

A Mixed-Methods Investigation of XR Security Warnings—Lessons Learned

Junyi Zou

*School of Computing & Mathematical Sciences
University of Greenwich
London, United Kingdom
j.zou@greenwich.ac.uk*

Ali Hamza

*School of Computing & Mathematical Sciences
University of Greenwich
London, United Kingdom
a.hamza@greenwich.ac.uk*

Riccardo Bovo

*School of Computing & Mathematical Sciences
University of Greenwich
London, United Kingdom
r.bovo@greenwich.ac.uk*

Georgios Loukas

*School of Computing & Mathematical Sciences
University of Greenwich
London, United Kingdom
g.loukas@greenwich.ac.uk*

Abstract—As immersive XR environments become more prevalent, timely and effective security warnings are essential to protect users from cyberattacks that compromise performance and well-being. This paper investigates how users perceive and respond to in-headset alerts triggered during Denial-of-Service (DoS) attacks. We developed a real-time warning system and evaluated its effectiveness across three pilot studies ($n = 46$) in healthcare and industrial training scenarios. Using self-report measures (IDSQ, SAM) and behavioural categorization, we assessed alert comprehension, urgency perception, and user action. We distil three design lessons emphasizing the importance of visual salience, modality coordination, and urgency calibration. These findings offer practical guidance for designing effective XR security notifications that support user awareness and action during immersive threats.

Index Terms—Extended Reality (XR), Security Warnings, Intrusion Detection, Denial-of-Service (DoS), Multimodal Alerts, Human-Computer Interaction (HCI), User Attention, Cybersecurity, Immersive Environments, Usable Security

I. INTRODUCTION

Extended Reality (XR) technologies are being increasingly adopted across domains such as healthcare, training, and productivity. However, XR’s immersive nature—which isolates users from their physical surroundings—also makes them vulnerable to cybersecurity threats that compromise not just system integrity, but user well-being. In particular, Denial-of-Service (DoS) attacks can degrade performance, induce VR sickness, and disrupt cognitive flow by overwhelming rendering pipelines [1]. Under such conditions, users must be alerted promptly to exit the environment or remove their headset. Designing effective in-headset security warnings is therefore critical to maintaining both safety and trust in XR systems.

This work has been supported by Social and hUman ceNtered XR (SUN) project that has received funding by the Horizon Europe Research & Innovation Programme under Grant agreement N. 101092612. Views and opinions expressed in this paper are those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the European Commission can be held responsible for them.

While research from mobile and automotive human-computer interaction (HCI) offers valuable insights into urgency design, spatial placement, and cross-modal alerts [2]–[5], these studies are rooted in planar, interruptible contexts. XR introduces new challenges due to its spatial attention demands, high embodiment, and limited access to external cues [6], [7]. Recent XR studies have examined general alert perception [8], [9], but lack empirical research on how users perceive and respond to security warnings during immersive cyberattacks, limiting practical design guidance. To address this gap, we developed a real-time Intrusion Detection System (IDS) that detects GPU-based DoS attacks and triggers multimodal alerts inside XR applications. Across three pilot studies in healthcare and training scenarios, we simulated these attacks and examined how users reacted to our security warning. To comprehensively evaluate user responses to XR security warnings, we combined self-report and behavioural measures: the Intrusion Detection System Questionnaire (IDSQ) [10] assessed users’ perceived clarity and actionability of alerts, the Self-Assessment Manikin (SAM) [11] captured emotional arousal and valence, and observed participant behavior was categorized into one of three response types: immediate compliance, clarification seeking, or inaction. Through this process, we derive three key design lessons. First, visual alerts must remain salient despite XR-specific constraints like occlusion and dynamic gaze. Second, integrating visual and auditory cues ensures detection even under high task load. Third, the urgency conveyed must be calibrated consistently across channels to avoid misinterpretation. Together, these findings offer concrete guidance for building attention-resilient, user-centred XR security notifications.

II. BACKGROUND

A. Cybersecurity Risks and XR

As XR technologies become more integrated into everyday applications, they not only present a growing target for cyberattacks [12], but also offer new opportunities to support

timely and context-aware cybersecurity interventions [13]. These systems are vulnerable to common threats like malware, Remote Code Execution (RCE), and data breaches, which can compromise user privacy and system integrity across both AR and VR environments [14]. Attackers can exploit vulnerabilities in crucial XR parameters, including network protocols to display and sensor data, to manipulate or disrupt immersive experiences [15]. Such cyberattacks in AR and VR not only affect system functionality but also critically impact user well-being through psychological manipulation, and extensive privacy violations [16]. While a wide range of cyber threats can target XR environments, our research focuses specifically on Denial-of-Service (DoS) attacks. Unlike data theft or injection attacks, DoS attacks directly target XR’s real-time performance, exploiting GPU and network vulnerabilities to cause frame drops and lag that disrupt immersion and trigger VR sickness [1]. Moreover, attacks that mimic information overload or create persistent interruptions can diminish the value of security alerts themselves, causing users to tune out important warnings over time [3]. Given these risks, we argue that DoS attacks represent a critical vector for research, especially when studying how users perceive and respond to cybersecurity threats in immersive systems.

