

VRにおける視点操作攻撃：空間音響を悪用した攻撃の実証と考察

大塚 航世^{1,a)} 倉崎 翔大¹ 藤田 真由¹ 金岡 晃¹

概要：VR や AR を含む XR 技術は急速に発展しており、今後、社会全体への普及が見込まれる重要な技術である。XR はユーザに高い没入感を提供する一方で、感覚や知覚に強く作用する特性を有し、それに起因する新たなセキュリティリスクの顕在化が懸念されている。知覚に対するリスクや攻撃手法については一定の議論が進んでいるものの、その多くは視覚刺激に着目したものであり、聴覚刺激を用いたリスク評価や実証的研究は極めて限定的である。本研究では、VR 環境において空間音響刺激によってユーザの注意を特定方向に誘導し、頭部の向きを操作する視点操作攻撃の可能性に着目する。本攻撃の有効性を評価するため、ユーザ実験としてタスク遂行中の行動指標の観察と半構造化インタビューによる定性的評価を実施した。その結果、主観的な評価において視点操作効果は限定的であることが示された一方で、攻撃が成立し得る条件や環境的要因が明らかとなった。

キーワード： XR Security, Virtual Reality, 聴覚刺激攻撃, 知覚操作攻撃

Viewpoint Manipulation Attacks in Virtual Reality: Empirical Investigation and Analysis of Attacks Using Spatial Audio

KOUSEI OTSUKA^{1,a)} SHODAI KURASAKI¹ MAYU FUJITA¹ AKIRA KANAOKA¹

Abstract: XR technologies, including VR and AR, have been rapidly advancing and are expected to become a key foundational technology with widespread adoption across society in the near future. While XR provides users with a high level of immersion, it also strongly influences sensory and perceptual experiences, raising concerns about the emergence of new types of security risks. Although various risks and attack methods targeting human perception have been discussed, most prior studies have focused on visual stimuli. In contrast, risk assessment and empirical research involving auditory stimuli remain extremely limited. This study focuses on the potential for Viewpoint Manipulation Attacks in VR environments, where spatial audio is used to direct the user's attention in a specific direction and manipulate head orientation. To evaluate the effectiveness of this attack, we conducted a user study, accompanied by a qualitative assessment through semi-structured interviews. As a result, although the subjective evaluations based on the interviews indicated that the manipulation effect was limited, the analysis revealed specific conditions and environmental factors under which such attacks could potentially succeed.

Keywords: XR Security, Virtual Reality, Auditory Stimulus Attacks, Perceptual Manipulation Attacks

1. はじめに

Virtual Reality (VR) や Augmented Reality (AR) に代表される XR (Extended Reality) 技術は急速に発展し、多様

な分野での応用が広がっている。

XR がもたらす高い没入感やリアリティは、従来のコンピューティング環境には存在しない新たなセキュリティリスクを引き起こすことが近年の研究により明らかとなっている。得に、視覚や聴覚、触覚といった感覚刺激を悪用してユーザの知覚を操作する Perceptual Manipulation Attack (PMA) は、ユーザの認知や行動を意図せず誘導し、身体

¹ 東邦大学
Toho University

^{a)} 6524002o@st.toho-u.ac.jp

的危害につながる可能性が指摘されている [1-3, 9, 19]。しかし、既存の PMA 研究の多くは視覚刺激に焦点を当てており、聴覚刺激による PMA に関する議論は依然として不十分である。

本研究では、聴覚刺激を用いた攻撃モデルとして、VR 環境において空間音響を悪用しユーザの視点方向を意図的に誘導する Auditory Viewpoint Manipulation Attack (AVMA) の可能性を実証的に評価する。

既存の聴覚 PMA の実証的研究は、使用刺激やタスク設計が限定的であった [2]。そのため、本研究ではタイプの異なる複数種類の聴覚刺激を用い、現実的な操作負荷と生態学的妥当性を満たすタスク環境を設計した上で、以下の Research Questions (RQs) を設定する。

RQ1: AVMA により視点の誘導は可能か。

RQ2: AVMA を受けたユーザは、それを攻撃 (悪意) として認識するか。

RQ3: AVMA はタスクパフォーマンスに影響を与えるか。

これらの RQ に答えるため、30 名の参加者を対象にユーザ実験を実施した。実験では、VR ゲームタスク実行中の参加者に対して 5 種の AVMA を行い、半構造化インタビューから質的分析を行った。その結果、AVMA による視点誘導は認められたものの効果は限定的であり、人間の潜在的な AVMA への認知的耐性が示された。

本研究の貢献は以下の 3 点である。

- (1) VR 環境下における AVMA の成立可能性と限界の実証的評価
- (2) AVMA に対する人間の認知的耐性の存在とその要因の検討
- (3) 視点誘導を評価するための生態学的妥当性を満たす実験タスクの提案・設計

