



OPEN ACCESS

EDITED BY

Justine Saint-Aubert,
Inria Rennes - Bretagne Atlantique Research
Centre, France

REVIEWED BY

Timothy John Pattiasina,
Institut Informatika Indonesia, Indonesia
Yoshiko Arima,
Kyoto University of Advanced Science (KUAS),
Japan

*CORRESPONDENCE

Yang Guang,
✉ yangguang@vogue.is.uec.ac.jp

RECEIVED 28 June 2025

REVISED 02 October 2025

ACCEPTED 21 October 2025

PUBLISHED 13 November 2025

CITATION

Guang Y, Sakurai S, Matsumura K and Okafuji Y
(2025) The impact of audience avatar
movements on presence in virtual live events.
Front. Virtual Real. 6:1655545.
doi: 10.3389/frvir.2025.1655545

COPYRIGHT

© 2025 Guang, Sakurai, Matsumura and Okafuji.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

The impact of audience avatar movements on presence in virtual live events

Yang Guang^{1*}, Sho Sakurai¹, Kohei Matsumura ² and
Yuki Okafuji²

¹Hirota Laboratory, University of Electro-Communications, Chofu, Tokyo, Japan, ²Playful Laboratory, Ritsumeikan University, Osaka, Japan

Introduction: Virtual live (VL) events allow audiences to participate as avatars in immersive environments, but uniform and repetitive audience avatar motions can limit emotional engagement and presence. This pilot study investigates whether varying proportions of user-controlled versus dummy (scripted) avatars affect perceived presence and synchrony.

Methods: Seven participants experienced a VR concert under three conditions (0%, 50%, 100% user-controlled avatars) using the Oculus Quest 2. Presence was measured using the Igroup Presence Questionnaire (IPQ), and avatar-movement synchrony was evaluated using autocorrelation and cross-correlation analyses. A Synchronization Index (SI) was defined as cross-correlation between each participant's motion and the group-averaged motion.

Results: Presence scores increased with a higher proportion of user-controlled avatars. SI also tended to increase in the 100% condition. Although correlations between SI and presence were moderate ($r = .42$) and did not reach statistical significance due to the pilot sample size, both subjective and objective trends supported the hypothesized direction.

Discussion: User-controlled audience avatars enhanced presence compared to scripted avatars, suggesting movement diversity contributes to immersive VL experiences. This study demonstrates a feasible evaluation pipeline for avatar synchrony and highlights design implications for VR live events. Future work will expand sample size, incorporate additional behavior modalities, and refine synchrony metrics.

KEYWORDS

virtual reality, virtual live events, avatars, user interaction, synchronization in VR, immersive presence, emotional engagement, cross-correlation

1 Introduction

In virtual reality (VR) research, immersion is commonly defined following Slater and Wilbur's framework as the extent to which a system delivers an enveloping and coherent sensory environment that supports user involvement (Slater and Wilbur, 1997). Presence refers to the subjective sensation of "being there" in the virtual environment, as conceptualized by Lombard and Ditton and measured via instruments such as the Presence Questionnaire by Witmer and Singer (Lombard and Ditton, 1997; Witmer and Singer, 1998). In this study, we adopt these definitions to ensure consistency: immersion in the sense of environmental enveloping (Slater and Wilbur, 1997) and presence as the subjective experience measured by validated questionnaires (Witmer and Singer, 1998).

Virtual Live events (VL events) refer to music concerts, theater performances, virtual tours, and similar experiences that take place entirely within virtual environments rather than physical venues. Unlike traditional live streaming, VL events typically leverage VR avatars to enable deeper immersion and novel forms of user interaction unique to VR platforms. Key characteristics include participants' interactive engagement via avatars, a heightened sense of self-projection in the virtual space, and the potential for synchronous, co-located experiences in distributed settings. During the COVID-19 pandemic, VL events emerged as an important alternative for entertainment under social distancing constraints; surveys and industry reports indicate a surge in participation and development of VR-based performances (Swarbrick et al., 2021; XR Association, 2023). Consequently, VL performances have drawn research interest as platforms for immersive user experiences.

Building on this context, our work examines one of the critical challenges reported by users—namely, the perceived lack of presence in audience roles—and investigates how avatar movement diversity may mitigate this issue.

While capable of rendering live stages and supporting avatar-based participation, existing VL platforms such as VRChat (2022), Cluster (2023), and Project Sekai (SEGA Co., Ltd, 2024) have been reported to exhibit limitations in user-perceived presence in the audience role. Industry surveys and user studies suggest that audience avatars often lack sufficient expressiveness, and interactions are constrained by uniform movement patterns that fail to convey individual emotions or personality cues (Waltemate et al., 2018).

One promising direction is to adjust audience interaction in virtual live events to more closely approximate real-world social dynamics, thereby improving subjective presence (Gonzalez-Franco and Lanier, 2017; Sebanz et al., 2006). Although full-body motion capture systems (e.g., Xsens or Perception Neuron) can offer richer individual movements and expressive potential (VRChat, 2023), such approaches entail high costs and technical barriers and do not guarantee enhanced social interaction among audience avatars by themselves. Research on group synchrony indicates that merely enabling realistic motion is insufficient. Rather, context-sensitive and diversity-oriented synchronization mechanisms are required to foster emotional engagement (Tarr et al., 2012; Hatfield et al., 1993). Therefore, improving presence requires the exploration of cost-effective technical solutions in tandem with carefully designed interaction conditions. In this exploratory pilot study, we examine how varying the diversity of audience avatar movements and synchronization patterns affects the subjective presence in VL event settings.

Despite advances in immersive VR performances, systematic methods to quantify avatar synchronization effects remain scarce. To address this gap, we propose a mixed-methods framework combining validated presence scales with a synchronization index, and conduct a pilot study to test how user-driven versus controlled (human-driven) avatars affect presence. Based on these findings, we outline initial design implications for large-scale VR audiences.

In this study, we expand the definition of presence to encompass not only spatial immersion but also emotional engagement elicited through avatar interactions. Building on established instruments such as the Slater-Usoh-Steed questionnaire and the Igroup

Presence Questionnaire (Usoh et al., 2000; Schubert et al., 2001a), we incorporate additional exploratory measures related to avatar movement diversity and synchronization patterns. Details of these measures, including the combination of subjective presence scores with motion-capture data to assess avatar-driven interactions, are described later in the paper.

In this study, we assess presence in VL events through subjective immersion and emotional engagement measures collected via validated questionnaires and complemented by exploratory motion-capture metrics. Here, presence is conceptualized primarily as the audience's emotional involvement and sense of co-presence facilitated by avatar interactions, with a focus on movement diversity and synchronization quality rather than mere interaction frequency. Building on established methodologies, our questionnaire-based approach employs adapted items from the Presence Questionnaire (Witmer and Singer, 1998) alongside additional questions tailored to capture perceptions of avatar movement realism and social connection. Details of the questionnaire design and analysis procedures are described later in the paper.

The objective of this exploratory pilot study is to examine how varying the diversity of audience avatar movements, from uniform scripted patterns to richer, user-like variations, affects subjective presence and emotional engagement in VL event settings. By addressing the challenge of limited audience interaction in current VL platforms, we investigate movement diversity as a factor that influences presence. Although conducted with a small sample and focused interaction modality (e.g., synchronized light-stick gestures), this study aims to generate preliminary insights and propose initial design guidelines for VR event experiences that support enhanced presence. These findings will inform future large-scale studies and platform design strategies.

Building on foundational frameworks of social synchrony, notably the joint action theory (Sebanz et al., 2006) and emotional contagion models (Hatfield et al., 1993), we posit that coordinated and diverse avatar movements can amplify co-presence beyond mere visual immersion. Recent advances in entertainment computing further demonstrate that interactive media platforms generate powerful illusions of social unity (Gonzalez-Franco and Lanier, 2017). In this study, we (1) introduce a robust evaluation framework combining psychometric presence scales with a novel motion-capture-based Synchronization Index, (2) test how varying proportions of synchronized versus dummy (scripted) avatars modulate emotional engagement and presence, and (3) extract design principles for scalable, cost-effective audience interaction in VL events. Together, these contributions both extend VR presence theory and provide practical guidelines for next-generation digital entertainment systems.

2 Related work: research on face-to-face and avatar movement coordination

Research on remote group synchrony indicates that simple gestures, such as virtual “high-fives” or collective waves, can foster a shared sense of unity among distributed viewers and boost immersive engagement (PutturVenkatraj et al., 2024;

Sebanz et al., 2006). In large physical audiences, emergent synchronization of handclapping has been linked to collective emotional alignment (Néda et al., 2000), and synchronized arousal between performers and spectators has been demonstrated in ritual contexts (Konvalinka et al., 2011). In VR sports viewing, studies show that these synchronized non-verbal cues convey emotional states even in the absence of speech, thereby enhancing subjective presence (Kimmel et al., 2024; Hatfield et al., 1993). Nevertheless, the precise pathways through which real-time emotional expression via avatar movement heightens presence remain underexplored.