B. Human Factors in Time-Critical Warnings

Given that XR cyberattacks like Denial-of-Service incidents severely degrade user experience and well-being, making rapid threat perception and reaction fundamental, effective time-critical warnings in immersive environments must be designed with understanding of human factors. For instance, Li et al. explored for the urgent notifications, users prefer to focus on the prominent displayed warning content rather than referring to the relevant context [17]. Visuri et al. found that users primarily judge warnings by their perceived significance and alignment with personal priorities, rather than solely by their arrival time [7]. Furthermore, Abello et al. demonstrated that even minor design modifications in warnings, such as changes to layout or colour, can significantly influence user likelihood of taking action [3]. Additionally, Louis and Zeng used eye tracking to propose dynamic supraliminal and subliminal warnings that optimize alert timing and placement without disrupting task flow [18]. Alternative modalities like affective haptics have also been explored to improve urgency perception and reduce reliance on saturated channels, such as in Do et al.’s wrist-based “Spidey Sense” system for cybersecurity alerts [19]. Collectively, these studies show that designing time-critical notifications requires conveying urgency and importance to prompt action, beyond simply delivering messages faster. Meaning that for XR, alerts must be developed to effectively capture the user’s attention and prompt appropriate responses within immersive and often highly demanding contexts.

Automotive interface design offers valuable XR parallels, as both domains require responses under high cognitive visual load. Yamin’s (2024) research on in-car display visibility provides relevant color and legibility guid-

ance [5]. Complementing this, studies by Betancur (2024) on head-up displays for automotive warnings found that combination of color/placement/duration lead to faster recognition/reactions [4].

C. Human Factors in Mobile Notification Design

Beyond urgency, effective XR security warnings depend on how users interpret and prioritize notifications based on their goals, attention, and context. To inform this, we draw on mobile HCI research, which offers insights into content framing, spatial placement, and cognitive interference.

For instance, studies on spatial relative positioning suggest that placing notifications away from a user’s immediate visual focus can be less disruptive while still ensuring visibility when needed [2]. Moreover, A notification’s content, frequency, and design format collectively influence user action; subtle variations in elements like color, animation, and urgency levels significantly affect whether users acknowledge or ignore alerts [3]. Crucially, understanding how users balance their internal goals with environmental context and the perceived urgency of an alert is fundamental to creating systems that elicit appropriate and timely user responses [20]. Research by [6] further confirms that alert perception critically depends on timing and situational context.

While valuable, these insights stem from planar, interruptible contexts like smartphones and desktops. In XR, where attention is spatial, embodiment is high, and user focus is deeply immersive, existing notification strategies must be re-evaluated. Critically, while multimodal alerts have been studied in AR, mobile, and automotive domains, there is currently no research directly examining how users perceive and respond to security-specific warnings in XR environments. This absence motivates our investigation into whether existing design principles generalize—or fail—within immersive, time-critical XR security scenarios.

D. Multimodality in Notification Design

Visual alerts alone are often insufficient in XR, where users may be deeply focused or looking away from on-screen cues [8]. Research across XR mobile and automotive domains consistently shows that combining visual elements with spatialized audio significantly improves detection rates and response times [21], [22]. For instance, in safety-critical automotive contexts, audio-visual combinations enhanced urgency perception and action speed [23]. In AR guidance tasks, pitch variation was used to direct attention effectively [24], while ambient and modulated cues have been shown to improve awareness without overwhelming the user [9].

Together, these findings underline the value of multimodal alerts, especially in immersive environments where perceptual bandwidth is limited and user attention is fragmented. This motivates our design of always-perceptible XR security warnings that integrate auditory and visual cues to ensure robust alert detection, even under high cognitive load or degraded rendering conditions from a DoS attack.



Fig. 1: Visual representation of the experimental setup across the three pilot sites. Left: Pilot 1—Versilia Hospital (Lido di Camaiore, Italy); Centre: Pilot 2—FACTOR (Valencia, Spain); Right: Pilot 3—Biotech campus, École polytechnique fédérale de Lausanne (EPFL), Geneva, Switzerland.

E. Background Summary

XR’s unique perceptual and interaction characteristics render traditional notification paradigms insufficient, particularly during attack-degraded system conditions. Our work addresses this by developing and evaluating a multimodal alert system tailored for immersive, time-sensitive XR security contexts.

III. SYSTEM

To investigate user responses to cybersecurity threats in XR we developed a modular Unity package that simulates a GPU-based DoS attack and renders a real-time in-headset security warning. This Unity package continuously monitors runtime performance and extracts a set of low-level system features associated with rendering performance degradation. These telemetry features are transmitted in real-time to a remote Intrusion Detection System (IDS), which processes the data using an unsupervised anomaly detection model. Based on the IDS classification, the Unity package triggers a multimodal security warning if an active GPU-based attack is detected. This end-to-end system allows for a closed-loop simulation of an attack scenario, including both the attack execution and the user-facing response mechanism, all within an immersive XR environment.