2. 関連研究

2.1 XR 特有の知覚操作 (Perceptual Manipulation) による脅威

PMA は XR 環境における新たな脅威として重大なリスクが懸念されている [2, 19]。

Casey ら [1] は、空間の微小なシフトでユーザの移動を無意識に誘導できることを実証した。Tseng ら [19] は、転倒を誘発する SteppingOn 攻撃や、顔への衝突を誘発する HittingFace 攻撃など、安全性に直結するシナリオを提案している。さらに Cheng ら [2] は、視覚や聴覚、状況認識の各チャネルに対する PMA を実証し、ユーザの判断や行動に多角的な影響を与えることを示した。

2.2 聴覚 PMA の初期研究と課題

Cheng らは通知音や着信音を用いた聴覚 PMA を実証し、MR 環境における聴覚刺激がタスクパフォーマンスに影響

を与える可能性を示した [2]。しかし、刺激は 2 種類に限定されており、採用タスクは単純で実際の XR 利用場面における複雑な行動や意思決定状況を十分に反映していない。

大塚らは聴覚刺激の特性に基づき XR 環境下の 21 種類の潜在的脅威と 5 種類の攻撃シナリオを整理し、方向性や距離感といった空間音響特性の悪用可能性を指摘した [13]。さらに、de Haas らの体系的レビューは、空間音響を含む設計が注意や行動操作に直結しうることを示している [3]。また、音は視野外からも作用し、攻撃面が拡張される可能性や、物理空間の実在音源として知覚されることで現実感を増幅し、PMA の実効性を高める可能性を指摘している。

また Rothe ら [15] は Cinematic VR において方向性をもつ音が視点移動を誘発することを示したが、用いられた刺激は文脈依存 (diegetic) 音に限定されており、文脈に依存しない刺激や複雑な行動状況下での影響は十分に評価されていない。

3. 脅威モデル

本研究で対象とする攻撃は、VR 環境 (パススルー AR を除く) においてシングルユーザゲームをプレイしているユーザに対し、空間的特性をもつ聴覚刺激を提示し、ユーザの注意および視点方向を意図的に誘導するものである。VR は HMD により実環境の視野を遮るため、物理環境由来の手掛かりや環境音の混入を最小化し、聴覚刺激のみが与える影響を単離しやすい。

3.1 攻撃前提：聴覚刺激の分類

XR 環境において利用され得る聴覚刺激は、その性質に応じていくつかの観点から分類できる。我々は、ユーザが利用している XR コンテンツの文脈依存度に着目し、聴覚刺激を「完全文脈依存」「準文脈依存」「非文脈依存」の三つに分類した (表 1)。文脈依存度に応じて、ユーザが刺激を VR 由来のものと判断するか、現実の出来事と判断するかが異なり、その結果として聴覚刺激に対する違和感や視点誘導への影響も変化すると考えられる。

その他に、音の種類による分類の観点がある。Loughrey らは International Affective Digitized Sounds (IADS) を拡張した IADS-E を構築し、聴覚刺激を 10 カテゴリに整理している [10]。

さらに、聴覚刺激の言語的意味や、それに伴う心理的反応もユーザの知覚や行動に大きく関与し得る。したがって、AVMA のような攻撃の利用を検討する際には、音の文脈的適合性、種類、意味・心理的效果といった要素を踏まえて議論する必要がある。

3.2 攻撃者モデル

本攻撃の実現手段として、「Virtual 空間への不正オブジェ

表 1: 聴覚刺激の文脈依存度分類とその説明

文脈依存度	説明
完全文脈依存	XR 空間内やコンテンツの事象と同期して発生する音
準文脈依存	XR 空間内やコンテンツに関連するが、同期していない音
非文脈依存	XR 空間内やコンテンツに全く関連しない音

クトの設置」があり、そしてそれらを可能にする「サプライチェーン攻撃」や「サイドローディングによる悪性アプリや悪性スクリプトの侵入」が考えられる。

Virtual 空間への不正オブジェクトの設置は、攻撃者は Virtual 空間に不正なオーディオオブジェクトやスクリプトを配置し、Virtual 空間上で聴覚刺激攻撃を出力する攻撃手法である。

さらに、これらの攻撃を成立させるための侵入経路の 1 つとして、サプライチェーン攻撃が考えられる。例えば、Unity Asset Store は第三者が作成したアセットを Unity の審査を経て公開できるマーケットプレイスであるが、審査を通過しても悪意あるスクリプトの混入を完全に排除できるとは限らない。その結果、開発者が悪意のあるアセットを取り込むことで、攻撃が広範に拡散する可能性が指摘されている [7]。

加えて、攻撃者が無害なアプリに偽装して不正機能を含むアプリを配布し、公式ストアの審査を回避する目的でユーザをサイドローディングへ誘導する手口も考えられる。

3.3 視点操作によるリスク

AVMA によって引き起こされるリスクは以下のものが考えられる。

- 認知負荷の増大や作業ミス: 不必要な聴覚刺激や視点変化は、タスク集中力の低下や作業ミスを誘発する可能性がある [2]。
- 重要情報の見逃し: Rothe らの研究 [15] では、視界外にある重要情報の見逃しを防ぐ方法として、誘導目的の聴覚刺激を提示しているが、逆にタスクに必要な情報が視野外に移動してしまう危険性がある。
- プライバシー侵害や機密情報漏洩: 視点誘導により頭部方向、すなわち外向き HMD カメラの向きを操作すれば、視野内情報の取得リスクが増大し得る。HMD 固有ではないが、室内画像一般から多数の個人情報が抽出可能と報告されている [12]。

