Poster: Auditory Viewpoint Manipulation Attacks in VR: An Empirical Study

Kousei Otsuka
Toho University
6524002o@st.toho-u.ac.jp

Shodai Kurasaki
Toho University
7525001k@st.toho-u.ac.jp

Mayu Fujita
Toho University
6524010f@st.toho-u.ac.jp

Akira Kanaoka
Toho University
akira.kanaoka@is.sci.toho-u.ac.jp

I. Introduction

XR technologies directly affect human perception through sensory channels such as vision and audition, providing strong immersion and a sense of presence, while also introducing risks when such channels are exploited [1]. While prior work has discussed risks and the potential for attacks exploiting auditory stimuli [2], [3], only limited empirical validation exists [1]. In particular, spatial audio is commonly supported in XR devices and can naturally lead users to misattribute the source of sounds to the physical world [1].

We define Auditory Viewpoint Manipulation Attacks (AVMA) as the adversarial use of spatial-audio cues to steer users' attention and viewpoint direction in VR, and we empirically evaluate their feasibility and stealth. In our study, we design stimuli to elicit physical-world misattribution and test whether it promotes viewpoint steering. If an adversary can steer users' attention and head orientation toward a specific direction using spatial-audio stimuli, the impact may extend beyond task disruption and operational errors. It can also increase privacy/confidentiality leakage by raising the likelihood that outward-facing cameras or passthrough views capture sensitive surroundings in the sensor feed [4], including personal information in residential spaces [5]. Moreover, AVMA can be introduced through the software supply chain, and because auditory stimuli can act on users in a natural manner [2], the attack may be difficult to detect or report subjectively and may persist unnoticed.

However, it remains unclear to what extent AVMA succeeds in practice, whether users can recognize it as adversarial, and under what conditions it is suppressed. To address these questions, we conducted a user study with 30 participants to assess the effectiveness and stealth of AVMA.

We find that viewpoint manipulation can occur without physical-world misattribution. From semi-structured interviews (N=30), half of the participants reported AVMA-induced viewpoint manipulation, and the stimuli were often not interpreted as an “ attack. ” Interview responses further suggested cognitive resistance to context mismatches in the surrounding environment, indicating that attacks relying on misattribution may be suppressed under certain conditions.

II. Attacker Model

The attack considered in this study targets users playing a single-user game in VR. An adversary-controlled in-app component presents spatially structured auditory stimuli to steer the user's attention and head orientation.

TABLE I
AVMA stimuli used in the study

Stimulus	VR contextual dependence	Potential reaction
PV: Phone Vibration Sound	Non-dependent	Everyday response
GB: Glass Breaking Sound	Non-dependent	Defensive response
HV: Human Voice Sound	Non-dependent	Social response
BC: Baby Crying Sound	Non-dependent	Affective response
DC: Dependent Context Sound	Dependent	In-game response

a) Assumptions: VR applications can obtain outward-facing camera video (passthrough) via platform-provided APIs, subject to permissions [4].

b) Adversary objective: The adversary aims to manipulate the user ’ s head orientation to increase the exposure of sensor-captured surroundings and enable large-scale privacy leakage, including personal information in private residential environments [5].

c) Attack vector via the software supply chain: We consider a supply-chain setting in which an adversary distributes a malicious asset that appears to be a benign 3D object, and a developer unintentionally incorporates it into a project. Given widely used asset distribution platforms such as the Unity Asset Store, such assets can be disseminated at scale and subsequently reach a broad user population.

d) Potential impact: Because the attack can propagate through the supply chain and may be difficult for users to interpret as adversarial, it can persist over time and increase privacy exposure risk when deployed widely.

III. Study Overview

We conducted a user study in which participants were exposed to AVMA stimuli while performing a VR game task. All participants were exposed to all five stimuli in a counterbalanced order. Participants were 30 students from the authors' university (mean age = 21, SD = 2.017).

a) Task: We developed a Tetris-like falling-block VR puzzle game in Unity and presented AVMA stimuli during a 10-minute gameplay session using a Meta Quest 3 HMD.

b) Attack Stimuli: We used five auditory stimuli in total: four context-non-dependent sounds intended to be misattributed as originating from the physical environment, and one context-dependent in-game sound effect (Table I).