A. GPU-Based DoS Attack

Attacks such as denial of service (DoS) can prevent users from accessing a VR environment seamlessly, disrupt social presence, and potentially lead to VR sickness. To test our XR security warning design, we developed GPU-based cyberattacks inspired by the approach of Odeleye et al. [1]: GPU-based attack simulation was implemented within a Unity VR environment using a compute shader. Odeleye et al’s malware leveraged OpenGL to overload the GPU with large textures and long-running operations, whereas our implementation exploits Unity’s compute shader pipeline to simulate a denial-of-service scenario in an interactive XR setting. Specifically, we designed a dynamic shader that performs high-frequency wave calculations on a resizable RenderTexture, with resolutions up to $10,560 \times 10,560$ pixels. We built this scenario as a Unity executable and assumed a post-compromise model, in which an attacker can execute arbitrary Unity code on the target system. We could then evaluate GPU-based threats across a

wider range of consumer and enterprise XR devices. As the shader animates over time, it monopolizes GPU resources, causing significant frame drops and increased latency within the VR runtime. Similar to the findings in [1], we observed dropped frames, judder, and system instability. Our approach shows that even high-level rendering APIs in game engines can be abused to mount GPU resource exhaustion attacks that degrade or crash immersive experiences. A key motivation behind using Unity and compute shaders is to ensure cross-platform compatibility, allowing our malware simulation to be deployed on a variety of stand-alone XR headsets, including the Meta Quest 3 and Microsoft HoloLens 2.

B. Intrusion Detection System (IDS)

To detect GPU-based attacks in immersive environments, we adopt an unsupervised anomaly detection approach based on the Isolation Forest algorithm, following the method introduced by Odeleye et al. [1]. We monitor the same set of low-level system metrics derived from the VR runtime: average framerate, framerate standard deviation, average frametime, frametime standard deviation, and framerate entropy change. These features are computed using a two-second sliding window, producing time-series data suitable for early-stage anomaly detection. The IDS is implemented as a standalone Python service, containerised using Docker and deployed on a separate server. It communicates with the Unity application via a lightweight socket-based protocol, receiving live telemetry and returning warning flags when anomalous behaviour is detected. We refer readers to the original paper for a detailed description of the feature extraction process and model configuration.

C. Multimodal Security Warning Design

We designed our XR security warning (see Figure 2) to integrate synchronized visual and auditory cues, ensuring salience even under distraction or degraded rendering [25]. This aligns with prior work showing visual alerts alone are often missed due to occlusion or high task load [4], [7], [8]. To improve detection and reaction times, we added spatialized audio using Unity’s Audio Source with 3D positioning [21], [22], and anchored the visual warning to the user’s gaze using a billboard approach—consistent with strategies shown to enhance noticeability [26]. We implemented the warning using a simple



Fig. 2: Screenshots showing the visual appearance of cybersecurity warnings across XR environments in Pilots 1, 2, and 3.

and consistent layout to support interpretability, aligning with work that emphasizes the need for clear, structured, and user-friendly explanations in cybersecurity warnings [27].

Our visual design draws on research in time-critical notification. We used a green background and yellow text for fast recognition in daylight [4], [5], with white outlines for legibility and a restricted colour palette to reduce cognitive load [5]. The warning appears at eye level, persists for over two seconds, and uses a 24 pt font for clarity [4], [5]. Given the severity of GPU-based DoS attacks—known to impair system performance and induce VR sickness—we calibrated the audio warning for high urgency [24].

While some research supports user control (e.g., snoozing [5], [23]), our design intentionally restricts such options to enforce immediate action. Although timing during low cognitive load improves acceptance [2], we prioritized immediacy upon threat detection.

D. Pilot Applications

Our custom Unity package—including the GPU-based DoS attack, IDS, and immersive security warning—was integrated into each pilot application [28]–[30]. Pilots 1 and 2 were deployed on the Microsoft HoloLens 2 (AR), while Pilot 3 was a VR application deployed on the Meta Quest 3 (see Figure 2).

1) *Pilot 1 – Rehabilitation Scenario:* This pilot uses the application described by Sarri et al. [28], which explores embodied augmented reality for lower limb rehabilitation. In this rehabilitation setting, participants perform a “grasp and place” task, reaching for a virtual object and placing it on a dynamically positioned shelf. The task is designed to simulate upper limb rehabilitation, with difficulty adjusting to the patient’s progress (see Figure 2, left).

2) *Pilot 2 – VR Training Scenario:* This pilot uses the industrial safety training application presented by Mula et al. [29], which leverages extended reality technologies to support workplace training and risk awareness. Participants engage with immersive educational content and interactive questionnaires aimed at reinforcing workplace safety protocols. The application simulates standard safety procedures and risk identification tasks (see Figure 2, centre).

3) *Pilot 3 – Motivational Therapy Scenario:* This pilot uses the motivational therapy scenario introduced by Greci et al. [30], which targets patients with apathy following brain injury. The VR scenario combines motivational coaching with motor tasks. Participants reach toward multiple targets in

a gamified environment intended to encourage engagement. Unlike the AR-based Pilots 1 and 2, this is a fully immersive VR application (see Figure 2, right).