4. 実験デザイン

本実験の目的は、RQ に基づいて、VR 環境下での AVMA の効果を明らかにすることである。

4.1 タスク設計

4.1.1 タスク選定理由

タスクには、テトリスを基盤としたゲームを採用し、VR

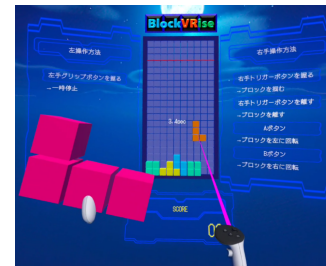


図 1: AVMA 評価のために開発した実験アプリケーション『BlockVRise』のプレイ画面

環境用に「BlockVRise (図 1)」を開発した。テトリスは広く普及しているゲームであり、VR 対応版 [4] も存在することから、参加者に受容されやすいと考えられる。

BlockVRise は両手のコントローラによる操作を特徴とし、没入感と身体性を伴った自然な行動中に AVMA を介入させることで、実際の VR 利用場面に近い状況下でユーザ反応やリスクを評価でき、生態学的妥当性が高いと考えた。

4.1.2 タスク概要

BlockVRise では、参加者はテトリスと同様に、落下してくるブロック (ミノ) を盤面に積み重ね、横一列が揃うとその列が消去されるというルールでプレイする。実験参加者は、画面上部の赤色のラインまでブロックが積み重ならないように配置することが求められる。操作は VR 環境に適応させており、図 1 に示すように、右手コントローラから発せられるピンク色のレーザーをミノに照射し、特定のボタンを押すことで「掴み」、移動や回転を行う。また、左手コントローラの位置には次に落下するミノを表示している。

本システムではタスクを一定時間継続的に実施するために、赤色のラインより上にミノが設置された場合にはグリッド上の全ブロックを消去してゲームを継続可能とした。

4.1.3 実験条件への配慮

タスクの難易度調整のため、ミノを「掴める」時間を各ミノにつき合計 5 秒間に制限した。

また、AVMA による攻撃音とは別に、タスク中は BlockVRise から聴覚刺激として BGM と効果音を出力し、自然なゲーム体験を提供した。いずれも同一のフリー音源サイトから取得し、BGM はタスク全体を通じてループ再生される。

また、BlockVRise では盤面を正面に配置し、タスク遂行で大きな頭部運動を必要としないことから、タスクに伴う自然な動作と AVMA による誘導を切り分け、視点操作効果を純粹に評価できる環境を構築した。

4.2 本実験で用いた聴覚刺激

本研究で提案する AVMA では、5 種類の聴覚刺激を用いた。以下に各刺激の概要を示す。

- **Phone Vibration Sound (PB):** 日常的に聞かれる持続的

表 2: 実験で用いる聴覚刺激の分類

具体的音	文脈依存度	音の種類	誘発され得る反応
Phone Vibration Sound	非文脈依存	日常	日常的反応 [17]
Glass Breaking Sound	非文脈依存	破壊	防御反応 [16]
Human Voice Sound	非文脈依存	人	社会的反応 [18]
Baby Crying Sound	非文脈依存	人	情動的反応 [14]
Dependent Context Sound	準文脈依存	効果音	

なスマートフォンのバイブレーション音^{*1}。

- **Glass Breaking Sound (GB):** ガラスが割れる音^{*2}。
- **Human Voice Sound (HV):** 成人女性による「お疲れ様です」という呼びかけ音。本実験では、実験実施者が男性であったため実験実施者の呼びかけと誤認することを防止する観点から、女性の声を録音して使用した。
- **Baby Crying Sound (BC):** 乳幼児の泣き声の音^{*3}。
- **Dependent Context Sound (DC):** BlockVRise においてブロックを横一列に揃えた際に出力される効果音^{*4}。文脈に依存する音を使用することで、違和感を軽減しつつ、視点誘導を実行できる可能性がある [15]。

本研究では、物理空間からの音と誤認させることを目的として、主に非文脈依存の聴覚刺激を中心に採用した。また、準文脈依存の刺激を加えることで、文脈的な関連性が視点誘導に与える影響についても検証可能とした。各聴覚刺激の文脈依存度、音の種類、および想定される反応を表 2 に示す。

本研究では音像位置の知覚が比較的容易な左右方向 [13] を選択し、ユーザの真左または真右、Virtual 空間上で 5m 離れた位置からこれらの聴覚刺激を出力した。

4.3 実験手順

本実験では欺瞞的アプローチを採用した。参加者には、実験目的や取得情報、謝礼、中断に関する説明を行い、同意を得た上で録画・録音を開始した。

次にタスクのルールを説明し、ゲーム内の全効果音の発生条件を明示した。プレイ動画とともに効果音を提示し、タスク中に AVMA とゲーム効果音が混同されないよう配慮した。

