for the Friedman tests range from 0.23 to 0.51, indicating a small to moderate levels of agreement. Effect sizes for the Wilcoxon Signed Rank comparisons range from 0.75 to 0.90, indicating large and substantial difference between the paired conditions. A summary of key results follows.

Comfort around others (Figure 3(c)) was rated lowest in conditions including *Speech* (*No Visual-Speech*: $M = 2.72$, $SD = 1.13$; *Icon-Speech*: $M = 2.83$, $SD = 1.25$; *Text-Speech*: $M = 3.39$, $SD = 1.38$) due to concern for private messages being overheard, e.g. P5 stated, “really not great for privacy... hearing someone’s messages on the train second-hand would be horrible”. Some participants found conditions with *Ping* socially unacceptable, fearing it might annoy others. As P1 noted, “I get annoyed by pings easily... my father has games on his phone that send him notifications all the time [and] I get easily annoyed by it”. Nonetheless, many participants reported being comfortable using *Ping* around others (*No Visual-Ping*: $M = 3.33$, $SD = 1.24$; *Icon-Ping*: $M = 4.00$, $SD = 0.91$; *Text-Ping*: $M = 4.47$, $SD = 0.62$).

Informativeness ratings (Figure 3(d)-(g)) showed *No Visual-No Audio* performed the worst, being significantly different from all other conditions for being (d) **easy to notice** ($M = 1.27$, $SD = 0.46$) and from all conditions but *Icon-No Audio* for feeling of (f) **not missing out** ($M = 1.33$, $SD = 0.77$). The *Icon-No Audio* condition was reported as harder to notice than others ($M = 3.00$, $SD = 1.14$). Participants felt *Icon* alone was not distinct enough when focusing on the experimental task. P15 commented, “Visually, I’m focused on the task so it gets lost.” However, participants suggested if paired with *Ping* ($M = 4.28$, $SD = 0.67$) or *Speech* ($M = 4.00$, $SD = 0.97$) that the noticeability could increase; e.g. P8 observed, “[With] ping it’s more noticeable... it makes it more distinguishable, especially when doing other stuff.”

The desired information ((g) **the information I want**) was rated highest for conditions containing *Text*: *Text-No Audio* ($M = 4.39$, $SD = 0.78$); *Text-Ping* ($M = 4.44$, $SD = 0.78$); *Text-Speech* ($M = 4.17$, $SD = 1.10$). Participants preferred receiving the message content alongside the notification, allowing them to make an informed decision about how to respond. For example, P15 noted, “Message is always great because it’s a big blob of information you can decide what to do next with”.

Usability was rated highest for the most (h) **useful** and (j) **natural** conditions for *Icon-Ping* (useful: $M = 3.94$, $SD = 0.93$; natural: $M = 3.89$, $SD = 0.76$), *Text-No Audio* (useful: $M = 4.33$, $SD = 0.68$; natural: $M = 4.11$, $SD = 0.83$), *Text-Ping* (useful: $M = 4.50$, $SD = 0.79$; natural: $M = 4.22$, $SD = 0.88$). *Icon-Speech* ($M = 3.11$, $SD = 1.28$) and *Text-Speech* ($M = 3.06$, $SD = 1.35$) were seen as the most (i) **urgent** with significant differences mainly observed between *No Visual-No Audio* and *Icon-Ping*, and between *No Visual-No Audio* and *Icon-Speech*. None of the conditions were perceived as being particularly (k) **a hindrance** to the experience. While *Icon-Speech* ($M = 3.89$, $SD = 1.32$) was reported as the most (l) **disruptive**, no significant differences were found regarding disruption.

5.4 Rankings

Figure 5 summarizes the ranking distributions and average ranking scores for all conditions. For general preference, *Text-Ping* scored highest (7.61 out of 9.00) and was the first choice of 44.4% of participants. In contrast, *No Visual-No Audio* performed worst (1.11 out of 9.00), with 94.4% of participants ranking it last. Participants favored *Text-Ping* due to the combination of *Ping* for drawing attention (e.g. P1: “with a ping you notice there’s something you should pay attention to”) and *Text* for its informativeness (e.g. P8: “you can see what its about and what’s the urgency to respond”). *No Visual-No Audio* was ranked last because participants found it frustrating to frequently interrupt the task to check the message window, leading to feelings of disconnection from their messages.