IV. EXPERIMENT

This study employed a repeated-measures design where participants experienced a GPU-based malware attack and a subsequent security warning from an Intrusion Detection System (IDS) during a continuous AR session. The malware attack degraded the AR experience through frame rate drops and jitter, which the IDS detected, triggering the warning. User responses to the security warning were evaluated by measuring security warning acceptance via the Intrusion Detection System Questionnaire (IDSQ) and by categorizing the effectiveness of their subsequent actions. The emotional responses of the participants to the warning were further assessed using the Self-Assessment Manikin (SAM), focusing on valence and arousal levels.

A. Participants

A total of 46 participants volunteered across the three pilots: 22 at *Ospedale di Versilia* in Viareggio, Italy (Pilot 1), 9 at *Factor* facilities in Valencia, Spain (Pilot 2), and 15 at *EPFL Campus* in Geneva, Switzerland (Pilot 3).

B. Procedure

In each pilot, healthy participants engaged in a scenario-specific VR task reflecting real-world applications of XR in rehabilitation, training, or therapy (see Section III-D). These tasks constituted the “normal phase” of the study, during which an unannounced cyberattack was triggered (t_0). Before the task, participants completed demographic questionnaires and a short familiarization session with the VR system (5 minutes). Once the attack was detected by the Intrusion Detection System (IDS), a multimodal security warning was displayed (t_1), and participants’ behaviour and action latency were recorded. After the session, participants completed the IDSQ and SAM questionnaires, followed by a debriefing.

C. Questionnaires and Participants Response categorization

Table I presents the Intrusion Detection System Questionnaire (IDSQ) [10], used to assess security warning acceptance via a 5-point Likert scale, and the Self-Assessment Manikin (SAM) [11], used to measure emotional responses (valence and arousal). Additionally, participant actions were categorized based on their response to the warning (see I Categorization).

TABLE I: Measures: IDSQ, SAM, and Action Categorization

Measure	Description / Scale
IDSQ	
S1	The warning message effectively captured my attention.
S2	The warning messages are clear and understandable. (1 = Strongly Disagree to 5 = Strongly Agree)
S3	I understand the actions I need to take in response to the alerts. (1 = Strongly Disagree to 5 = Strongly Agree)
S4	The warning message displayed early enough. (1 = Strongly Disagree to 5 = Strongly Agree)
S5	The alerts generated are relevant and actionable. (1 = Strongly Disagree to 5 = Strongly Agree)
SAM	
Valence	(1 = Unhappy, unpleasant, dissatisfied ... 5 = Happy, pleasant, satisfied)
Arousal	(1 = Relaxed, calm, sluggish, dull ... 5 = Excited, agitated, wide awake, aroused)
Categorization	
R1	Followed the alert Participant immediately acted on the warning (e.g., removed headset).
R2	Asked for clarification Participant sought confirmation from the experimenter before taking action.
R3	Ignored the warning Participant took no action or showed no response to the alert.

V. RESULTS

A. IDS Detection Latency and Accuracy

Although our work does not focus on advancing intrusion detection techniques, the effectiveness of in-headset warnings depends on the IDS triggering alerts with minimal delay and high accuracy. IDS performance was evaluated by measuring the *detection latency*, defined as the time between attack onset (t_0) and alert issuance (t_1). System logs were also analysed for false positives, i.e., alerts raised in the absence of an actual attack. Across all three pilots, the system successfully detected GPU-based attacks and triggered warnings within 0.27 to 1.25 seconds—fast enough to support timely user responses before significant performance degradation occurred.

- **Pilot 1:** Median latency = 0.27 s (SD = 0.03 s)
- **Pilot 2:** Median/mean latency \approx 1.0 s (SD = 0.45 s)
- **Pilot 3:** Mean latency = 1.25 s (SD = 0.43 s)

Importantly, no false positives were observed, and transient delays (1–2 seconds) in a few cases did not interfere with the warning’s effectiveness. These results confirm that the IDS provided a reliable and low-latency signal for triggering warnings, and thus did not limit user response or the validity of our findings.

B. Participants’ comprehension of Security Warnings (IDSQ)

To analyse IDSQ responses and assess user perceptions of the warning, we conducted a Wilcoxon signed-rank test to compare participant responses against the neutral midpoint (3) on the Likert scale. Visual distributions of responses are shown in Figure 4, and corresponding statistical results are summarised in Table II.