その後、参加者は HMD を装着し、4 分間のチュートリアルを経て本番タスクを開始した。本番では開始 2 分後から AVMA を実施し、以降 100 秒間隔で音刺激を提示した。実施者はプレイ画面をリアルタイムで確認しながら、刺激提示時の様子を観察した。

終了後にデブリーフィングを行い、本来の目的を説明して同意を再確認した。その後、アンケートとインタビューを実施し、謝礼を渡して実験を終了した。

^{*1} https://otologic.jp/free/se/cell_phone02.html

^{*2} <https://taira-komori.jp/daily01.html>

^{*3} https://vsq.co.jp/plus/sound/category_sub/baby/

^{*4} <https://musmus.main.jp/se.html>

4.4 実験参加者

実験は 2025 年 5 月 1 日から 2025 年 7 月 15 日に実施し、31 名（男性 18 名、女性 13 名、非回答 0 名）が実験に協力した。このうち女性 1 名は音量設定を誤り音量が低い状態で実験を行ったため、分析対象から除外した。以下の参加者属性は、分析対象の 30 名に基づくものである。

実験参加者の年齢は 18 歳から 26 歳で、平均年齢は 21 歳（SD = 2.017）であった。全員が東邦大学理学部または理学研究科に所属している。また、テトリスゲーム経験についてのアンケートでは、参加者全員がテトリスのルールを事前に理解しており、19 名の参加者はテトリスゲームのプレイ経験者であった。

参加条件として、視力（矯正時 0.7 未満）や聴覚に問題のある者は参加を控えるよう呼びかけた。これは、VR タスクの遂行において十分な視覚・聴覚機能が確保されていない場合、タスクパフォーマンスや AVMA の効果に影響を及ぼし、実験結果に偏りが生じる可能性があるためである。

本実験の参加者には謝礼として 2000 円分の Amazon ギフトカードを提供した。謝礼の内容は、大学所在地の最低賃金を基準とし、実験の所要時間（約 70 分）および来訪に伴う移動時間等を考慮して算出した。

4.5 実験環境・システム構成

本実験では Meta Quest 3 を使用した。アプリケーションの開発には Unity 2021.3.25f1 を使い、音響表現には、DearVR Unity（バージョン 1.6.0）^{*5}を使用した。

参加者は椅子に座った状態でタスクを遂行した。実験中は室外に「実験中」の掲示を行い、周囲に静粛を促すことで外部音が実験に影響しないよう配慮した。

4.6 実験の統制条件

すべての AVMA は Unity の Audio Mixer を使い、レベルメーターを目視で確認し同程度の音量に調整した。

また、ミノの出現順序による結果の偏りを防ぐため、全参加者に対して同一の出現パターンを適用した。

加えて、順序効果や一次キャリーオーバー効果を排除するため、Williams Design を採用した。最小 10 通りの組み合わせを作成し、30 名の参加者に 3 名ずつ割り当てて効果を均等化した。

さらに、作成した 10 通りの出力順序を、順序効果が均等化される前半 5 通りと後半 5 通りに分割した。前半 5 通りでは、PB・HV・BC の 3 種類を左方向、GB・DC の 2 種類を右方向から出力し、後半 5 通りでは各出力方向を反転した。

最終的な出力順序と出力方向を表 3 に示す。

^{*5} 当該製品は 2024 年 9 月に販売を終了している。

表 3: AVMA の出力順序および左右出力方向の割り当て。
表中の括弧内は出力方向 (L: 左方向, R: 右方向) を表す。

順目	1st AVMA	2nd AVMA	3rd AVMA	4th AVMA	5th AVMA
1	PB (L)	GB (R)	BC (L)	HV (L)	DC (R)
2	GB (R)	HV (L)	PB (L)	DC (R)	BC (L)
3	HV (L)	DC (R)	GB (R)	BC (R)	PB (L)
4	DC (R)	BC (L)	HV (L)	PB (L)	GB (R)
5	BC (L)	PB (L)	DC (R)	GB (R)	HV (L)
6	DC (L)	HV (R)	BC (R)	GB (L)	PB (R)
7	BC (R)	DC (L)	PB (R)	HV (R)	GB (L)
8	PB (R)	BC (R)	GB (L)	DC (L)	HV (R)
9	GB (L)	PB (R)	HV (R)	BC (R)	DC (L)
10	HV (R)	GB (L)	DC (L)	PB (R)	BC (R)

4.6.1 Inter-Stimulus Interval (ISI) 設定

本実験では、各参加者に対して 5 回の AVMA を実施した。このとき、刺激提示間隔 (Inter-Stimulus Interval: ISI) は AVMA の効果評価において重要な要素であるため、本研究ではその設定を特に考慮した。

既存研究では、20 秒の ISI と比較して 100 秒の ISI を設定した場合、長期的な馴化 (慣れ) の進行が抑制され、刺激の新規性が維持されることが報告されていることから、本実験では ISI を 100 秒に設定した [6]。