In parallel, seminal work by (Dahl and Friberg, 2007) established that movement parameters, particularly velocity and amplitude, critically shape emotional transmission in physical performances, however their findings have to be fully translated to VR settings yet. Subsequent VR research has demonstrated that avatar motions aligned with musical or performative cues can function as proxies for emotional expression and elevate presence levels (Shlizerman et al., 2018), but systematic investigation of these effects under varying synchronization conditions is lacking.

Extending beyond lab-scale experiments, large-scale virtual concerts such as Ariana Grande's Fortnite performance have exemplified how massive, real-time avatar synchronization can produce powerful collective experiences, suggesting design principles for next-generation VL platforms (Hatmaker, 2021). Meanwhile, (Rogers et al., 2022) compared real-time full-body avatar interactions with face-to-face communication and found comparable levels of user comfort and engagement, yet their study did not directly measure presence or unity in live-event contexts.

Entertainment computing research further underscores that interactive VR environments can generate compelling illusions of presence (Gonzalez-Franco and Lanier, 2017), and that user satisfaction in simulated live performances correlates with avatar responsiveness and environmental interactivity (Liaw et al., 2020). Together, these bodies of work point to a gap in understanding how real-time diversification and synchronization of audience avatar movements jointly contribute to emotional expression, social unity, and the subjective sense of "being there." Cross-correlation approaches to quantify movement coordination have been established in prior work (Cornejo et al., 2018; 2023). Our study extends this line by applying avatar-audio coupling in VL audience settings and by integrating beat-locked motion features into a Synchronization Index (SI). Recent advances in AI-driven avatar animation further illustrate this potential: (Ullal et al., 2021; Ullal et al., 2022) proposed multi-objective optimization frameworks for expressive gestures in AR/MR; Juravsky et al., 2024 introduced SuperPADL, enabling scalable language-directed physics-based motion control; and Bourgault et al. (2025) presented Narrative Motion Blocks, combining natural language with direct manipulation for animation creation. Our study therefore investigates how varying degrees of movement diversity among audience avatars modulate presence and unity in VL events, addressing this critical research gap.

Based on previous studies, the real-time reflection of human movements in audience avatars during VL events may create experiences similar to those of actual live events. This study aims to explore how diversifying the movement patterns of audience

avatars enriches emotional expression and influences presence in VL events.

3 Experimental verification of avatar proportion and presence in digital entertainment

To investigate how varying the proportion of avatar types influences the presence of audience avatars, this section outlines the experimental design and methods employed in this study.

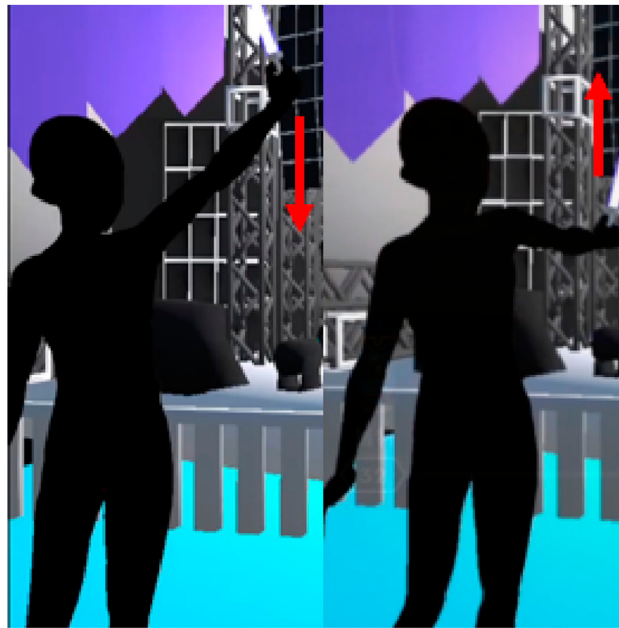
Terminology: In this paper, we use consistent terms for audience avatars. Program-controlled avatars are referred to as "controlled (human-driven) avatars" (replacing earlier labels such as "pre-programmed" or simply "dummy"). Avatars driven by live participants are referred to as "user-controlled (human-driven) avatars" (replacing earlier terms such as "user avatars" or "human avatars"). These unified terms are applied consistently throughout the manuscript, including figure captions.

3.1 Evaluation framework

In order to rigorously assess avatar synchronization effects, we developed a dual-track evaluation framework. First, subjective presence was measured using the Igroup Presence Questionnaire (IPQ) (Schubert et al., 2001b), comprising Spatial Presence, Involvement, and Experienced Realism subscales scored on a seven-point (0–6) scale. Internal consistency in our pilot yielded Cronbach's $\alpha = 0.84$. In addition, we included exploratory "unity" and "empathy" items, which were not part of the validated IPQ but were designed to probe potential interpersonal aspects of VL presence. Second, we define a Synchronization Index (SI) as the maximum normalized cross-correlation coefficient between each participant's glow-stick motion and the *group-averaged audience motion* time series, computed within a fixed lag window and normalized to [0,1]. In parallel, we report beat-alignment CCFs between individual motion and the 140 BPM reference sequence as a separate, exploratory objective metric (not part of SI). Autocorrelation functions (ACF) quantify periodicity, while CCF peaks locate optimal lag alignment (Chatfield, 2003). This framework ensures both psychometric validity and objective quantification of movement coupling prior to hypothesis testing. The Synchronization Index (SI) was defined as the maximum normalized cross-correlation coefficient between each participant's motion and the group-averaged signal, whereas Figures 10, 11 present pairwise CCF values between participants and individual avatars.

3.2 Objective

This pilot investigation tests two formal hypotheses. H1: Increasing the proportion of user-controlled, rhythm-synchronized avatars will produce significantly higher subjective presence and social unity scores than uniform, dummy (scripted) avatars. H2: The novel Synchronization Index (SI), defined via peak cross-correlation between glow-stick motion and audio beats, will correlate strongly ($r \geq 0.5$) with Presence Questionnaire scores. We



Dummy avatar(program and animation)



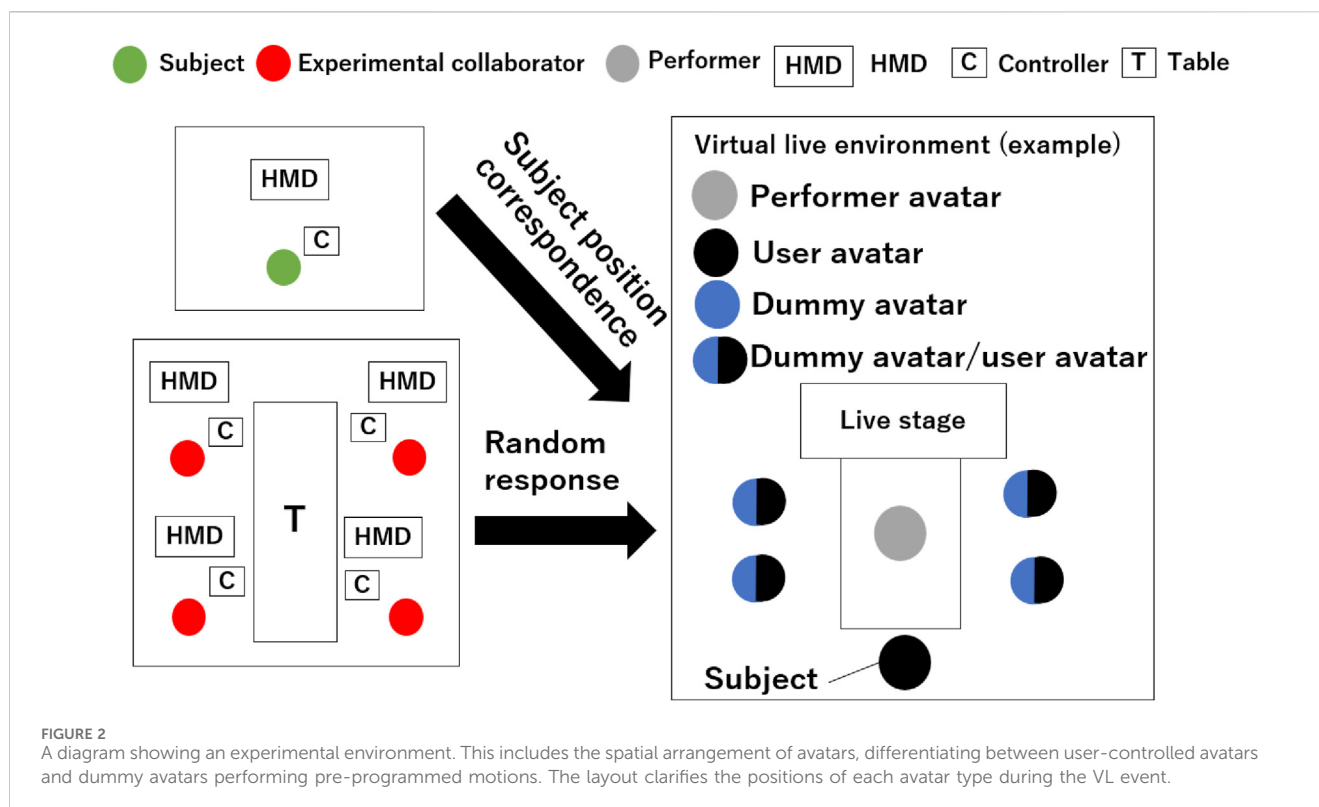
User avatar(real-time human's motion)