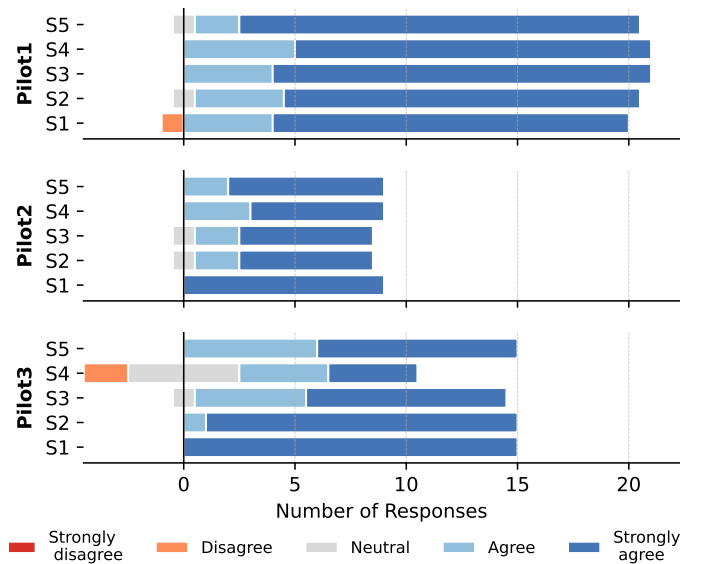


Fig. 3: Responses to the IDSQ across three pilot studies. The y-axis lists the five IDSQ items (S1–S5), each representing a statement rated by participants. The x-axis shows the number of responses on a 5-point Likert scale.

- **Pilot 1** ($n = 22$): All items showed p between .005 and .010, with rank-biserial correlations between 0.974 to 1.000, despite the small sample size.
- **Pilot 2** ($n = 9$): All five items showed $p < .001$ with rank-biserial correlations of 1.000 meaning all answers are equal or above the midpoint (3).
- **Pilot 3** ($n = 15$): All items showed significant positive deviation from the neutral midpoint. Three items (S2, S3, S5) had $p < .001$; S4 reached $p = .018$. S1 showed no variance (all participants gave the same rating), indicating unanimous agreement.

Results from all three pilots consistently showed strong agreement with each statement. Most participants rated the alerts as clear, timely, and actionable.

TABLE II: Wilcoxon Signed-Rank Test Results for IDSQ Scores Compared to Neutral (3)

Item	Pilot 1		Pilot 2		Pilot 3	
	p	r	p	r	p	r
S1	.005	0.974	–	–	–	–
S2	.010	1.000	< .001	1.000	< .001	1.000
S3	.010	1.000	< .001	1.000	< .001	1.000
S4	.007	1.000	< .001	1.000	.018	0.745
S5	.006	1.000	< .001	1.000	< .001	1.000

Note. Wilcoxon signed-rank test. The alternative hypothesis specifies that the median is greater than 3. p = significance level; r = rank-biserial correlation. S1 in Pilots 2 and 3 had zero variance and were excluded.

C. Participants Emotional Response to Security Warning (SAM, pilot 3 only)

To evaluate participants’ affective response to the security warning, we conducted Wilcoxon signed-rank tests comparing

SAM ratings against the neutral midpoint of 3. **Arousal ratings** were significantly higher than the midpoint ($V = 55.000$, $p = .001$), indicating heightened alertness or excitement in response to the warning. **Valence ratings** did not significantly differ from the midpoint ($V = 3.000$, $p = .065$), suggesting a neutral emotional tone. Rank-biserial correlations are reported as effect sizes. These findings suggest that the warning was highly attention-grabbing, as reflected in the elevated arousal ratings, but did not trigger strong emotional valence in either direction. Participants remained physiologically alert without experiencing the warning as distressing or emotionally charged, indicating it was perceived as clear, engaging, and actionable rather than threatening.

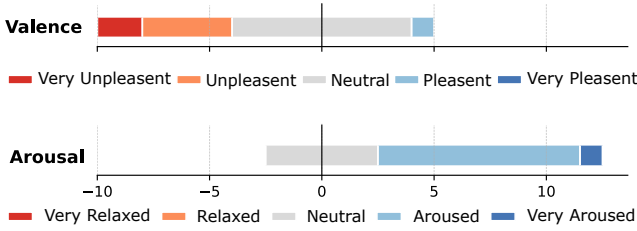


Fig. 4: Diverging bar charts visualising responses to the SAM across pilot 3 only. The x-axis shows the number of responses on a 5-point Likert scale .

TABLE III: Wilcoxon Signed-Rank Test Results for SAM

Dimension	V	p	Rank-Biserial Correlation	SE (RBC)
Valence	3.000	.065	-0.786	0.399
Arousal	55.000	.003	1.000	0.342

Note. For the Wilcoxon test, the effect size is given by the matched rank-biserial correlation. The alternative hypothesis specified that the median is different than 3.

D. Participant Responses to Security Warning

To evaluate how participants responded to the IDS-triggered security alert, their actions were observed and categorized into three response types (R1–R3) as defined in Table I.

As shown in Table IV, most participants (77–100%) across all pilots either responded directly or asked for confirmation, indicating that the warning was generally recognized and taken seriously. These behavioural findings align closely with the IDSQ results, suggesting that the alert was not only understood, but also motivated action or engagement from most users.

TABLE IV: Distribution of participant responses (R1–R3) to IDS-triggered security warnings across pilot studies.