TABLE II
KEY RESULTS

Outcome	Value
Participants with viewpoint manipulation	15/30 (50.0%)
Manipulated by context-non-dependent stimuli	13/30 (43.3%)
Recognized as adversarial/malicious	8/30 (26.7%)

TABLE III
CONTINGENCY TABLE BETWEEN VIEWPOINT MANIPULATION AND SOURCE JUDGMENT (PHYSICAL ENVIRONMENT VS. VR)

Viewpoint manipulation	Physical environment (count)	VR (count)
Yes	5	17
No	17	71

One stimulus was recorded by the authors, and the remaining four were obtained from license-compliant free sound-effect sources.

c) Ethical Considerations: This study was approved by the authors' institutional review board (IRB).

IV. KEY RESULTS AND DISCUSSIONS

A. Key Results

a) KR1: Feasibility of viewpoint manipulation: 15 out of 30 participants self-reported at least one instance of viewpoint manipulation (Table II). In addition, 13 out of 30 participants reported manipulation induced by at least one of the context-non-dependent stimuli (PV, GB, HV, BC), which were designed to be misattributed as physical-world sounds. These results indicate that audio outputs by XR devices can influence users' head orientation, suggesting that the attack surface of XR systems may extend beyond conventional network intrusions and malware to include human-mediated channels.

b) KR2: High stealth and low adversarial attribution: 8 out of 30 participants interpreted the stimuli as adversarial (Table II). Moreover, participants often attributed the stimuli to internal factors such as in-game system sounds or bugs rather than to an external attacker. This stealth property can undermine user-driven mitigation/reporting, and may allow attacks to persist and accumulate over time as detection and response are delayed.

c) KR3: Limited physical-world misattribution: Our attack scenario assumes that users misattribute VR-originated sounds as physical-world sounds and, as a result, change their viewpoint. However, across 110 stimulus instances, source judgment (physical environment vs. VR) and viewpoint manipulation (yes/no) were independent ($\chi^2(1, N = 110) = 0.04$, $p = .84$, $\phi \approx .02$, Table III). Consistently, manipulation rates were comparable between stimuli judged as originating from the physical environment (22.7%, 5/22) and those not misattributed (19.3%, 17/88). These findings suggest that viewpoint manipulation can occur without explicit physical-world misattribution.

d) KR4: Security-relevant task impact: 17 participants reported reduced concentration, 5 paused the task, and 7 reported operational errors during gameplay. These findings suggest that auditory stimuli can impair task performance and induce interruption or errors even when viewpoint manipulation is not the primary outcome. That such errors were observed even in a relatively simple game task implies that the impact may be amplified in high-load or high-assurance settings.

B. Discussion

a) Implications: Our results indicate that physical-world misattribution is not a prerequisite for viewpoint manipulation. Even when users attribute a sound to the VR system, salient auditory cues can still reorient attention and trigger head movements. Interview responses further suggest cognitive resistance to context mismatches: when the implied situation conflicted with the surrounding environment, participants often rejected a physical-world interpretation. Because viewpoint manipulation can occur without explicit physical-world misattribution, Defenses that focus only on preventing misattribution or improving user awareness may miss attacks, motivating system-driven detection and mitigation.

b) Defense: A key characteristic of this attack is that it can infiltrate a wide population via the software supply chain while remaining highly stealthy, making user-driven detection and reporting unreliable. Accordingly, at the distribution layer, it is important to strengthen vetting and analysis for assets that include capabilities likely to expand the attack surface, such as audio output, sensor access, and external communications, and to enforce least-privilege dependency design based on explicit capability and permission declarations. At runtime, a promising direction is to audit consistency between audio-output logs and VR context logs that record virtual-space and content events, enabling detection and suppression of context-incongruent audio outputs without relying on user reports.

c) Future work: As future work, we will corroborate our qualitative findings with quantitative evidence by analyzing objective telemetry such as HMD head-pose and position logs. We also plan to develop methods for computing consistency between VR-context logs and audio-output events to better characterize when AVMA succeeds and how it can be detected.

ACKNOWLEDGMENT

This work was supported by JST CREST (JPMJCR22M4 and the AIP Challenge Program 2024) and by NICT, Japan.

REFERENCES

[1] Kaiming Cheng et al., “ Exploring User Reactions and Mental Models Towards Perceptual Manipulation Attacks in Mixed Reality, ” in Proc. USENIX Security '23 (32nd USENIX Security Symposium), 2023.