4.7 倫理的配慮

本実験は、東邦大学理学部生命倫理審査委員会の承認を得て実施した。

本実験は欺瞞実験として設計され、参加者には「VR 体験が集中力や疲労感に与える影響を評価する」と説明した。また、VR HMD 利用により不快感が生じる可能性を伝え、実験中は理由を問わず中断・辞退でき、いずれの場合も謝礼を満額支給することを事前に明示した。

メインセッション終了後にデブリーフィングを行い、本来の目的と倫理的配慮を説明した。すべての参加者が最終的に継続に同意し、辞退者はいなかった。

音刺激は NIOSH の基準 [5] に基づき 85dB 以下に制限した。この水準は最大 8 時間連続使用しても聴覚に悪影響を及ぼさず、本実験はこれより大幅に短時間で実施していることから、聴覚へのリスクは極めて低いと判断した。

取得したデータは匿名化し、研究室内の鍵付きキャビネットに厳重に管理した。

4.8 分析手法

本研究では、AVMA の評価を目的として半構造化インタビューを実施した。詳細な質問項目は GitHub 上^{*6}で公開している。インタビューでは、まず実験中に不自然な聴覚刺激を知覚したかを確認し、その後、実験実施者側から使用した刺激を提示して AVMA の認知の有無を確認した。

^{*6} <https://github.com/kanaoka-laboratory/AVMA-XR-CSS2025>

その後、実験参加者が認知した AVMA に対して以下の観点から質問を行った。

- **Virtual と物理空間の混同:** 聴覚刺激を HMD スピーカー出力か実験室音かを判断したか。
- **攻撃認識:** AVMA のより攻撃や悪意を感じたか。
- **視点操作:** AVMA により視点が操作されたと感じたか。
- **タスクへの影響:** AVMA がタスク遂行に影響したか。

インタビューは録音音声を文字起こしし、Nvivo を用いて著者 2 名が帰納的コーディングを行った。両者はまずそれぞれ独立にコードブックを作成し、議論を通じて統一コードブックを策定した。その後、再度コーディングを行い、全参加者についてコードの一致率を評価した結果、Cohen の Kappa 係数は 0.745 となり、信頼性が確保された [11]。

4.9 制限事項

本実験には以下の制限事項が存在する。

1 つ目に、対象は著者らの大学に所属する若年層学生に限られ、全年齢層や一般社会人への一般化には限界がある。加齢による感覚や認知機能の変化を踏まえると、より多様な集団での検証が望まれる。

2 つ目に、視覚/聴覚の異常有無は募集案内で確認するとどまり厳密なスクリーニングは行っていない。ただし実験中に異常の訴えはなく、AVMA 認知率の結果からも刺激は十分知覚されていたと考えられる。

3 つ目に、BlockVRise は右手操作を前提としており、利き手によって操作性に差が生じた可能性がある。

4 つ目に、AVMA で用いた音源はフリー素材 4 種類と独自録音 1 種類を併用しており、録音環境の違いが影響した可能性がある。

5. 結果

5.1 AVMA について

5.1.1 AVMA の認知

実験の結果、すべての参加者が少なくとも 1 つ以上の AVMA を認知していた。インタビュー結果から、刺激提示前に自発的に言及されたものを「記憶に保持された AVMA」、刺激提示後に想起されたものを「刺激提示後に想起された AVMA」、実験中に全く認知されなかったものを「認知されなかった AVMA」として分類し、各 AVMA の認知率を表 4 に示す。

PB、HV、BC は総じて 90%を超える認知率を示した。一方で、PB については、認知した参加者 28 人のうち 19 名 (67.8%) がインタビュー時には思い出せなかったが、研究者から刺激を提示された際に思い出した。

一方、DC は 13 名 (41.4%) の参加者が認知せず、タスク終了後も記憶に保持されていたのは 4 名 (13.8%) にとどまった。

表 4: AVMA の認知率 (%)

AVMA 種類	PB	GB	HV	BC	DC
実験中に認知された AVMA	93.3	76.7	100.0	96.7	56.7
記憶に保持された AVMA	30.0	50.0	66.7	86.7	13.3
刺激提示後に想起された AVMA	63.3	26.7	33.3	10.0	43.3
認知されなかった AVMA	6.7	23.3	0.0	3.3	43.3

表 5: AVMA に対する攻撃・悪意認識率 (%)

AVMA 種類	全て	PB	GB	HV	BC	DC
攻撃悪意認識率	26.7	3.7	17.4	6.7	13.8	5.9

表 6: 音の現実性評価 (高/低) と音源判 (実験室/VR) のクロス集計表

現実性	実験室 (人)	VR (人)
高	22	36
低	0	36

なお「聞こえた気がするが覚えていない」など回答が曖昧な場合は「実験中に認知されなかった AVMA」として扱った。また、BC を認知していた 29 名中 12 名は猫や動物の鳴き声と誤認していたが、刺激提示後に訂正した場合は「実験中に認知された AVMA」として分類した。