Pilot	Sample Size (n)	R1 (%)	R2 (%)	R3 (%)
Pilot 1	22	27.3	50.0	22.7
Pilot 2	9	22.2	66.7	11.1
Pilot 3	14	35.7	64.3	0.0

E. User Attention to the Multimodal Security Warning

At the end of each session, participants were asked which component of the security warning—audio, visual, or both—first captured their attention. This question aimed to evaluate the effectiveness of the multimodal alert design in AR/VR environments across three pilot studies. The findings reveal distinct patterns. A substantial proportion of participants reported noticing the visual component first: 47.8% in Pilot 1, 55.6% in Pilot 2, and 47.1% in Pilot 3. These participants often described the visual warning as prominently placed and immediately noticeable. Meanwhile, 43.5% in Pilot 1, 44.4% in Pilot 2, and 35.3% in Pilot 3 identified the audio cue as their initial point of awareness, noting that it effectively cut through visual distractions. A smaller but significant group reported perceiving both modalities simultaneously: 8.7% in Pilot 1 and 17.6% in Pilot 3, while no participants in Pilot 2 selected this option. This group emphasized the reinforcing role of synchronized audio and visual elements. These results confirm that multimodal alerts improve the likelihood of timely user recognition and response, validating the effectiveness of combining audio and visual cues in immersive environments.

TABLE V: User Attention Modality Saliency Across Pilots

Attention Modality	Pilot 1 (%)	Pilot 2 (%)	Pilot 3 (%)
Visual first	47.8	55.6	47.1
Audio first	43.5	44.4	35.3
Both simultaneously	8.7	0.0	17.6

VI. LESSONS LEARNED

A. Lesson 1: Maintaining Visual Saliency for Security Warnings in XR Spaces

The dynamic nature of user movement in immersive XR environments—where individuals actively explore and shift their head/gaze within a 3D space—undermines the reliability of static visual cues for maintaining awareness of critical security alerts. This limitation arises from factors such as restricted peripheral vision, visual distraction from the primary task, and potential occlusion by other 3D objects.

In Pilot 1, practical visibility issues emerged, where visual elements were sometimes obscured by assets from the XR application itself. This necessitated technical adjustments, including the use of custom shaders to render the warning above other content and a head-locked orientation system (billboarding) to ensure the warning consistently appeared in front of the user’s view.

After implementing these changes, we observed a noticeable improvement in the saliency of the visual warning. In Pilot 2, a higher proportion of participants (55.6%) reported noticing the visual element first, compared to 47.8% in Pilot 1. By Pilot 3, despite increased multimodal integration, visual-first responses remained high (47.1%), suggesting that the technical adjustments improved the visibility and perceptual prominence of visual alerts across sessions (see Table V).

However, in Pilot 2, a new problem emerged: while users more consistently noticed the visual alert, a few perceived

it as intrusive. One participant remarked, *"It was like too much in my face"* (P6), suggesting that overly aggressive visual placement may disrupt user experience or cause discomfort. Furthermore, XR-specific technical constraints must be acknowledged. In particular, developers working within platforms such as Oculus on Android cannot trigger system-level UI overlays, requiring that alerts be rendered entirely within the application's own context. This highlights the importance of managing occlusion risks, layering, and modality coordination within the app itself.

B. Lesson 2: Visibility Requires Multimodal Cues

While prior work emphasizes the benefits of multimodal alerts [9], [17], our findings provide empirical validation of this principle in immersive XR settings. Visual-only alerts often failed to capture attention, particularly when users were facing away from the notification or deeply focused on the primary task. This was evident across all three pilots, where participants reported noticing either the audio, the visual, or both components of the alert (see Section V-E, Table V).

Even after improving the visibility and positioning of the visual warning, audio cues remained critical for capturing attention, especially under conditions of distraction or occlusion. These patterns reflect variability in user focus and perceptual availability, supporting Wickens' Multiple Resource Theory [31], which suggests that users draw on separate cognitive channels for visual and auditory input.

Multimodal alerts are therefore not just a redundancy mechanism, but a necessary adaptation to the perceptual and attentional demands of immersive environments. Audio often served as the initial trigger: as noted by P8 in Pilot 2, *"The audio prompted me to look at the warning."* In practice, Alert designers cannot assume visual cues will be perceived, even when centrally placed. Our results extend prior theory by showing that multimodal alerting improves noticeability in XR specifically when visual bandwidth is exceeded or task immersion is high.

C. Lesson 3: Multimodal Urgency Balance

While striving for balanced urgency across modalities like audio and visual is crucial for effective threat communication, poorly designed alerts can communicate different levels of importance and urgency across the modalities. As feedback from Pilot 3 indicated, the visual warning sometimes lacked the urgency conveyed by the audio, with participants noting that *"audio is more alarmist, visual part provide more information"* (P12). This imbalance was further highlighted by observations that the visual design, such as the use of green color, might reduce the perceived urgency, with one participant suggesting that *"Sign Green is okay, but it would convey higher urgency if it was Red"* (P4). This highlights the need to carefully consider how different modalities contribute to the overall perception of threat. Furthermore, accessibility considerations are paramount; for instance, deaf users relying solely on visual cues might perceive a different level of urgency. Future iterations should explore design strategies that balance urgency

across modalities while ensuring accessibility. Our findings contribute to the literature on security alerts by highlighting the importance of calibrating urgency not only for the overall warning but also for its individual components across different modalities (e.g., audio and visual). This nuanced approach, supported by direct user feedback indicating discrepancies in perceived urgency between audio and visual elements, suggests that an imbalance in perceived urgency between warning components can influence user response, underscoring the need for a modality-specific calibration strategy beyond the global urgency level of the alert itself.