[2] Esmée H. A. de Haas et al., “ Deceiving Audio Design in Augmented Environments: A Systematic Review of Audio Effects in Augmented Reality, ” in Proc. ISMAR-Adjunct 2022 (IEEE Int'l Symp. on Mixed and Augmented Reality Adjunct), 2022.

[3] Kousei Otsuka et al., “ Auditory Stimulus Attack in XR: Stimulus Characteristics and Technical Background Considerations,” in Proc. MetaCom 2025 (3rd IEEE Int'l Conf. on Metaverse Computing, Networking and Applications), 2025.

[4] Meta Platforms, Inc., “ Passthrough Camera API Overview, ” https://developers.meta.com/horizon/documentation/spatial-sdk/spatial-sdk-pca-overview/, Meta Horizon OS Developers, 2025. (Accessed: 2026-1-20)

[5] Rachel McAmis et al., “ The Writing on the Wall and 3D Digital Twins: Personal Information in (not so) Private Real Estate, ” in Proc. USENIX Security '23 (32nd USENIX Security Symposium), 2023.

Auditory Viewpoint Manipulation Attacks in VR: An Empirical Study

(This work was supported by JST CREST (JPMJCR22M4 and the AIP Challenge Program 2024) and by NICT, Japan.)

©Kousei Otsuka, Shodai Kurasaki, Mayu Fujita, Akira Kanaoka (Toho University)

Introduction

- XR technologies directly affect human perception through sensory channels such as vision and audition, enabling strong immersion while also opening **new attack surfaces** when sensory channels are exploited [1].
- While prior work has discussed the potential for attacks using auditory stimuli [2,3], **empirical studies remain limited** [1].

AVMA (Auditory Viewpoint Manipulation Attack)

AVMA exploits spatial-audio stimuli in VR to **steer users' attention and head orientation**, and can spread at scale via the software supply chain to indiscriminate users.

RISK: **Privacy leakage** via unauthorized access to passthrough camera data (e.g., by abusing passthrough APIs [4]).

Objective: Empirically evaluate the **feasibility** and **stealth** of AVMA through a user study.

Research Overview Video (YouTube)

Attacker Model

Assumption

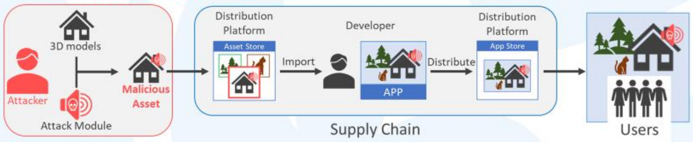
VR applications can obtain outward-facing camera/passthrough video via **platform-provided APIs** [4].

Objective

Manipulate head orientation to increase exposure of sensitive surroundings and enable privacy (e.g., personal information in residential spaces [5]) leakage.

Feature

AVMA malicious asset distributed via Unity Asset Store-style **supply chains.**



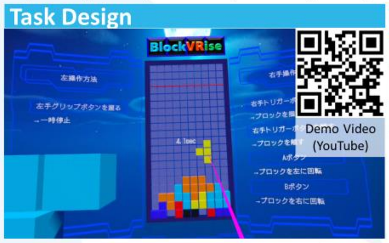
Impact

Stealthy propagation can lead to large-scale, persistent privacy/confidentiality leakage.

Study Overview

30 participants (students; mean age=21, SD=2.017), Meta Quest 3, 10-min gameplay.

Task Design



Demo Video (YouTube)

Tetris-like VR game with VR-specific interactions.

Auditory Stimuli (5 Stimuli) [*1]

PV: Phone Vibration Sound

GB: Glass Breaking Sound

HV: Human Voice Sound

BC: Baby Crying Sound

DC: Dependent Context Sound[*1]

Selected to elicit diverse reactions and be **plausible as physical-world sounds** (PV, GB, HV, BC,).

[*1] Sourced from free audio sites and recorded by the authors

Measures

Semi-structured interviews

- Viewpoint manipulation
- Physical-world misattribution
- Recognition of malicious intent

Ethical Considerations

- Authors' institutional **IRB** approval
- **Safe sound levels** based on NIOSH guidelines
- **Debriefing** to address deception

Key Results & Discussion [*2]

[*2] Results/discussion are based on subjective self-reports from semi-structured interviews.

Table1: Key Results (P=30).