Text-Speech also stood out, with the second highest number of participants (22.2%) ranking it first. Participants who preferred this condition appreciated *Speech* for allowing them to stay focused on the experimental task while maintaining awareness of messages. For example, P3 remarked, “I didn’t have to look away or look up or down”. However, participants acknowledged a potential limitation of speech-based delivery for longer messages to be accurately retained when delivered through *Speech*. P2 commented, “if it’s a long message, or a really long message, then it’s too much to remember when delivered in the speech”. Hence, participants felt pairing *Speech* with *Text* provided the best of both worlds: *Text* for checking detailed information and *Speech* for shorter messages that could be delivered aurally without distracting from the task. However, 55.5% ranked *Text-Speech* fifth or worse, finding *Speech* too disruptive (e.g. P11: “the speech was intrusive, it was disrupting”). Additionally, participants pointed out *Speech* might be unsuitable for tasks involving audio, such as “listening to talking, reading, mu-

sic, or something” (e.g. P11: “*Most things I’d be doing would have some sort of interference from speech*”).

For frustration, *No Visual-Speech* (7.22 out of 9.00, with 38.8% ranking it first) and *No Visual-No Audio* (7.17 out of 9.00, with 55.5% ranking it first) were the most frustrating conditions. Participants found *No Visual-Speech* disruptive, particularly due to the unexpected onset of speech, which was seen as confusing in some contexts (e.g. P18: “*if I’m in a social situation, I’d just confuse the speech*”). Some suggested a preceding audio cue (such as *Ping*) could mitigate this issue by signaling the start of *Speech* (e.g. P2: “*I think you need it to ping you and then starts talking to you... if it just starts talking, it takes you a while to work out what’s actually going on and why it started to talk to you*”). Meanwhile, *No Visual-No Audio* was frustrating for the same reasons it was ranked lowest in general preference: the need to constantly check the message window and the sense of disconnection from messages.

For noticeability, *Text-Speech* (8.44 out of 9.00, with 66.7% ranking it first) and *Text-Ping* (7.28 out of 9.00, with 28.8% ranking it first) were the most noticeable conditions. In general, conditions that combined visual and audio elements were considered more noticeable than those with only visual or audio. Participants expressed concern that a single modality could be easily masked by either the virtual or real environment (e.g. P2: “*Speech could be the application or the TV in the room or something else*”). Having both ensured if one was missed or masked the other would still make participants aware of the notification.

6 RESULTS: ASSESSING MESSAGE AWARENESS NEEDS

The counts of notifications selected in our assessing message awareness needs task are summarized in Figure 6. All participants used multiple types of notifications within the task (Section 6.1), highlighting a preference for varying notification types depending on the message content. All participants agreed their preferred notification depended on the message’s urgency and content (Section 6.2). Additionally, participants discussed other factors influencing their notification preferences (Section 6.3) and suggested alternative notification designs not included in our study (Section 6.4).

6.1 Use of Multiple Notifications

We examined how many unique notifications each participant chose for the sample messages. Across all six messages, 11 participants selected 3 unique notifications, 3 participants selected 2 notifications, 2 participants selected 4 notifications, and 2 participants selected 5 notifications. The most common range was 3 notifications used by 11 participants. Of these, 6 participants used *Icon-Ping*, *Text-Ping*, and *Text-Speech*. The remaining 5 participants chose: *Text-Speech*, one of *No Visual-No Audio*, *Icon-No Audio*, or *No Visual-Ping*, and a third notification.

Despite variations, all 11 participants expressed thinking about notifications in terms of “*low, medium, or high*” urgency. *Text-Speech* was universally seen as the highest urgency option (e.g. P2: “*It’s the most awareness of both visual and audio options*”). However, four participants wanted an even more disruptive option for high urgency (e.g. P17: “*I’d probably want it to just pause the experience and show this important message*”).