5.1.2 AVMA の攻撃/悪意認識

インタビューにおいて、認知した音を聞いた際にサイバー攻撃を受けたと感じたか、あるいは悪意を感じたかを確認した結果、実験を通じて何らかの攻撃性または悪意を感じ取った参加者は 8 名であった。AVMA 毎の攻撃/悪意の認識率を表 5 に示す。

5.1.3 物理現実との混同

インタビューでは、物理現実との混同を狙った AVMA (PB、GB、HV、BC) を聞いた際に、それが実験室から発生したものか、あるいは VR ゴーグルのスピーカーからの出力であったかを確認し、混同の有無を評価した。さらに、各音刺激について再現度や音の現実性が高いか低いかを尋ねた。

音の現実性 (高/低) と音源判断 (実験室/VR) の関連性を検証として、クロス集計表 (表 6) を作成し、 χ^2 検定を実施した。その結果、 $\chi^2(1, N = 94) = 15.77, p < .001$ となり、両者は独立ではなく、有意な関連があることが認められた。

観測度数 (表 6) をみると、現実性が低いと判断された音について実験室からの発生と回答した参加者は存在しなかった。一方、現実性が高いと判断された音については、37.9% (22/58) が実験室からの音と回答したのに対し、62.1% (36/58) は VR からの音と回答していた。

5.1.4 視点誘導の有無

実験の結果、いずれかの AVMA によって視点が誘導された参加者は 15 名 (50.0%) であった。誘導が認められた参加者における各 AVMA の誘導割合を表 7 に示す。誘導

表 7: 視点誘導が認められた参加者における AVMA 別誘導割合 (%)

AVMA 種類	PB	GB	HV	BC	DC
誘導割合	26.7	40.0	53.3	26.7	33.3

が確認された 15 名のうち半数を超える 9 名が HV によって視点を誘導されていた。

5.2 タスクへの影響

インタビューにおいて、不正な聴覚刺激によってタスクへの影響を調査した結果、17 名 (56.7%) の参加者は、AVMA によって意識が音やタスク以外の事象にそらされる、あるいは集中力が低下したと回答した。また、5 名の参加者は AVMA により動作が一時的に停止したり、思考が途切れたと述べた。さらに、7 名の参加者は、AVMA によってタスクにミノの設置ミスといった操作ミスが発生したと回答した。

5.3 タスクに対する意識

タスクへの集中は、AVMA の誘導効果に影響する可能性がある。インタビューでは、22 名 (73.3%) の参加者が実験およびゲームに集中していたと回答した。また、視点誘導が生じなかった AVMA について理由を尋ねたところ、15 名が「実験やゲームに集中していた」「ゲームが楽しかった」と述べた。

6. 考察

本章では、前章で得られた結果とコーディング結果を用いてさらに深く考察し、RQ への回答とともにインタビューで得られたその他の内容について議論を行う。

6.1 視点誘導の可能性 (RQ1)

本研究では、5 種類の AVMA のいずれかによって視点が誘導されたと答えた参加者は 16 名であった。さらに、物理空間からの実在音と誤認させることを狙った PB、GB、HV、BC による誘導は 14 名が確認された。

一方、インタビュー調査では、同じ音が日常生活環境で発生した場合には 28 名 (93%以上) の参加者が振り向くと回答しており、音そのものには潜在的に強い誘導力があることが考えられる。この結果との乖離を説明する要因として、以下の要素が考えられる。

6.1.1 物理環境文脈が視点誘導に与える影響と認知的耐性

インタビューでは、音源判断において VR と回答した参加者や、誘導が生じなかった理由として、23 名の参加者が「実験室に赤ちゃんはいない」「割れる物がない」といった物理環境に基づく発言が確認された。これらは、ユーザが音源を判断する際に音の質のみならず、周囲の物理環境という文脈の手がかりを参照していることを示している。

本実験では参加者が実験室に入室してから VR デバイスを装着するまでに約 15 分の時間があり、この間に得られた実験室環境の視覚的/聴覚的特徴が短期記憶として保持され、物理環境文脈を形成していたと考えられる。聴覚は視覚よりも場面文脈や短期記憶に依存した期待に強く拘束されることが報告されており [8]、このことは参加者が物理環境の文脈的不一致を敏感に検出した理由を説明できる。

たとえ自然な音が提示されたとしても、人間はわずかな文脈の齟齬を手がかりに違和感を察知し、結果として AVMA に対して高い耐性（認知的耐性）を示す可能性がある。すなわち、人間は AVMA に気づきやすい側面を有している一方で、音が環境文脈と強く整合する場合には、それを物理現実由来のものと混同し視点誘導が成立する可能性が高まると考えられる。

6.1.2 タスク集中度と攻撃成立の困難性

インタビューではタスク中や誘導不成立の理由としてタスクへの集中や没頭に関する言及があり、それが誘導効果を抑制した可能性が考えられる

本実験タスクである BlockVRise は、その性質上注意が分散しにくく、視覚的焦点が正面の盤面に集約される。

本実験タスクである BlockVRise は、その性質上注意が分散しにくく、視覚的焦点が正面の盤面に集約されるため、視覚的焦点が一箇所に強く集約されるタスク環境では、AVMA の視点誘導効果は限定的になると考えられる。一方で、リラックスした VR 体験や探索型タスクのように、視覚的焦点が分散しやすい環境下では、聴覚刺激による視点誘導が成立する可能性が高まると考えられる。