VII. LIMITATION AND FUTURE WORK

A limitation of our current study is the exclusive reliance on audio and visual modalities for XR security warnings. Future research could explore incorporating haptic feedback to enhance the salience and effectiveness of alerts. Haptics, could reduce reliance on potentially overloaded visual and auditory channels [18]. Additionally, our evaluation primarily utilized self-reports (e.g., IDSQ) and emotional ratings via the Self-Assessment Manikin (SAM), as well as observed behaviors, future research could introduce quantitative metrics, such as real-time physiological indicators of stress (e.g., heart rate variability or galvanic skin response). This combined approach would enable a deeper understanding of both the self-reported and physiological dimensions of user experience, leading to more nuanced and effective warning system designs in XR contexts.

VIII. CONCLUSION

This paper presented a mixed-methods investigation into how users perceive and respond to XR security warnings during immersive cyberattacks. Through three pilot studies, we showed that multimodal alerts significantly improve noticeability and response effectiveness. Our findings underscore the need for salience-aware, urgency-balanced, and modality-sensitive designs, offering practical guidance for future XR cybersecurity systems.

REFERENCES

- [1] B. Odeleye, G. Loukas, R. Heartfield, and F. Spyridonis, "Detecting framerate-oriented cyber attacks on user experience in virtual reality," in *Proceedings of the 1st International Workshop on Security for XR and XR for Security (VR4Sec)*. Vancouver, B.C., Canada (virtual): VR4Sec Workshop Organizers, August 2021, pp. 1–5. [Online]. Available: <http://bura.brunel.ac.uk/handle/2438/23880>
- [2] S. Hueber, E. Jang, and J. Borchers, "Attentive notifications: Minimizing distractions of mobile notifications through gaze tracking," in *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction, MobileHCI 2023 Companion*. Association for Computing Machinery, Inc, 9 2023.
- [3] H. M. Abello, M. B. Badiola, M. J. Custer, L. B. Fausto, P. J. Leonida, D. B. Yongco, and J. A. Deja, "Simon says: Exploring the importance of notification design formats on user engagement," *arXiv preprint arXiv:2412.00531*, 2024. [Online]. Available: <http://arxiv.org/abs/2412.00531>
- [4] J. A. Betancur, H. Vargas, C. Sanchez, and F. Merienne, "Visual guidelines integration for automotive head-up displays interfaces," *International Journal on Interactive Design and Manufacturing*, 4 2024.