Opinions differed for low-urgency messages. Six participants chose *Icon-Ping*, preferring to be alerted of new messages so they could decide when to attend to them (e.g. P3: “*I definitely want to know they are still there, then can decide when to look at them whenever I have time*”). Five participants preferred no (*No Visual-No Audio*) or minimal disruption (*Icon-No Audio* or *No Visual-Ping*) for low-urgency messages, as they did not consider immediate attention necessary (e.g. P10: “*if I was doing something immersive, and the message I got was not important, I probably would not want to be notified about it and would just get it whenever*”). The choice of a medium notification depended on partici-

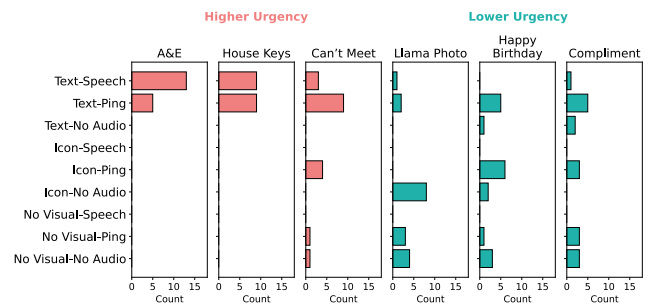


Figure 6: Counts of notifications selected for each message in the message awareness needs assessment tasks (Section 4.6).

pants’ low-urgency choice. Those who picked a more disruptive option for low urgency (i.e. *Icon-Ping*) selected *Text-Ping* for medium urgency, while the others varied based on personal preferences.

Participants who selected fewer or more than 3 notifications also matched their choices to message urgency. The 3 participants using only 2 notifications felt the sample messages were either high or low urgency and did not require a medium step. The 2 participants selecting 4 notifications wanted the option to use no notification (*No Visual-No Audio*), particularly for less significant messages (e.g. P1: “*If it’s your birthday, you get loads of those, so I just want no visual, no audio, nothing*”). The 2 participants who used 5 notifications aimed for a more nuanced approach to urgency, rather than generalizing it as low, medium, or high (e.g. P10: “*It still also depends on the context of what I was doing in AR as well. So there’s really just a lot of range when to notify me and how*”).

6.2 Influence of Message Content

All participants agreed that their preferred notification depended on the message content, with more urgent and important messages requiring more disruptive notifications, as illustrated in Figure 6. Messages such as *A&E* (*accident and emergency*) and *House Keys* were seen as very urgent and justified immediate attention (e.g. P17: “*If it’s an emergency scenario, I would want the experience to be disturbed fully*”), and all participants chose either *Text-Speech* or *Text-Ping* for these messages. While *Can’t Meet* was considered time-sensitive, it was not viewed as critical (e.g. P18: “*It’s important, but I’ve got more time to attend to it*”), with some participants choosing less disruptive notifications and 1 participant going as far as to select *No Visual-No Audio*.

Time sensitivity was not the only factor; message type also influenced notification choice. For example, 3 participants wanted *No Visual-No Audio* for *Happy Birthday* messages, due to the large number they might receive on that day (e.g. P1: “*If it’s your birthday you get loads of those so I just want no visual, no audio*”). However, 8 participants preferred a more noticeable notification (i.e. *Text-Ping* or *Icon-Ping*) to ensure they could respond and thank the sender (e.g. P13: “*Happy birthday messages are nice and I want to make sure I thank everyone*”). Similarly, participants preferred more disruptive notifications (i.e. *Text-Ping* or *Icon-Ping* or *Text-Speech*) for positive messages like *Compliment*, as they valued seeing and responding to uplifting content in a timely manner (e.g. P13: “*it’s nice to read something positive, somebody is appreciating me, obviously I need to reply*”).

Finally, *Llama Photo* was considered the least urgent message, with many participants opting for lower-priority notifications. Most felt that missing the notification would be acceptable (e.g. P2: “*If you miss it, it’s no big deal. You’ll see the llama eventually*”).