6.1.3 聴覚刺激の意味的性質と誘導効果

聴覚刺激そのものが持つ誘導能力の差も、視点誘導の成否に影響したと考えられる。たとえば、PB については、4 名が「今すぐ確認する必要はない」と述べており、行動化の優先度が低い刺激であると認識されていた。このような認識は、PB における刺激提示までの忘却率が比較的高かった要因の一つであると考えられる。

一方で、GB や HV は、他の刺激と比較してわずかではあるが相対的に高い誘導効果が確認された（表 7）。これは、GB が示す「危険や破損の兆候」としての意味的緊急性や、HV が含む「呼びかけへの応答義務」といった社会的関与度が高いため、誘導効果を強めたものと考えられる。ただし、本実験の物理環境文脈と整合しない場合には、参加者が周辺情報を即座に参照し、誘導が抑制される傾向がみられた。

DC は、コンテンツ側の文脈に整合するため違和感は小さい一方、その意味は「消去の通知」であるため緊急性や行動化可能性に乏しく、視点シフトには結びつきにくいことが考えられる。

すなわち、視点誘導の成立には音の「文脈的適合性」に

加えて、その音を持つ「意味的強度」も重要である。

6.1.4 AVMA 成立の条件と制約

以上の考察をまとめると、本研究で AVMA の誘導効果が限定的であった理由として、(1) 参加者が事前に得た物理環境の文脈記憶により現実混同が抑制されたこと、(2) タスクへの集中度が高く視覚的焦点が強く集約していたこと、(3) 刺激音の意味的強度の違いの三点が挙げられる。すなわち、単に物理環境の音を模倣するだけでは視点誘導は成立しにくく、物理環境に整合する文脈的適合性を備え、注意が分散しやすいタスク環境において、さらに緊急性や社会的関与度の高い意味をもつ音が提示される場合に、AVMA はより強く成立する可能性がある。

しかし、そのような成立条件を満たすためには、ユーザの視覚的焦点や注意状態のリアルタイム推定、タスク文脈に応じた最適な提示タイミングの決定、高精度な空間音響再現、さらに物理環境との文脈整合性の同時実現といった高度な技術が必要となる。したがって、物理現実との混同を狙う AVMA を安定的に成立させることは困難であるといえる。

6.2 AVMA の認識 (RQ2)

本実験では、すべての参加者が少なくとも 1 つ以上の AVMA を認知しており、特に、非文脈依存の PB、HV、BC の認知率は 90% 超の認知率を示した（表 4）。一方、DC の認知率は、56.7% にとどまり、さらにタスク終了後も記憶に保持していた参加者は 4 名のみであった。すなわち、文脈依存度により「気づかれやすさ」が異なることが示唆される。

また、認知された音全体に対して、攻撃性/悪意を感じた参加者は 26.7% であった。インタビューでも、24 名が認知したいずれかの AVMA を「ゲームの不具合」「何らかのシステム/効果音」と解釈したと述べており、AVMA の認識が攻撃などの外的要因ではなく、利用コンテンツやシステム内部の要因へと帰属されやすい傾向が示された。これらの結果は、聴覚刺激による「攻撃」がユーザに気づかれにくいという点で、高い秘匿性を有し得ることを示唆している。

6.3 タスクへの影響 (RQ3)

インタビューでは AVMA によって集中力の低下や、一時停止、ミスといった影響が確認された。また、視点誘導が直接の原因でミスが生じたと述べた参加者はいなかったものの、視点がタスクと無関係な方向へ移動することにより誤操作を誘発し得る可能性は十分に考えられる。

これらの所見は、聴覚刺激による攻撃が視点誘導を目的とせずとも、タスク集中を阻害し、中断や誤操作を引き起こすリスクを内包することを示す。とりわけ、XR 技術が適用される医療、航空、軍事等の高信頼領域では、短時間

の中断や小規模なミスが重大な結果に直結し得る。

6.4 VR と AR による影響の差異

インタビューでは、2名の参加者が「振り向いても確認できない」と回答し、誘導が生じなかった理由となった。本実験では VR HMD の装着により物理空間の視認ができず、行動が抑制された可能性がある。一方で、透過型 AR グラスやパススルー表示のように物理環境を視認できる状況では、聴覚刺激に対する確認行動は抑制されにくいと考えられる。

6.5 実験アプリケーションや実験が与えた影響

アンケートによる BlockVRise の SUS スコアは 74.42 であり、インタビューでも BlockVRise 自体に対する違和感やストレスの指摘は確認されなかった。以上より、本実験における AVMA の影響評価に対し、実験アプリケーションの完成度が作用した可能性は低いと考えられる。

また、実験の欺瞞手続により、不正な聴覚刺激が視点誘導を目的としている点に気づいた参加者はいなかった。しかし、実験という状況や、連続した不自然な聴覚刺激について言及した参加者がいたことから、実験環境特有のバイアスが AVMA の誘導効果や刺激認知に影響を与えた可能性がある。