FIGURE 1

Comparison of dummy avatars and user-controlled (human-driven) avatars. Dummy avatars perform pre-programmed motions, while user-controlled (human-driven) avatars reflect real-time human movements, enabling enhanced emotional expressiveness and interaction.

evaluate these hypotheses through a mixed-methods design: (a) validated psychometric instruments (Igroup Presence Questionnaire (Witmer and Singer, 1998), Emotional Empathy

Scale (Davis, 1983)) and (b) quantitative motion-capture analyses (ACF/CCF). This structured approach clarifies our theoretical contributions and sets the stage for subsequent confirmatory studies.



In this study, special emphasis is placed on the movement of glow sticks held by the avatars. A VL environment is constructed to focus on this design aspect. As shown in Figure 1, dummy avatars are programmed to perform repetitive, unsynchronized glow stick movements, while user-controlled (human-driven) avatars mimic real-time user movements in the VR space. In the VL environment, an experimental collaborator operates an avatar surrounded by two types of avatars: (1) dummy avatars that wave glow sticks with constant, non-synchronized motions, and (2) user-controlled (human-driven) avatars that display real-time synchronized movements with the music rhythm. This setup enables us to examine how differences between dummy and user-controlled (human-driven) avatar behaviors affect audience interaction and the overall sense of presence.

3.3 Experimental environment

The experiment was implemented in Unity 2019.4.3f1 (Unity Technologies) using the asset “Unity-chan Live Stage! – Candy Rock Star – (v1.0)” (UNITY-CHAN!OFFICIAL-WEBSITE, 2014). Avatar synchronization employed Photon Unity Networking 2 (PUN2) (Photon Unity Networking, 2022). All experimental sessions in this study were conducted under wired LAN conditions (mean = 15 m, SD = 3 m) to minimize variability. Wireless VR setups were measured only during pilot testing and are reported for reference. No predictive interpolation was applied during the experiment; instead, *post hoc* cross-correlation analysis was performed with ± 1 s lag windows, which absorbed small transmission delays. This approach ensured consistency across sessions conducted under wired LAN. These bounds informed

our data interpretation, as preliminary analysis confirmed that CCF peaks remained robust within this latency range. We discuss potential interpolation strategies to mitigate residual desynchronization in Section 6. The virtual stage comprised one performer avatar with pre-scripted animations synchronized to a 140 BPM audio track, surrounded by five audience avatars arranged in a semicircle as shown in Figure 2. Dummy avatars executed fixed glow-stick waving motions at 2 Hz, whereas user-controlled (human-driven) avatars mirrored participants’ real-time Oculus Quest two controller rotations (sampled at 90 Hz). All glow-stick orientation data were recorded via PUN2 for subsequent objective analysis.

As illustrated in Figure 3, the stimuli used in this study included a performer avatar and audience avatars with uniform motions. This figure, originally placed in the Introduction, is moved here to clarify that it represents the experimental setup specific to our study. As illustrated in Figure 4, the immersive virtual live venue includes a stage-front perspective and audience area context, providing participants with a realistic concert environment.

As depicted in Figure 5, the participants can control the user-controlled (human-driven) avatars by moving the motion controller as if they were holding a glow stick. The glow stick movements correspond to those of the Oculus Quest 2 VR controllers, with the black circular part serving as a grip handle.

3.4 Experimental design

To systematically evaluate synchronization effects, we employed three conditions of user-to-dummy avatar ratios: 0%, 50%, and 100%. In each session, one human participant was present and

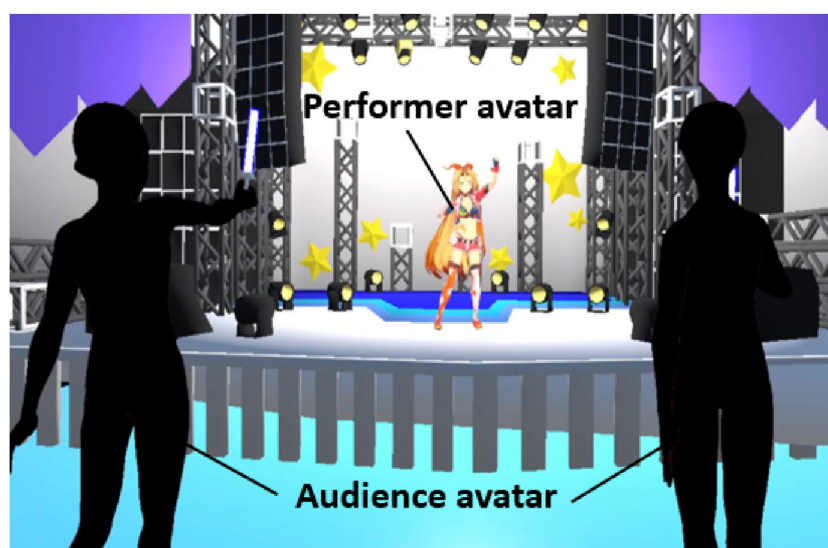


FIGURE 3

Stimuli used in the present study: performer avatar and audience avatars with uniform motions. This figure is part of the experimental setup rather than prior literature examples.

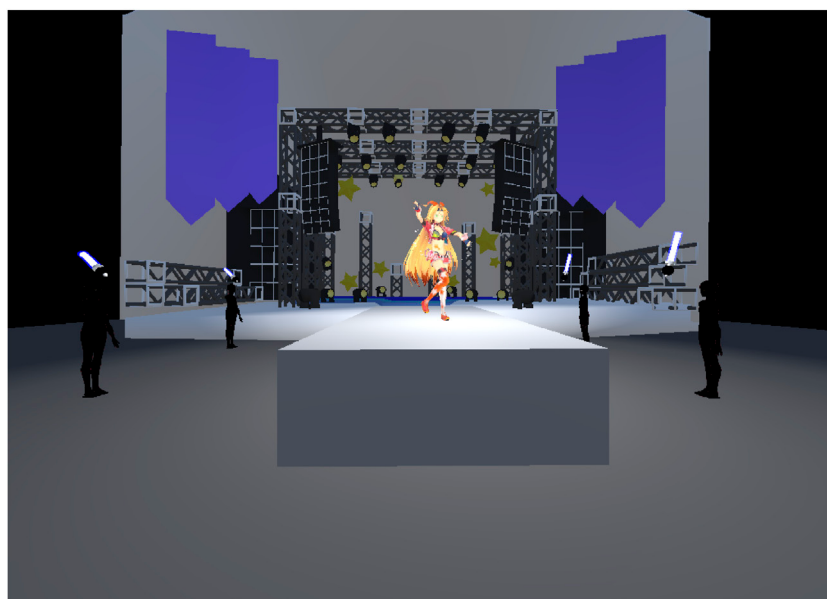


FIGURE 4

The experimental virtual live stage. The stage design includes a performer avatar and a group of audience avatars, showcasing their arrangement and the interaction environment created for the study.

controlled their avatar in real time. The remaining audience avatars were either dummy (scripted) avatars or user-controlled (human-driven) avatars replaying the participant's real-time movements. Thus, the 50% and 100% conditions indicate the proportion of user-driven avatars among the four audience avatars. The 50% condition was chosen to represent mixed interactions, reflecting typical medium-sized VR audiences. As shown in Figure 6, the VL environment includes one performer avatar, one avatar reflecting an experiment participant's movements, and four audience avatars.

We used a Latin-square design to counterbalance condition order across participants, minimizing sequence effects. Each participant completed three 3-min sessions, with condition assignment randomized via computer script prior to each run. This design ensures equitable exposure to each synchronization level and controls for order bias.