6.3 Additional Factors Influencing Notification Choice

Besides message content, most participants noted that their current task would influence how they preferred to be notified and of what messages (e.g. P9: “It is very context-dependent... if I am at home watching a movie that’s different than at work and P10: “If I’m at work, then I only want critical messages or work-related messages”). Additionally, participants said the identity of the sender mattered (e.g. P8: “Who it is can influence if it’s more urgent, if its your close friends and family that’s top priority, then say your housemates, and so on, but there’s groups of urgency by people too”). One participant also noted a difference in their response to group chat versus one-on-one messages (e.g. P7: “I don’t get notifications for group chats, but I get notifications for direct one to one chats, because if people now if they want something from me to go direct, that’s a bit more urgent”). These insights highlight that future work is needed to explore factors beyond the urgency of a message’s content which influence notification preferences.

6.4 Suggested Alternative Notifications

Some participants proposed alternative notifications for message delivery. Four participants wanted more disruptive designs for the highest urgency messages, ranging from more disruptive audio alerts (e.g. P18: “if it is an emergency then it should ring [like a phone call]”), to halting the application in use and forcing the user to engage with the message (e.g. P17: “I’d probably want it to just pause the experience and show this important message”). Two participants suggested subtle notifications using haptic feedback (e.g. P5: “definitely an option could be haptic feedback, especially in the controllers”), although previous research by Ghosh et al. has shown this to be ineffective for XR notifications [17]. One participant suggested embedding notifications within the current experience to match its theme (e.g. P11: “If I was watching a movie or playing a game then I might want to be somehow integrated into the theme of what I’m doing... like the pop-up has a fantasy theme while watching Game of Thrones”), aligning with research by Christie et al. on content-themed notifications in AR [10].

7 DISCUSSION

Our study examined the effects of audio and visual notification designs in XR environments and explored user preferences for different types of notifications. Based on our findings, we identify key factors in designing effective notifications for XR user interfaces, including the importance of context-aware delivery, sensory management, and message content. These insights offer new directions for developing adaptive notification systems in XR that balance noticeability with user comfort and privacy.

7.1 Impact of Audio Elements on XR Notifications

Building on prior research [52] that examined notification placements in VR, our study explored how the addition of audio elements and visual representations of message notifications influenced response times, comfort, informativeness, usability, and preferences. Our findings suggest that the effect may depend on the combination of the two modalities. Pairing *No Visual* notifications (similar to the *In-Situ* placement in Rzyayev et al. [52]) with audio elements (*Ping* and *Speech*) make notifications easier to notice and reduce response times.

When a visual cue is present (*Icon* or *Text*), comparable to the *Floating* and *Head-Up Display* notifications of Rzyayev et al., the effect may depend on whether the audio cue contains more information than the visuals. Adding more detailed auditory than visual information, e.g. *Icon-Speech*, reduced the response time by 1.66 s on average ($p \leq 0.01$). Yet, the differences were marginal when less detailed or equally detailed audio was paired with visual cue: between *Icon-No Audio* and *Icon-Ping* (average difference -0.86 s,

ns); between *Text-No Audio* and *Text-Ping* (average difference 0.28 s, ns); and between *Text-No Audio* and *Text-Speech* (average difference -0.57 s, ns). Similarly, adding more detailed visual than audio shortened the response time (average difference: 1.4 s, $p \leq 0.001$) while marginal differences (ns) were observed when less or equally detailed visual cue was paired with audio. However, the lack of significant differences in these comparisons requires careful interpretation and further investigation to validate these effects in broader XR contexts.

In terms of privacy and comfort, some participants expressed concerns about audio notifications being overheard in public places. The use of discreet audio methods, such as headphones and directed speakers, can help address these concerns, ensuring that notifications remain private while maintaining their effectiveness in attracting user attention.

7.2 Adaptive Multimodal Notifications in XR

Our results emphasize the importance of offering diverse combinations of multimodal notifications that adapt to the urgency of the message, the user’s available sensory channels, and preferences surrounding message awareness. In doing so, we expand on Zenner et al.’s concept of visual pop-ups with escalating disruptiveness [62], highlighting the benefit of adaptive multimodal notification mechanisms for messaging systems in future XR systems. XR, in particular, is well suited to such personalization and context-aware adaptation due to its inherent dynamism and the flexibility it enables in users use of applications, interfaces, and interactions aligned around user tasks and priorities [9].