7. おわりに

本研究は、VR 環境における聴覚刺激による視点誘導攻撃 (AVMA) の影響を評価した。結果として、ユーザには一定の耐性が見られる一方で、攻撃を自覚しにくい高い秘匿性が確認された。これは、聴覚刺激を悪用した攻撃がユーザに気付かれず影響を及ぼす可能性を示し、VR 普及に伴い重要なセキュリティ課題となり得る。

謝辞 本研究は、JST CREST (JPMJCR22M4、AIP チャレンジプログラム 2024 を含む) および国立研究開発法人情報通信研究機構 (NICT) のリサーチアシスタント制度の支援を受けて実施された。

参考文献

[1] Casey, P. et al.: Immersive Virtual Reality Attacks and the Human Joystick, *IEEE Transactions on Dependable and Secure Computing*, Vol. 18, No. 2, pp. 550–562 (online), DOI: 10.1109/TDSC.2019.2907942 (2019).

[2] Cheng, K. et al.: Exploring User Reactions and Mental Models Towards Perceptual Manipulation Attacks in Mixed Reality, *32nd USENIX Security Symposium (USENIX Security 23)* (2023).

[3] de Haas, E. H. A. et al.: Deceiving Audio Design in Augmented Environments: A Systematic Review of Audio Effects in Augmented Reality, *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct 2022)* (2022).

[4] Enhance Experience Inc. - TETRIS® EFFECT: CON-

NECTED — Official Website (English). <https://www.tetriseffect.game/> (Accessed: 2025-08-22).

[5] National Institute for Occupational Safety and Health (NIOSH) - Noise-Induced Hearing Loss. <https://www.cdc.gov/niosh/noise/about/noise.html> (Accessed: 2025-08-22).

[6] Gatchel, R. J. et al.: Effects of interstimulus interval length on short- and long-term habituation of autonomic components of the orienting response, *Physiological Psychology*, Vol. 3, No. 2, pp. 133–136 (1975).

[7] Include Security Research Blog - Hacking Unity Games with Malicious GameObjects. <https://blog.includesecurity.com/2021/06/hacking-unity-games-malicious-unity-game-objects/> (Accessed: 2025-08-22).

[8] Kaganovich, N. et al.: Event-related potentials evidence for long-term audiovisual representations of phonemes in adults, *Eur J Neurosci*, Vol. 54, No. 11, pp. 7860–7875 (2021).

[9] Kurasaki, S. et al.: Image Movement Attacks on Optical See-Through HMDs: Covert Gaze Manipulation and Privacy Risks in AR/MR Systems, *3rd IEEE International Conference on Metaverse Computing, Networking and Applications (MetaCom 2025)* (2025).

[10] Loughrey, D. G. et al.: Association of Age-Related Hearing Loss With Cognitive Function, Cognitive Impairment, and Dementia: A Systematic Review and Meta-analysis, *Neuropsychology Review*, Vol. 28, No. 2, pp. 181–199 (online), DOI: 10.1007/s11065-018-9381-6 (2018).

[11] MedCalc Software Ltd. - Inter-rater agreement (kappa), MedCalc Manual (Online documentation) (2025). Accessed: 2025-08-21; describes computation of Cohen's κ and weighted κ with linear and quadratic weights.

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

[13] Otsuka, K. et al.: Auditory Stimulus Attack in XR: Stimulus Characteristics and Technical Background Considerations, *3rd IEEE International Conference on Metaverse Computing, Networking and Applications (MetaCom 2025)* (2025).

[14] Rigo, P. et al.: Brain processes in women and men in response to emotive sounds, *Social Neuroscience*, Vol. 12, No. 2, pp. 150–162 (online), DOI: 10.1080/17470919.2016.1150341 (2017).

[15] Rothe, S. et al.: Diegetic Cues for Guiding the Viewer in Cinematic Virtual Reality, *19th ACM Symposium on Virtual Reality Software and Technology (VRST '17)* (2017).

[16] Soh, C. et al.: Unexpected Sounds Nonselectively Inhibit Active Visual Stimulus Representations, *Cerebral Cortex*, Vol. 31, No. 3, pp. 1632–1646 (online), DOI: 10.1093/cercor/bhaa315 (2021).

[17] Stohart, C. et al.: The Attentional Cost of Receiving a Cell Phone Notification, *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 41, No. 4, p. translation into pages if available (online), DOI: 10.1037/xhp0000100 (2015).

[18] Tillmann, J. et al.: The Effect of Visual Perceptual Load on Auditory Awareness of Social vs. Non-social Stimuli in Individuals with Autism, *Journal of Autism and Developmental Disorders*, Vol. 51, No. 4, pp. 1028–1038 (online), DOI: 10.1007/s10803-020-04587-0 (2021).

[19] Tseng, W.-J. et al.: The Dark Side of Perceptual Manipulations in Virtual Reality, *2022 ACM CHI Conference on Human Factors in Computing Systems (CHI '22)* (2022).