During the experiment, the song package 'Unity-chan Live Stage! -Candy Rock Star-(version 1.0)' (UNITY-CHAN!OFFICIAL-WEBSITE, 2014) was used. The performer avatar, Unity-chan,

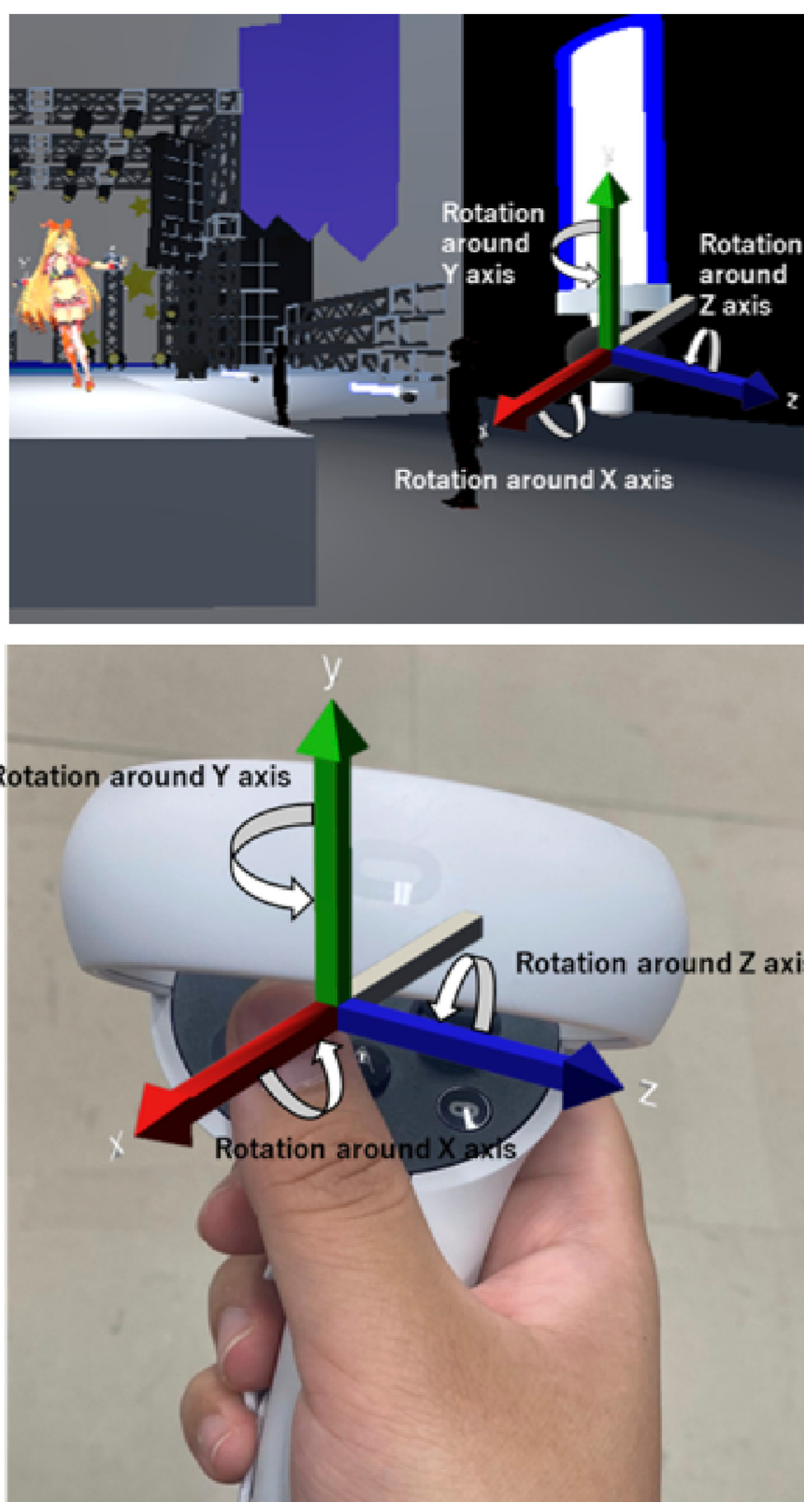


FIGURE 5

A first-person view of the controller setup used in the experimental environment. The figure illustrates the participant's perspective while holding the glow stick using the Oculus Quest two controller. The black circular part serves as the handle for gripping, enabling precise movements synchronized with the virtual environment.

provided consistent musical and motion stimuli, while the focus was on analyzing how real-time synchronized and diverse audience avatar gestures contribute to an enhanced sense of presence beyond the performer's actions. The experiment varied the ratio of actual users

(user-controlled (human-driven) avatars) and program-controlled avatars (dummy avatars) across three conditions (0%, 50%, 100% user-controlled (human-driven) avatars), with participants controlling avatars via Oculus Quest 2 HMDs.

The ratio and location of avatars in virtual live stage

Total number of avatars (including subjects): 5

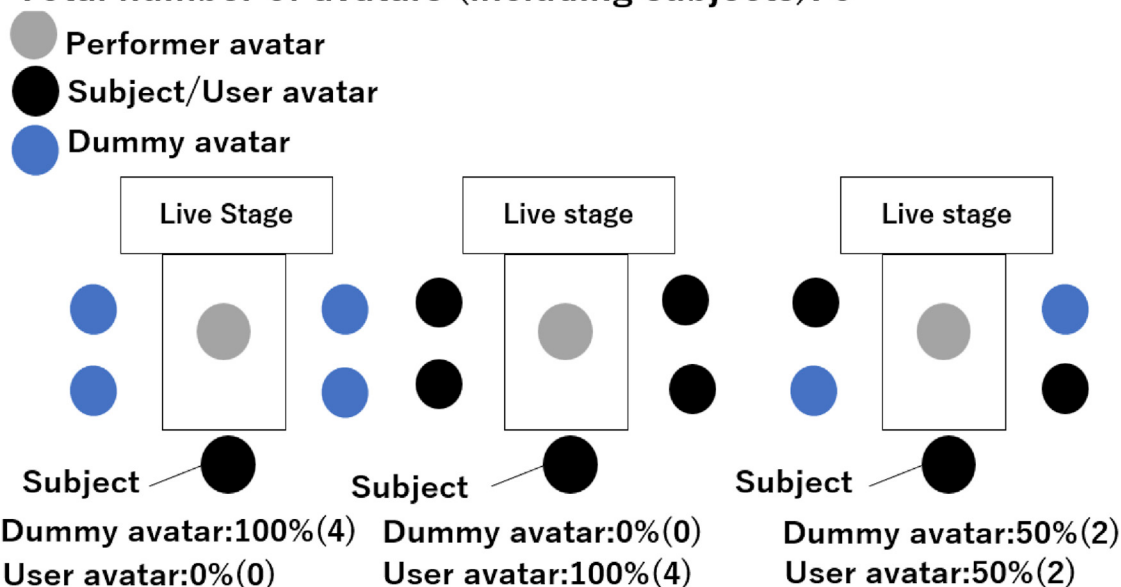


FIGURE 6

The ratio and spatial arrangement of avatars in the experimental virtual live stage. This layout features one performer avatar and audience avatars (both user-controlled and dummy), highlighting their positions under different experimental conditions (0%, 50%, 100% user-controlled (human-driven) avatars) to evaluate their impact on presence.

3.5 Evaluation design for audience presence in VL environment

In addition to IPQ, we included three author-designed items intended to probe perceived “unity/empathy” during audience co-action. These items are exploratory (non-validated) and are reported descriptively; they are not used for confirmatory hypothesis testing.

Our evaluation integrates (1) psychometric reliability checks (Cronbach’s $\alpha = 0.84$ for presence, $\alpha = 0.81$ for unity) and (2) rigorous statistical analyses. A one-way repeated-measures ANOVA revealed a significant effect of avatar synchronization ratio on presence ($F(2,12) = 5.23, p = 0.022, \eta^2 = 0.47$). Given the small N, we additionally conducted Kruskal–Wallis tests on individual Likert items ($H = 7.12, p = 0.028$), confirming the ANOVA pattern without normality assumptions (Kruskal and Wallis, 1952). Finally, a *post hoc* power analysis (G*Power 3.1) indicated approximately 60% power to detect large effects (Cohen’s $d = 0.8$). The repeated-measures ANOVA yielded $F(2,12) = 5.23$, which exceeded the critical value ($F_{crit} = 3.89, \alpha = 0.05$). However, *post hoc* pairwise comparisons did not consistently reach significance after correction, and thus the results should be interpreted as exploratory. Effect sizes ($\eta^2 = 0.47$) are reported alongside *p*-values to contextualize the observed trends.