7.2.1 Eye Gaze Tracking and Adaptive Audio

One direction for future work involves integrating eye gaze tracking to determine whether users have noticed a visual notification [30, 61]. Since our results show auditory cues are useful when notifications are out of the user’s view, eye tracking could trigger audio only when necessary. This adaptive approach would enhance noticeability while minimizing unnecessary auditory intrusions.

7.2.2 Managing Sensory Overload

Participants highlighted that context matters when deciding whether audio notifications are appropriate. If the user’s visual or auditory channels are already engaged [33], such as during a conversation or when immersed in a game, audio cues may be disruptive or masked. In such cases, less intrusive modalities, like subtle visual notifications or no notifications, may be preferred. Alternative approaches include ones which alter a user’s experienced audio without significantly impacting engagement, e.g. dynamically altering volume levels [42], embedding notifications and auditory content into the music users are currently listening to [57, 58], or removing a subset of audio sources [42, 50]. Systems could also consider adjusting the timing of notifications to prevent interruptions during high cognitive load tasks [5, 45].

7.2.3 Haptic Feedback for Subtle Alerts

Several participants suggested haptic feedback as a less intrusive option for low-urgency notifications. While haptics alone may not be as effective in XR [17], its combination with visual or audio cues may prove useful for subtle alerts. Haptics could offer a discreet alternative, ensuring that notifications are still noticed without intruding on ongoing tasks, particularly when both audio and visual channels are occupied.

7.2.4 Message Content as a Factor

Message content plays a critical role in determining the most appropriate notification modality. Our participants frequently categorized notifications based on message urgency (low, medium, or high). Urgent messages may warrant more noticeable audiovisual

cues, while less urgent messages could be delivered with minimal disruption. This may apply to not only interpersonal messaging but also various notifications in XR, such as the presence of bystanders [41] or other real-world events, feedback for gaze, gestural, or speech commands [7, 29], navigation [27], and guidance to XR objects [8]. Future XR systems should adapt notification delivery based on the message’s urgency and relevance, ensuring timely awareness without overloading the user. In addition, future work is necessary to identify other influential factors, and to determine how to balance and optimize for multiple factors at once, for example altering how notifications are delivered based on the sender’s identity, receiver’s on-going task, and message’s urgency.

7.3 Alternative XR Messaging Approaches

Some participants indicated a preference for receiving message content directly, either via *Text* or *Speech*, allowing them to make informed decisions about how to respond. Future work could enhance this by summarizing longer messages [59, 60], enabling users to assess message importance without processing its full content. For example, instead of receiving a detailed paragraph, users might see a summary like “*Julie – feedback on experimental design*”. However, care must be taken to avoid misrepresenting the sender’s intent, especially for sensitive messages. For instance, if a sibling sent a detailed message about a medical issue concerning a niece, summarizing it as “*Brother – your niece is probably dying*” would be tactless and could cause unnecessary distress or confusion. This highlights the importance of carefully tailoring summaries to both the content and the emotional weight of the message.

Future work could also explore giving message senders more control over notification delivery [6, 62] by leveraging XR’s advanced context sensing capabilities [18, 37, 38]. For instance, XR devices could assess the receiver’s context, such as activity, environment, and readiness to engage [15], to adapt how and when notifications are delivered [30]. This would enable senders to tailor notifications dynamically, ensuring timely communication without causing unnecessary disruption. However, any system that shares contextual data from the receiver must take privacy risks into account [40], ensuring such information is shared in a controlled, privacy-preserving manner to prevent potential misuse for surveillance, intrusion, or manipulation.

8 CONCLUSION

We explored the impact of audio and visual elements on notifications in XR environments, focusing on how different modalities and presentations affect response times, noticeability, and user preferences. Our findings demonstrate adding auditory cues, particularly to out-of-view visual notifications, enhances noticeability and reduces response times, while visual notifications alone suffice when within the user’s field of view. Participants expressed preferences for context-aware, adaptive notifications that take into account message urgency, task engagement, and sensory load. Future work should focus on adaptive, context-sensitive notification systems that balance noticeability, intrusiveness, and privacy, ensuring effective communication in immersive XR environments.