Outcome	Value
Participants with viewpoint manipulation	**15/30 (50.0%)**
Manipulated by PV, GB, HV, BC	13/30 (43.3%)
Recognized as adversarial/malicious	8/30 (26.7%)

KR1: Feasibility of Viewpoint Manipulation

- 15/30 (50.0%) participants self-reported AVMA-induced viewpoint manipulation (Table 1).
- 13/30 (43.3%) reported manipulation from context-non-dependent stimuli (PV, GB, HV, BC), which are plausible as physical-world sounds (Table 1).

Spatial-audio can induce viewpoint manipulation in VR.

KR3: Limited Physical-World Misattribution

- Across N=110 stimulus instances, source judgment (physical environment vs. VR) and viewpoint manipulation (yes/no) were independent ($\chi^2(1, N=110)=0.04$, $p=.84$, $\phi\approx.02$) (Table 2).
- Manipulation rates were comparable for stimuli judged as physical-world sounds (22.7%, 5/22) and those not misattributed (19.3%, 17/88) (Table 2).

Viewpoint manipulation can occur without explicit physical-world misattribution.

Table2: Viewpoint Manipulation and Source Judgment (N=110).

Viewpoint manipulation	Physical environment	VR
Yes	5	17
No	17	71

KR2: High Stealth and Low Adversarial

- 8/30 (26.7%) participants labeled the stimuli as adversarial/malicious (Table 1).
- Most instead attributed them to in-app causes (e.g., sound effects or bugs), rather than an external attacker.

This stealth can evade subjective detection and delay response, allowing attacks to persist and accumulate.

KR4: Security-Relevant Task Impact

- 17 participants reported **reduced concentration**.
- 5 participants reported **pausing the task**.
- 7 participants reported **operational errors** during gameplay.

Attacks using auditory stimuli can disrupt task performance and may induce interruptions or errors.

The impact may be amplified in high-workload or high-assurance settings.

D1: Implications

- Viewpoint manipulation can occur even when users attribute sounds to the VR system, without physical-world misattribution.
- AVMA often evades subjective detection; user-driven recognition and incident reporting is unreliable.
- Thus, mitigations focused only on misattribution prevention or **user awareness may miss attacks**, motivating system-driven detection and mitigation.

D2: Defense

Supply chain: Strengthen vetting and analysis of assets whose capabilities expand the attack surface, such as audio output, sensor access, and external communications.

Least privilege: Enforce explicit capability/permission declarations and least-privilege dependency design.

Runtime: Audit consistency between audio-output events and VR-context logs, enabling detection/suppression of context-incongruent audio outputs without relying on user reports.

Future Work

- Validate self-reports with **HMD telemetry logs**. [Analysis in progress]
- Develop runtime detection VR-context–audio consistency.

REFERENCES

[1] Kaiming Cheng et al., "Exploring User Reactions and Mental Models Towards Perceptual Manipulation Attacks in Mixed Reality," in Proc. USENIX Security '23 (32^{nd} USENIX Security Symposium), 2023.
[2] Esmée H. A. de Haas et al., "Deceiving Audio Design in Augmented Environments: A Systematic Review of Audio Effects in Augmented Reality," in Proc. ISMAR-Adjunct 2022 (IEEE Int'l Symp. on Mixed and Augmented Reality Adjunct), 2022.
[3] Kousei Otsuka et al., "Auditory Stimulus Attack in XR: Stimulus Characteristics and Technical Background Considerations," in Proc. MetaCom 2025 (3rd IEEE Int'l Conf. on Metaverse Computing, Networking and Applications), 2025.
[4] Meta Platforms, Inc., "Passthrough Camera API Overview," https://developers.meta.com/horizon/documentation/spatial-sdk/spatial-sdk-pca-overview/, Meta Horizon OS Developers, 2025. (Accessed: 2026-1-20).
[5] Rachel McAmis et al., "The Writing on the Wall and 3D Digital Twins: Personal Information in (not so) Private Real Estate," in Proc. USENIX Security '23 (32nd USENIX Security Symposium), 2023.

Toho University, Faculty of Science, Usable Security Laboratory https://www.klab.is.sci.toho-u.ac.jp/
JST CREST SSS Internet of Realities (IoR) Project https://www.internet-of-realities.org/
Contact: Kousei Otsuka（Toho Univ.）https://kouseiotsuka.github.io/
Mail: 6524002o@st.toho-u.ac.jp