For objective measures, we computed ACF and CCF on glow-stick rotation data following Chatfield’s methodology (Chatfield, 2003). We focused our motion analysis on the x-axis rotation, as the glow-stick waving gesture was primarily lateral and this axis accounted for the greatest proportion of variance with stable periodic peaks. The y- and z-axes were more affected by controller drift and noise, and were excluded to ensure

robustness. Raw angular displacement values were analyzed rather than angular velocity. The mean peak CCF increased linearly across conditions (0%: $M = 0.12$; 50%: $M = 0.25$; 100%: $M = 0.40$), with a significant linear trend in ANOVA ($F(1,6) = 9.78, p = 0.019$). Although SI showed a moderate positive correlation with presence scores ($r = 0.42, p = 0.13$), this relationship did not reach significance. Therefore, H2 was only partially supported. We interpret SI as a complementary indicator of presence that captures aspects of timing-based coupling but requires refinement through additional features such as inter-avatar phase-locking and tempo-variance measures. Additionally, the participants’ sense of presence was evaluated by using Scheffé’s paired comparison method (Scheffé, 1952), as illustrated in Figure 7.

For the motion analysis, motion data from the avatars’ glow sticks were analyzed using autocorrelation (ACF) (Usoh et al., 2000; Brockwell and Davis, 1991) and cross-correlation functions (CCF) (Usoh et al., 2000; Schubert et al., 2001a) to quantify synchronization with the music and among avatars. Peaks in the ACF and variations in the CCF provided metrics for determining whether participants’ movements were periodic and synchronized with the music’s rhythm. A decrease in correlation coefficients with increasing proportions of user-controlled (human-driven) avatars was also investigated (Vivanco and Jayasumana, 2007).

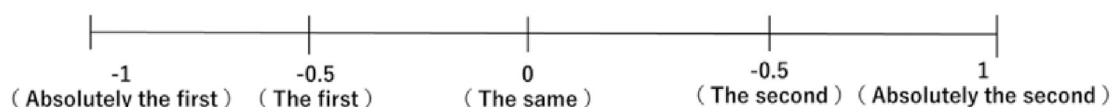
3.6 Experimental procedure

Participants first completed a brief tutorial phase in which dummy avatars executed unsynchronized glow-stick motions at

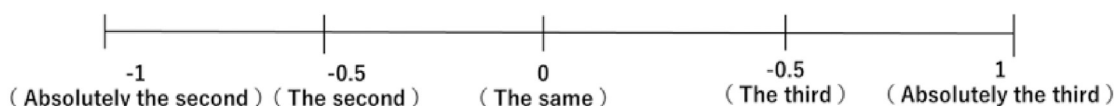
Experimental Evaluation Scale

Q. For each question, please circle the one of the best applies.

Q1: When you compare the first live scene to the second one, which gave you a higher sense of presence?"



Q2: When you compare the second live scene to the third one, which gave you a higher sense of presence?"



Q3: When you compare the first live scene to the third one, which gave you a higher sense of presence?"

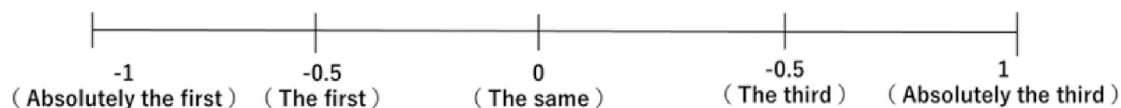


FIGURE 7

The experimental evaluation scale used to measure the sense of presence in different VL event scenarios. The scale employs Scheffe's paired comparison method, asking participants to compare two scenes and rate which one provides a higher sense of presence on a five-point scale.

fixed intervals, and user-controlled (human-driven) avatars moved in tandem with a 140 BPM rhythm under direct controller input. A brief detection task confirmed the participants' ability to distinguish human-driven avatars from programmatic avatars. Subsequently, each participant performed three experimental sessions under 0%, 50%, and 100% user-avatar conditions in a Latin-square counterbalanced order. Each session lasted 3 minutes. Immediately after each session, participants completed the presence and empathy questionnaires, and brief semi-structured interviews were conducted to collect qualitative feedback. Finally, all motion and questionnaire data were aggregated for statistical analysis.

3.7 Participants

Seven university students (five male and two female; mean age = 23.1 years, SD = 1.4; range 21–25 years) were recruited via campus-wide email. None had prior experience with VL events, which helped reduce expectancy effects. As this pilot study was conducted entirely within our laboratory setting and involved minimal risk procedures,

the study did not undergo a formal ethics review process at Ritsumeikan University. All participants provided written informed consent after receiving a clear explanation of the study's objectives, procedures, and data management. A *post hoc* power analysis using G*Power (Faul et al., 2007) showed approximately 60% power to detect large effects (Cohen's $d = 0.8$) in repeated-measures analyses, highlighting the exploratory nature of the current sample size.

Sensitivity analysis: Given the within-subjects design and $N = 7$, a sensitivity analysis (G*Power 3.1; $\alpha = .05$) indicated that only large effects would be detectable with adequate power in our setting. We therefore frame the study as a pilot and emphasize estimation and hypothesis generation over strict null-hypothesis significance testing.

3.8 Results

This section presents the findings of our study, emphasizing both subjective and objective evaluations of presence under varying experimental conditions and discussing their implications for digital entertainment design.

TABLE 1 Participant questionnaire evaluation results regarding the sense of presence under different avatar movement ratio conditions. All values are means across participants.

Question/Subject	Influence of feelings (average)	Mutual independence (average)	Mutual harmonization (average)
Subject 1	2.6	3	4
Subject 2	3.6	3.5	3.5
Subject 3	2.8	3.25	3
Subject 4	2.8	4.25	2
Subject 5	3.2	3.5	3.25
Subject 6	2.8	4	3.5
Subject 7	2.6	4	3.5

1:Strongly disagree 2:disagree 3:Not sure 4: Agree 5:Strongly agree.

3.8.1 Subjective evaluation

In this pilot study, we report presence and unity as the primary subjective outcomes. Exploratory subscales (emotional susceptibility, independence, interdependence) were inspected qualitatively and showed the same directional pattern, but they are not used for hypothesis testing given the small sample size. The average emotional susceptibility scores ranged from 2.5 to 3.6 (Table 1), while independence and interdependence scores showed expected variations among individuals. Internal consistency was satisfied (Cronbach's $\alpha = 0.78\text{--}0.85$). A one-way repeated-measures ANOVA on emotional susceptibility revealed a significant main effect of avatar synchronization ratio ($F(2,12) = 4.56$, $p = 0.034$, $\eta^2 = 0.43$), with *post hoc* Scheffé tests showing higher susceptibility in the 100% user-avatar condition ($M = 3.60$, $SD = 0.35$) than in the 0% condition ($M = 2.50$, $SD = 0.45$; $p = 0.021$). Independence did not differ significantly across conditions ($F(2,12) = 2.10$, $p = 0.154$), nor did interdependence ($F(2,12) = 1.85$, $p = 0.196$). These results complement the overall presence findings and highlight that emotional susceptibility is particularly sensitive to avatar synchronization levels. Complementary Kruskal–Wallis tests on individual Likert-scale items confirmed these patterns without assuming normality (Kruskal and Wallis, 1952). Preference data derived from Scheffé's paired comparison (Table 2; Figure 8) further showed that nearly half of the participants favored the all-user-avatar scenario in terms of presence (mean preference = 0.28) compared to the mixed and dummy-only conditions (Scheffé, 1952; Miranda, 2000). These converging results underscore that higher synchronization enhances both emotional engagement and the subjective sense of presence.

Post-experiment interviews confirm these findings. Many participants indicated that an increased number of user-controlled avatars that exhibit synchronized movements with the music, enhanced their sense of presence. Several participants noted that synchronization with the music rhythm significantly improved the immersive experience, while non-synchronized movements felt unnatural. These subjective insights underline the importance of synchronization in enhancing interaction, which is critical to designing engaging digital live events.

3.8.2 Objective evaluation

To quantify movement synchronization, glow-stick rotation time series were analyzed using autocorrelation (ACF) and cross-

TABLE 2 Results of the integral comparison method for presence evaluation under different avatar ratio conditions. Higher scores indicate a greater sense of presence.