REFERENCES

[1] Apple. CarPlay, 2021. Last Accessed 04/09/2024. 2
 [2] Apple. Apple Vision Pro, 2024. Last Accessed 04/09/2024. 2, 3
 [3] Apple. iMessage Icon, 2024. Last Accessed 04/09/2024. 2
 [4] Apple. iPhone Ping Noise, 2024. Last Accessed 04/09/2024. 2
 [5] K.-W. Chen, Y.-J. Chang, and L. Chan. Predicting opportune moments to deliver notifications in virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI ’22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3517529 1, 2, 8

[6] H. Cho, J. Oh, J. Kim, and S.-J. Lee. I share, you care: Private status sharing and sender-controlled notifications in mobile instant messaging. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW1), may 2020. doi: 10.1145/3392839 9
 [7] H. Cho, N. Sendhilkathan, M. Nebeling, T. Wang, P. Padmanabhan, J. Browder, D. Lindlbauer, T. R. Jonker, and K. Todi. Sonohaptics: An audio-haptic cursor for gaze-based object selection in xr. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–19, 2024. 1, 9
 [8] H. Cho, A. Wang, D. Kartik, E. L. Xie, Y. Yan, and D. Lindlbauer. Auptimize: Optimal placement of spatial audio cues for extended reality. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–14, 2024. 1, 9
 [9] H. Cho, Y. Yan, K. Todi, M. Parent, M. Smith, T. R. Jonker, H. Benko, and D. Lindlbauer. Minexr: Mining personalized extended reality interfaces. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI ’24. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613904.3642394 1, 2, 3, 8
 [10] I. Christie, M. McGill, and S. A. Brewster. Welcome to the jungle: Using artv to balance media immersion and reality awareness for passengers on public transport. In *Proceedings of the 2024 ACM International Conference on Interactive Media Experiences*, IMX ’24, p. 219–230. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3639701.3656321 8
 [11] S. A. L. Corbin J. M. *Basics of qualitative research: techniques and procedures for developing grounded theory*. SAGE Publications, Inc., 1998. 5
 [12] David Angel. The Four Types of Conversations: Debate, Dialogue, Discourse, and Diatribe, 2016. Last Accessed 04/09/2024. 3
 [13] J. E. Fischer, C. Greenhalgh, and S. Benford. Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI ’11, p. 181–190. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/2037373.2037402 2
 [14] J. E. Fischer, N. Yee, V. Bellotti, N. Good, S. Benford, and C. Greenhalgh. Effects of content and time of delivery on receptivity to mobile interruptions. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI ’10, p. 103–112. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1851600.1851620 2
 [15] P. E. Fortin, E. Sulmont, and J. Cooperstock. Detecting perception of smartphone notifications using skin conductance responses. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI ’19, p. 1–9. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300420 9
 [16] C. George, P. Tamunjoh, and H. Hussmann. Invisible boundaries for vr: Auditory and haptic signals as indicators for real world boundaries. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3414–3422, 2020. 2
 [17] 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. doi: 10.1109/TVCG.2018.2793698 2, 3, 8
 [18] Google. Learn more about Android XR, 2024. Last Accessed 06/01/2025. 9
 [19] Google. Text-to-Speech AI, 2024. Last Accessed 04/09/2024. 2, 3
 [20] C.-Y. Hsieh, Y.-S. Chiang, H.-Y. Chiu, and Y.-J. Chang. 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*, CHI ’20, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376228 2
 [21] S. H. Hsieh and T. H. Tseng. Playfulness in mobile instant messaging: Examining the influence of emoticons and text messaging on social interaction. *Computers in Human Behavior*, 69:405–414, 2017. doi: 10.1016/j.chb.2016.12.052 2
 [22] S. Imamov, D. Monzel, and W. S. Lages. Where to display? how interface position affects comfort and task switching time on glance-