- [5] P. A. Yamin, J. Park, and H. K. Kim, "In-vehicle human-machine interface guidelines for augmented reality head-up displays: A review, guideline formulation, and future research directions," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 104, pp. 266–285, 7 2024.
- [6] X. J. Chang, F. H. Hsu, E. C. Liang, Z. Y. Chiou, H. H. Chuang, F. C. Tseng, Y. H. Lin, and Y. J. Chang, "Not merely deemed as distraction: Investigating smartphone users' motivations for notification-interaction," in *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, 4 2023.
- [7] A. Visuri, N. van Berkel, T. Okoshi, J. Goncalves, and V. Kostakos, "Understanding smartphone notifications' user interactions and content importance," *International Journal of Human Computer Studies*, vol. 128, pp. 72–85, 8 2019.
- [8] P. C. Fernandes, F. P. Silvas, G. H. M. Rocha, B. N. Coelho, B. H. Adachi, and S. Delabrida, "Exploring human interaction in virtual reality: An experience report on users with and without visual impairment," in *ACM International Conference Proceeding Series*. Association for Computing Machinery, 12 2024.
- [9] S. Sigethy, S. Mayer, and C. Schneegass, "Learning in the wild-exploring interactive notifications to foster organic retention of everyday media content," in *Behaviour and Information Technology*. Taylor and Francis Ltd., 2024.
- [10] L. F. Cranor, "A framework for reasoning about the human in the loop," in *Proceedings of the 1st Conference on Usability, Psychology, and Security*, ser. UPSEC'08. USA: USENIX Association, 2008.
- [11] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0005791694900639>
- [12] S. Dastgerdy, "Virtual reality and augmented reality security: A reconnaissance and vulnerability assessment approach," *arXiv preprint arXiv:2407.15984*, 2024. [Online]. Available: <https://arxiv.org/abs/2407.15984>
- [13] A. Kanaoka and T. Isohara, "Enhancing smishing detection in ar environments: Cross-device solutions for seamless reality," in *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2024, pp. 565–572.
- [14] M. El-Hajj, "Cybersecurity and privacy challenges in extended reality: Threats, solutions, and risk mitigation strategies," *Virtual Worlds*, vol. 4, p. 1, 12 2024. [Online]. Available: <https://www.mdpi.com/2813-2084/4/1/1>
- [15] B. Odeleye, G. Loukas, R. Heartfield, G. Sakellari, E. Panaousis, and F. Spyridonis, "Virtually secure: A taxonomic assessment of cybersecurity challenges in virtual reality environments," *Computers & Security*, vol. 124, p. 102951, 2023.
- [16] H. Hadan, L. Zhang-kennedy, L. E. Nacke, L. Choong, and L. Zhang-Kennedy, "Deceived by immersion: A systematic analysis of deceptive design in extended reality," *ACM Comput. Surv.*, vol. 56, p. 25, 2024. [Online]. Available: <https://doi.org/10.1145/3659945>
- [17] T. Li, J. K. Haines, M. F. R. D. Eguino, J. I. Hong, and J. Nichols, "Alert now or never: Understanding and predicting notification preferences of smartphone users," *ACM Transactions on Computer-Human Interaction*, vol. 29, 1 2023.
- [18] L. Louis, "Reduction of cognitive load in immersive virtual reality with multisensory cues," *Tech. Rep.*, 2024. [Online]. Available: <https://louis.uah.edu/uah-theses/673>
- [19] Y. Do, L. T. Hoang, J. W. Park, G. D. Abowd, and S. Das, "Spidey sense: Designing wrist-mounted affective haptics for communicating cybersecurity warnings," in *DIS 2021 - Proceedings of the 2021 ACM Designing Interactive Systems Conference: Nowhere and Everywhere*. Association for Computing Machinery, Inc, 6 2021, pp. 125–137.
- [20] N. Gao, W. Shao, M. S. Rahaman, and F. D. Salim, "N-gage: Predicting in-class emotional, behavioural and cognitive engagement in the wild," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, 9 2020.
- [21] K. Yu, D. Du, D. Yu, J. Zhi, Y. Wang, and C. Jing, "Effects of a color gradient and emoji in ar-hud warning interfaces in autonomous vehicles on takeover performance and driver emotions," *Traffic Injury Prevention*, vol. 25, pp. 714–723, 2024.
- [22] H. Lee and W. Woo, "Exploring the effects of augmented reality notification type and placement in ar hmd while walking," in *Proceedings - 2023 IEEE Conference Virtual Reality and 3D User Interfaces, VR 2023*. Institute of Electrical and Electronics Engineers Inc., 2023, pp. 519–529.
- [23] M. Wang, J. Parker, F. Zhang, and S. C. Roberts, "A simulator study assessing the effectiveness of training and warning systems on drivers' response performance to vehicle cyberattacks," *Accident Analysis and Prevention*, vol. 203, 8 2024.
- [24] R. Guarese, E. Pretty, A. Renata, D. Polson, and F. Zambetta, "Exploring audio interfaces for vertical guidance in augmented reality via hand-based feedback," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, pp. 2818–2828, 5 2024.
- [25] M. Wang, J. Parker, F. Zhang, and S. C. Roberts, "A simulator study assessing the effectiveness of training and warning systems on drivers' response performance to vehicle cyberattacks," *Accident Analysis & Prevention*, vol. 203, p. 107644, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0001457524001891>
- [26] C. Gantiva and C. Mejia-Orjuela, "Yellow warnings labels on top are more effective to discourage smoking initiation: An experimental online study," *American Journal of Health Education*, vol. 55, pp. 239–244, 2024.
- [27] V. A. Kazakova, J. D. Hwang, B. J. Dorr, Y. Wilks, J. B. Gage, A. Memory, and M. A. Clark, "Splain: Augmenting cybersecurity warnings with reasons and data," *arXiv preprint arXiv:2311.11215*, 2023. [Online]. Available: <http://arxiv.org/abs/2311.11215>
- [28] F. Sarri, P. Kasnesis, S. Symeonidis, I. T. Paraskevopoulos, S. Diplaris, F. Posteraro, G. Georgoudis, and K. Mania, "Embodied augmented reality for lower limb rehabilitation," in *CLIFE 2024 - Creating Lively Interactive Populated Environments*. The Eurographics Association, 2024. [Online]. Available: <https://diglib.org>
- [29] J. Mula, R. Sanchis, R. de la Torre, and P. Becerra, "Extended reality and metaverse technologies for industrial training, safety and social interaction," *IFAC-PapersOnLine*, vol. 58, no. 19, pp. 575–580, 2024.
- [30] L. Greci, F. Bosco, and V. Croce, "The social and human centered xr: Sun xr project," in *International Conference on Extended Reality*. Springer, 2023, pp. 223–231.
- [31] C. D. Wickens, "Multiple resources and performance prediction," *Theoretical Issues in Ergonomics Science*, vol. 3, no. 2, pp. 159–177, 2002. [Online]. Available: <https://doi.org/10.1080/14639220210123806>