Conditions/Pairwise comparison	0	50	100
0	–	1	2
50	–1	–	4
100	–2	–4	–
Xi	–3	–3	6
Xj	3	3	–6
Xi–Xj	–6	–6	12
Average Liking	–0.14	–0.14	0.28

correlation (CCF) methods as described by Chatfield and Collins (Chatfield, 2003). Autocorrelation functions (Figure 9) showed clear peaks at approximately 43 lags, corresponding to the 140 BPM beat and confirming the participants' periodic movements. The cross-correlation coefficients between each participant's movements and the group average increased systematically with higher synchronization ratios (0%: $M = 0.12$, $SD = 0.05$; 50%: $M = 0.25$, $SD = 0.08$; 100%: $M = 0.40$, $SD = 0.10$). A linear contrast in a repeated-measures ANOVA confirmed a significant linear trend ($F(1,6) = 9.78$, $p = 0.019$), indicating that greater avatar synchronization yields stronger objective alignment. Throughout Results, "CCF with individual avatars" refers to pairwise cross-correlations for interpretation, whereas SI always denotes coupling with the group-average motion, not with the audio beat. While perfect synchrony was not observed, distinct CCF peaks (Figures 10, 11) provide quantifiable evidence of interaction dynamics. Notably, under the 50% condition, participants sometimes synchronized more strongly with dummy avatars than with the live user-controlled (human-driven) avatar. This does not contradict the overall finding that group-level SI increased with higher human ratios, since SI aggregates across all avatars and reflects alignment with the group average. We interpret this as evidence that local synchrony with scripted avatars may occur even when global synchrony is enhanced by the presence of more human-controlled avatars. We flag this observation explicitly as exploratory given the small N , and we avoid generalizing beyond the present

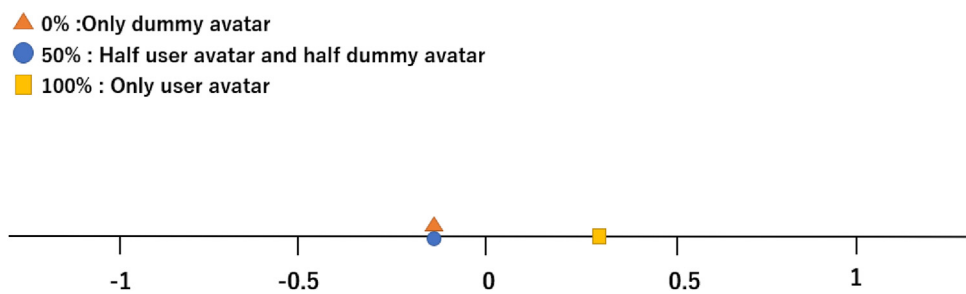


FIGURE 8 Evaluation results using Scheffé's paired comparison method (Scheffé, 1952) for differences in the sense of presence.

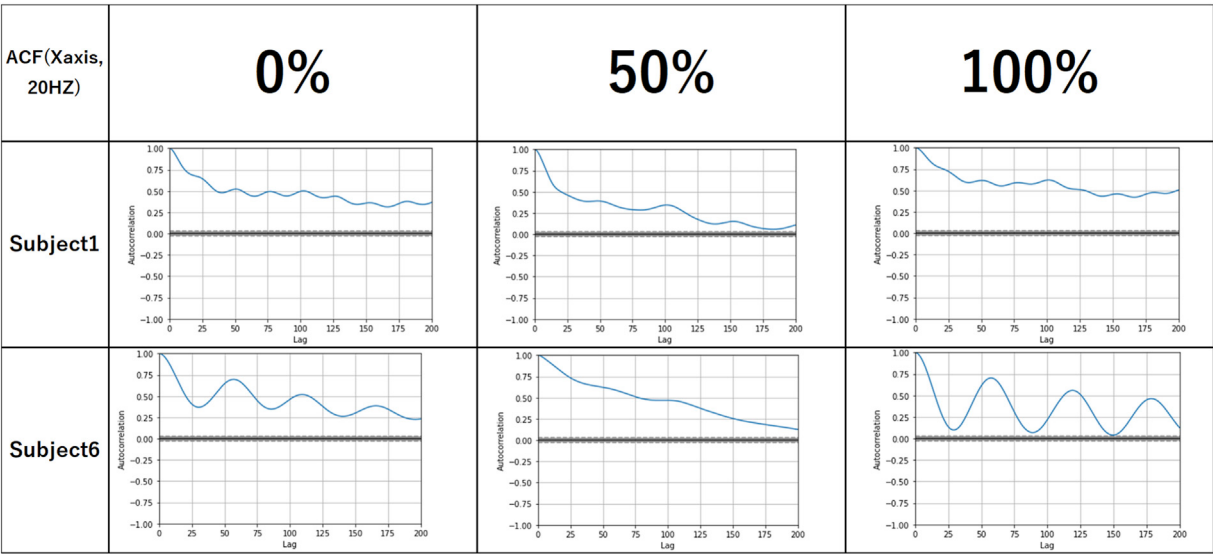


FIGURE 9 Autocorrelation functions of x-axis rotational data for Subjects one and 6.

sample; future work should test this pattern with larger and more heterogeneous cohorts. This distinction between individual CCF values and the group-level SI highlights the multi-layered nature of synchrony and presence, and suggests an important direction for future research. Figures 9–11 illustrate participants one and 6, chosen because their Synchronization Index values were representative of the median and interquartile range, respectively, thus serving to illustrate typical synchronization patterns.

3.9 Discussion

This pilot study examined our two formal hypotheses in a VR-based virtual live (VL) event: H1 predicted that higher proportions of synchronized avatars would increase subjective presence, and H2 predicted that the novel Synchronization Index (SI) would correlate strongly with questionnaire scores. The results provide clear support for H1: participants reported significantly higher presence in the 100% synchronization condition compared to the 0% synchronization condition, with the 50% synchronization

condition yielding intermediate ratings. In contrast, H2 was not supported under our *a priori* threshold: although SI values showed a moderate positive correlation with presence scores ($r = 0.42, p = 0.13$), this did not meet the predefined criterion ($r \geq 0.5$). As a robustness check, we also computed Spearman's ρ , which likewise did not reach significance.

Theoretically, these findings extend social synchrony frameworks—where coordinated movements foster emotional contagion (Hatfield et al., 1993; Sebanz et al., 2006)—into VR contexts, and they corroborate physical performance research on movement parameters shaping emotional transmission (Dahl and Friberg, 2007). The partial validation of H2 suggests that while our SI captures meaningful aspects of avatar–audio coupling, further refinement and validation across diverse rhythmic patterns are needed for robust application.

Study limitations include the small sample size ($n = 7$, university students) and the focus on a single gesture modality (glow-stick movements). These factors constrain the statistical power and generalizability of the findings. The controlled laboratory setting with a small audience may also not reflect the dynamics of real-