- able interfaces. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 851–858, 2020. doi: 10.1109/VR46266.2020.00110 2
- [23] N. Janaka, S. Zhao, and S. Chan. Notifade: Minimizing ohmd notification distractions using fading. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544549.3585784 2
- [24] N. N. P. K. Janaka, S. Zhao, and S. Sapkota. Can icons outperform text? understanding the role of pictograms in ohmd notifications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3580891 1, 2
- [25] 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*, ICMI '21, p. 470–477. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3462244.3479898 1, 2
- [26] 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, 2023. doi: 10.1109/VR55154.2023.00067 2
- [27] J. Lee, F. Jin, Y. Kim, and D. Lindlbauer. User preference for navigation instructions in mixed reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 802–811. IEEE, 2022. 1, 9
- [28] K. Lee, H. Li, M. R. Wellyanto, Y. J. Tham, A. Monroy-Hernández, F. Liu, B. A. Smith, and R. Vaish. Exploring immersive interpersonal communication via ar. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1):1–25, 2023. 2
- [29] M. Lee, M. Billingham, W. Baek, R. Green, and W. Woo. A usability study of multimodal input in an augmented reality environment. *Virtual Real.*, 17(4):293–305, Nov. 2013. doi: 10.1007/s10055-013-0230-0 1, 9
- [30] Z. Li, Y. F. Cheng, Y. Yan, and D. Lindlbauer. Predicting the noticeability of dynamic virtual elements in virtual reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2024. 8, 9
- [31] D. Lindlbauer, A. M. Feit, and O. Hilliges. Context-aware online adaptation of mixed reality interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, pp. 147–160. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3332165.3347945 1
- [32] S. Liu and D. Lindlbauer. TurnAware: motion-aware augmented reality information delivery while walking. *Frontiers in virtual reality*, 5:1484280, 18 Dec. 2024. doi: 10.3389/frvir.2024.1484280 2
- [33] X. B. Liu, J. N. Li, D. Kim, X. Chen, and R. Du. Human i/o: Towards a unified approach to detecting situational impairments. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–18, 2024. 8
- [34] F. Lu, S. Davari, L. Lisle, Y. Li, and D. A. Bowman. Glanceable ar: Evaluating information access methods for head-worn augmented reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 930–939, 2020. doi: 10.1109/VR46266.2020.00113 2
- [35] A. Mehrotra, V. Pejovic, J. Vermeulen, R. Hendley, and M. Musolesi. My phone and me: understanding people’s receptivity to mobile notifications. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pp. 1021–1032, 2016. 4
- [36] Meta. Quest 3, 2023. Last Accessed 04/09/2024. 2
- [37] Meta. Introducing Orion Our First True Augmented Reality Glasses, 2024. Last Accessed 22/12/2024. 9
- [38] Meta. Project Aria Datasets, 2024. Last Accessed 22/12/2024. 9
- [39] J. O’Hagan, J. Gugenheimer, J. Bonner, F. Mathis, and M. McGill. Augmenting people, places & media: The societal harms posed by everyday augmented reality, and the case for perceptual human rights. In *Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia*, MUM '23, p. 225–235. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3626705.3627782 2
- [40] J. O’Hagan, P. Saeghe, J. Gugenheimer, D. Medeiros, K. Marky, M. Khamis, and M. McGill. Privacy-enhancing technology and everyday augmented reality: Understanding bystanders’ varying needs for awareness and consent. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(4), jan 2023. doi: 10.1145/3569501 2, 9
- [41] J. O’Hagan and J. R. Williamson. Reality aware vr headsets. In *Proceedings of the 9TH ACM International Symposium on Pervasive Displays*, PerDis '20, p. 9–17. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3393712.3395334 1, 2, 9
- [42] J. O’Hagan, J. R. Williamson, M. Khamis, and M. McGill. Exploring manipulating in-vr audio to facilitate verbal interactions between vr users and bystanders. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*, AVI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3531073.3531079 8
- [43] J. O’Hagan, J. R. Williamson, F. Mathis, M. Khamis, and M. McGill. Re-evaluating vr user awareness needs during bystander interactions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581018 1, 2, 3, 4
- [44] J. O’Hagan, J. Gugenheimer, F. Mathis, J. Bonner, R. Jones, and M. McGill. A viewpoint on the societal impact of everyday augmented reality and the need for perceptual human rights. *IEEE Security & Privacy*, 22(1):64–68, 2024. doi: 10.1109/MSEC.2023.3333988 2
- [45] C. Park, J. Lim, J. Kim, S.-J. Lee, and D. Lee. Don’t bother me. i’m socializing! a breakpoint-based smartphone notification system. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 541–554, 2017. 2, 8
- [46] K. Pfeuffer, Y. Abdrabou, A. Esteves, R. Rivu, Y. Abdelrahman, S. Meitner, A. Saadi, and F. Alt. Attention: A design space for gaze-adaptive user interfaces in augmented reality. *Computers & Graphics*, 95:1–12, 2021. doi: 10.1016/j.cag.2021.01.001 2
- [47] M. Pielot, K. Church, and R. de Oliveira. An in-situ study of mobile phone notifications. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*, MobileHCI '14, p. 233–242. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2628363.2628364 2
- [48] 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*, SUI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3565970.3567701 1, 2
- [49] L. Plabst, A. Raikwar, S. Oberdörfer, F. R. Ortega, and F. Niebling. Exploring unimodal notification interaction and display methods in augmented reality. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*, VRST '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3611659.3615683 2
- [50] K. Rogers, G. Ribeiro, R. R. Wehbe, M. Weber, and L. E. Nacke. Vanishing importance: Studying immersive effects of game audio perception on player experiences in virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173902 8
- [51] 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*, NordiCHI '20. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3419249.3420095 1, 2, 3, 4
- [52] R. Rzayev, S. Mayer, C. Krauter, and N. Henze. Notification in vr: The effect of notification placement, task and environment. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '19, p. 199–211. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3311350.3347190 1, 2, 3, 4, 8
- [53] A. Sahami Shirazi, N. Henze, T. Dingler, M. Pielot, D. Weber, and A. Schmidt. Large-scale assessment of mobile notifications. In *Pro-*

- ceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, p. 3055–3064. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2556288.2557189 2
- [54] A. W. Salmoni and J. S. McIlwain. Fitts' reciprocal tapping task, a measure of motor capacity? *Perceptual and Motor Skills*, 49(2):403–413, 1979. 3
- [55] M. H. Vastenburg, D. V. Keyson, and H. de Ridder. Interrupting people at home. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, vol. 1, pp. 59–64. IEEE, 2004. 4
- [56] S. P. Walsh, K. M. White, and R. M. Young. The phone connection: A qualitative exploration of how belongingness and social identification relate to mobile phone use amongst australian youth. *Journal of Community & Applied Social Psychology*, 19(3):225–240, 2009. 2
- [57] A. Wang, Y. F. Cheng, and D. Lindlbauer. MARingBA: Music-adaptive ringtones for blended audio notification delivery. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 11 May 2024. doi: 10.1145/3613904.3642376 8
- [58] A. Wang, D. Lindlbauer, and C. Donahue. Towards music-aware virtual assistants. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, vol. 28, pp. 1–14. ACM, New York, NY, USA, 13 Oct. 2024. doi: 10.1145/3654777.3676416 8
- [59] P.-J. Wang, Y.-C. Lee, U.-D. Chen, and Y.-J. Chang. Notisummary: Exploring the potential of ai-driven text summarization on smartphone notification management. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing, UbiComp/ISWC '23 Adjunct*, p. 113–117. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3594739.3610702 9
- [60] G. Wu, J. Qian, S. Castelo Quispe, S. Chen, J. Rulff, and C. Silva. Artist: Automated text simplification for task guidance in augmented reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–24, 2024. 9
- [61] D. Yu, R. Desai, T. Zhang, H. Benko, T. R. Jonker, and A. Gupta. Optimizing the timing of intelligent suggestion in virtual reality. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–20, 2022. 8
- [62] 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*, CHI EA '18, p. 1–6. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3170427.3188505 2, 8, 9