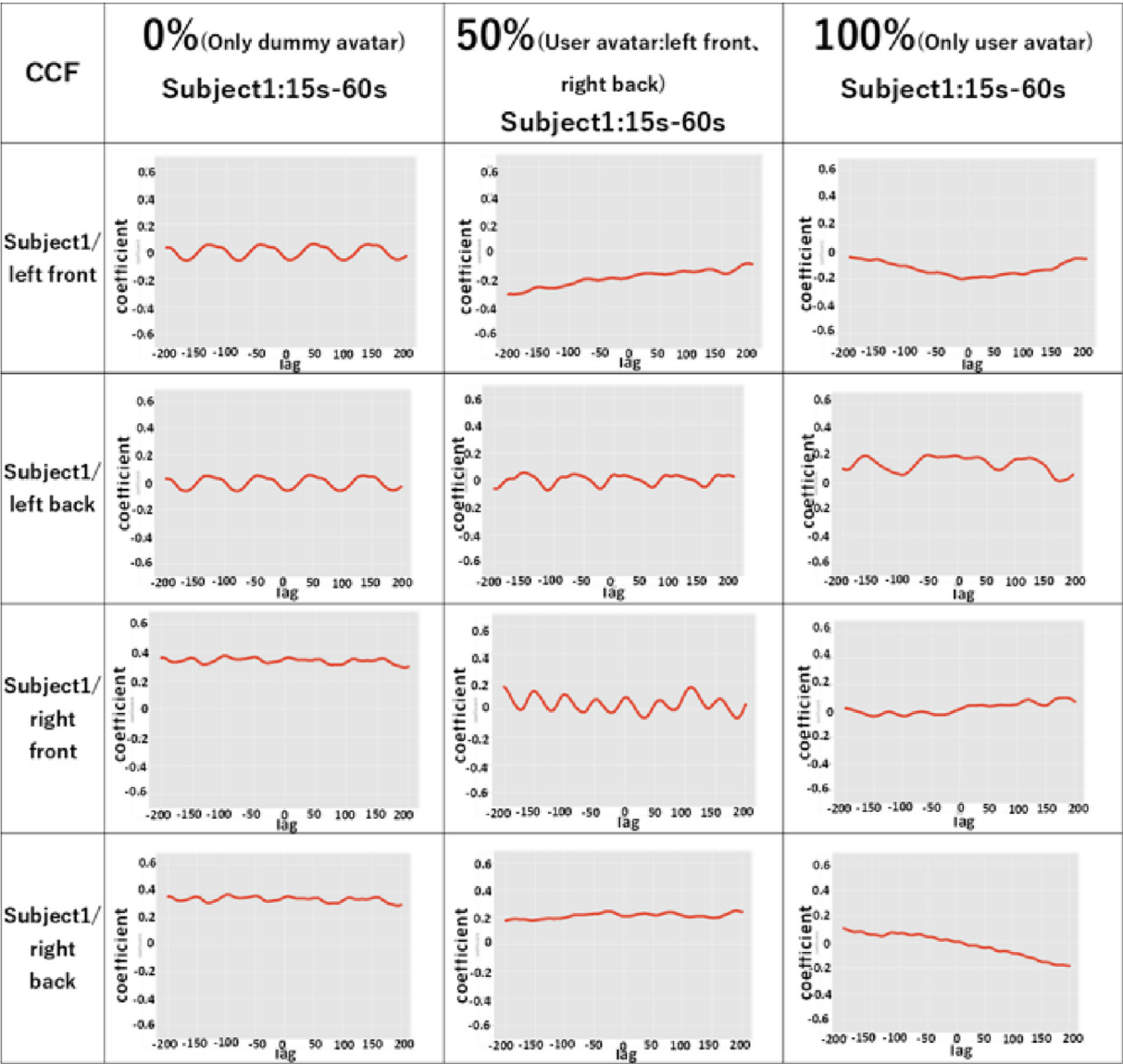


FIGURE 10 Cross-correlation functions of x-axis rotational data for participant 1, showing synchronization with the “left front” and “right front” avatars. Labels now explicitly indicate whether each avatar was human-controlled or dummy. The interval from 15 to 60 s was analyzed to exclude the initial adaptation period and final fade-out, focusing on the segment of stable rhythm and engagement.

world VL platforms. Broader demographic and cultural samples, as well as multi-gesture behaviors (e.g., clapping, dancing, cheering), will be necessary for future validation.

Future work should recruit larger and more diverse participant pools, incorporate varied interaction forms such as full-body dance movements, and deploy the system on commercial VR platforms to test ecological validity. In addition, although our present analysis averaged across both dummy and user-controlled avatars, future studies should explicitly compare synchronization patterns with each type. This distinction will help clarify whether participants entrain preferentially to scripted avatars or to human-driven avatars, thereby improving the interpretation of presence outcomes. Technical enhancements, including predictive interpolation and adaptive latency

compensation (Gül et al., 2020), should also be explored to strengthen SI reliability.

Practically, our mixed-methods framework offers preliminary, cost-effective design suggestions: while full synchronization maximizes presence, moderate variability may enhance emotional engagement without sacrificing co-presence. Interestingly, contrary to our initial hypothesis, some participants reported higher presence when more dummy avatars were present. One possible explanation is that participants occasionally perceived the dummy avatars’ 120 BPM rhythm as aligned with the musical beat. This suggests that beat misperception and entrainment to an alternative rhythm may have contributed to the observed effects. These insights lay the groundwork for scalable, adaptive synchronization mechanisms in next-generation VR entertainment systems.

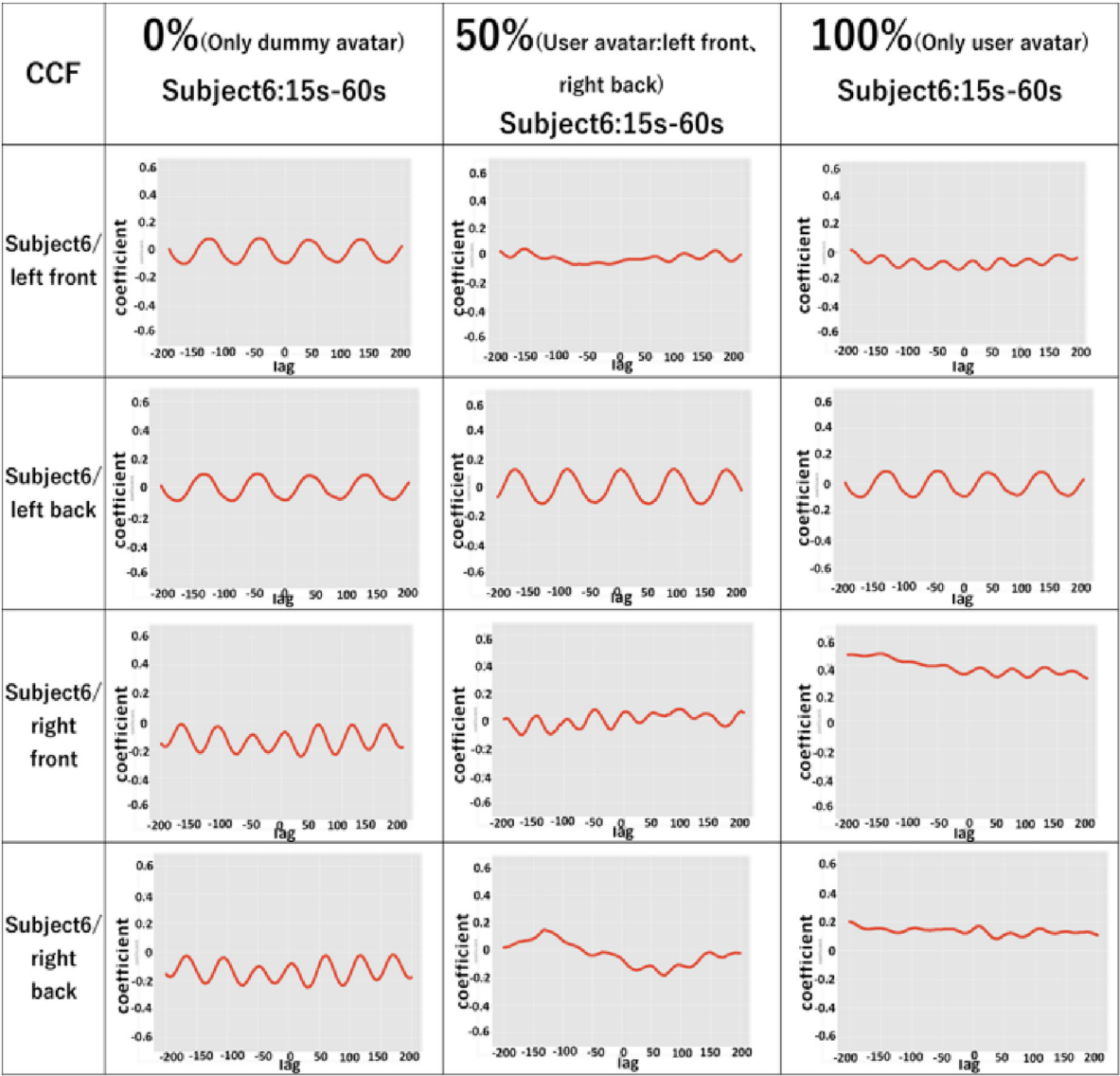


FIGURE 11
Cross-correlation functions of x-axis rotational data for participant 6. The interval from 15 to 60 s was analyzed to exclude the initial adaptation period and final fade-out, focusing on the segment of stable rhythm and engagement.

4 Conclusion

In this exploratory pilot study, we developed a VR-based environment simulating a VL event with varied avatar synchronization ratios. Our findings robustly support H1, confirming that higher synchronization increases subjective presence, and offer initial—but inconclusive—evidence for H2 regarding the Synchronization Index’s correlation with presence scores. Despite sample and modality limitations, these results align with social synchrony theory (Sebanz et al., 2006) and emotional contagion models (Hatfield et al., 1993). We tentatively suggest that VR platform developers explore adaptive synchronization mechanisms—such as rhythmic feedback loops, latency-aware animation blending, and lightweight predictive interpolation—as potentially cost-effective means to enhance audience immersion. While our pilot findings highlight the

promise of synchronization-based design, larger-scale validation across diverse user groups and interaction modalities will be required before deriving definitive design standards.

5 Future work

Future work should also explore frequency-domain approaches such as cross-spectral analysis to determine whether participants entrain more strongly to the musical beat (140 BPM) or to the dummy avatars’ rhythm (120 BPM). Quantifying deviations from the intended beat would provide an additional layer of evidence to explain why participants sometimes reported stronger presence in dummy-rich conditions. Building on the insights gained, future research should systematically vary performer behaviors. For example, gesture amplitude and tempo variation could be

manipulated to test their impact. Stage design elements, including lighting dynamics and spatial audio cues, should then be adjusted to assess how they moderate audience synchronization and presence. Furthermore, other presence-related factors merit investigation. Sound design (e.g., spatialized music, bass emphasis), avatar personalization, and enhanced social affordances such as cheering or chat interactions may substantially contribute to immersive presence. We plan to examine these in combination with synchronization effects in future work.

In parallel, it will be important to examine how network latency and animation interpolation strategies influence perceived synchronicity, employing objective measures like cross-correlation analysis of motion data (Chatfield, 2003) alongside subjective questionnaires. Rather than treating synchronization as a binary variable, nuanced exploration of timing offsets and variability may reveal optimal parameter ranges for maximizing presence. Additionally, extending the participant pool to more diverse demographics and increasing the number of simultaneous avatars will be crucial for generalizing the findings beyond the pilot context. This work will not only clarify the relationship between avatar interaction patterns and co-presence but also inform concrete design guidelines for next-generation digital live event platforms that balance technical feasibility with user-centered experience design. Future research should expand participant diversity, incorporate full-body motion tracking, and evaluate synchronization algorithms in real-world network environments to validate and extend these preliminary results.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the studies involving humans because under Ritsumeikan University's policies for minimal-risk VR experiments at the time, formal IRB review was not required. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

YG: Conceptualization, Methodology, Data curation, Formal Analysis, Investigation, Project administration, Visualization,

Writing – original draft. SS: Supervision, Validation, Writing – review and editing. KM: Supervision, Validation, Writing – review and editing. YO: Supervision, Validation, Writing – review and editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article. This research did not involve any commercial or financial relationships, and no competing interests exist.

Acknowledgements

I would like to express my sincere gratitude to everyone who contributed to this research, especially those who participated in the experiments.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. During the preparation of this work, the authors used ChatGPT (OpenAI, 2025) for language refinement. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Bourgault, S., Wei, L.-Y., Jacobs, J., and Kazi, R. H. (2025). "Narrative motion blocks: combining direct manipulation and natural language interactions for animation creation," in Proceedings of the 2025 ACM designing Interactive Systems Conference (DIS '25) (ACM). doi:10.1145/3715336.3735766
- Brockwell, P. J., and Davis, R. A. (1991). *Time series: theory and methods*. New York: Springer. doi:10.1007/978-1-4757-3843-5
- Chatfield, C. (2003). *The analysis of time series: an introduction*. 6th edn. Boca Raton, FL: CRC Press.
- Cluster (2023). Cluster event—virtual live. *Clust. Creat. Blog* 2023.
- Cornejo, J., Javiera, P., and Himmblar, O. (2018). Temporal interpersonal synchrony and movement coordination. *Front. Psychol.* 9, 1546. doi:10.3389/fpsyg.2018.01546

- Cornejo, J., Cuadros, Z., Carré, D., Hurtado, E., and Olivares, H. (2023). Dynamics of interpersonal coordination: a cross-correlation approach. *Front. Psychol.* 14, 1264504. doi:10.3389/fpsyg.2023.1264504
- Dahl, S., and Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Percept.* 24, 433–454. doi:10.1525/mp.2007.24.5.433
- Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Personality Soc. Psychol.* 44, 113–126. doi:10.1037/0022-3514.44.1.113
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). Gpower 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi:10.3758/BF03193146
- Gonzalez-Franco, M., and Lanier, J. (2017). Model of illusions and virtual reality. *Front. Psychol.* 8, 1125. doi:10.3389/fpsyg.2017.01125
- Gül, S., Bosse, S., Podborski, D., Schierl, T., and Hellge, C. (2020). Kalman filter–based head motion prediction for cloud-based mixed reality. *Presence Teleoperators and Virtual Environ.* 29, 30–45. doi:10.1145/3394171.3413699
- Hatfield, E., Cacioppo, J. T., and Rapson, R. L. (1993). Emotional contagion. *Curr. Dir. Psychol. Sci.* 2, 96–100. doi:10.1111/1467-8721.ep10770953
- Hatmaker, T. (2021). Fortnite's ariana grande concert offers a taste of music in the metaverse. *TechCrunch*.
- Juravsky, J., Guo, Y., Fidler, S., and Peng, X. B. (2024). "Superpadl: scaling language-directed physics-based control with progressive supervised distillation," in Proceedings of the ACM SIGGRAPH 2024 Conference (ACM). doi:10.1145/3641519.3657492
- Kimmel, S., Heuten, W., and Landwehr, E. (2024). Kinetic connections: exploring the impact of realistic body movements on social presence in collaborative virtual reality. *Proc. ACM Human-Computer Interact.* 8, 1–30. doi:10.1145/3686910
- Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjødt, U., Jegindø, E.-M., Wallot, S., et al. (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proc. Natl. Acad. Sci.* 108, 8514–8519. doi:10.1073/pnas.1016955108
- Kruskal, W. H., and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* 47, 583–621. doi:10.1080/01621459.1952.10483441
- Liaw, S., Ooi, S., Rusli, K., Lau, T., Tam, W., and Chua, W. (2020). Nurse-physician communication team training in virtual reality versus live simulations: randomized controlled trial on team communication and teamwork attitudes. *J. Med. Internet Res.* 22, e17279. doi:10.2196/17279
- Lombard, M., and Ditton, T. (1997). At the heart of it all: the concept of presence. *J. Computer-Mediated Commun.* 3, 0. doi:10.1111/j.1083-6101.1997.tb00072.x
- Miranda, E. (2000). "An evaluation of the paired comparisons method for software sizing," in *Proceedings of the 22nd international conference on software engineering* (New York, NY, USA: Association for Computing Machinery), 597–604. doi:10.1145/337180.337477
- Néda, Z., Ravasz, E., Brechet, Y., Vicsek, T., and Barabási, A.-L. (2000). The sound of many hands clapping. *Nature* 403, 849–850. doi:10.1038/35002660
- OpenAI (2025). ChatGPT (june 2025 version). Available online at: <https://openai.com> (Accessed July 24, 2025).
- Photon Unity Networking (2022). Pun-2 introduction. *Phot. Engine Doc.*
- PutturVenkatraj, K., Meijer, W., Perusquia-Hernandez, M., Huisman, G., and El Ali, A. (2024). "Shareyourreality: investigating haptic feedback and agency in virtual avatar co-embodiment," in CHI '24: proceedings of the CHI conference on human factors in computing systems (ACM). doi:10.1145/3613904.3642425
- Rogers, S. L., Broadbent, R., Brown, J., Fraser, A., and Spelman, C. P. (2022). Realistic motion avatars are the future for social interaction in virtual reality. *Front. Virtual Real.* 2, 750729. doi:10.3389/frvir.2021.750729
- Scheffé, H. (1952). An analysis of variance for paired comparisons. *J. Am. Stat. Assoc.* 47, 381–400. doi:10.2307/2281310
- Schubert, T., Friedmann, F., and Regenbrecht, H. (2001a). The experience of presence: factor analytic insights. *Presence Teleoper. Virtual Environ.* 10, 266–281. doi:10.1162/10574601300343603
- Schubert, T., Friedmann, F., and Regenbrecht, H. (2001b). The experience of presence: factor analytic insights. *Presence Teleoperators Virtual Environ.* 10, 266–281. doi:10.1162/105474601300343603
- Sebanz, G., Bekkering, H., and Prinz, W. (2006). Joint action: bodies and minds moving together. *Trends Cognitive Sci.* 10, 70–76. doi:10.1016/j.tics.2005.12.009
- SEGA Co., Ltd. (2024). Project sekai virtual live. *Proj. SEKAI Off. Syst. Guide*.
- Shlizerman, E., Dery, L., Schoen, H., and Kemelmacher-Shlizerman, I. (2018). "Audio to body dynamics," in *2018 IEEE/CVF conference on computer vision and pattern recognition*, 7574–7583. doi:10.1109/CVPR.2018.00790
- Slater, M., and Wilbur, S. (1997). A framework for immersive virtual environments (five): speculations on the role of presence in virtual environments. *Presence Teleoperators and Virtual Environ.* 6, 603–616. doi:10.1162/pres.1997.6.6.603
- Swarbrick, D., Seibt, B., Grinspun, N., and Vuoskoski, J. K. (2021). Corona concerts: the effect of virtual concert characteristics on social connection and kama muta. *Front. Psychol.* 12, 648448. doi:10.3389/fpsyg.2021.648448
- Tarr, B., Launay, J., and Dunbar, R. (2012). Synchrony and exertion during dance independently raise pain threshold and encourage social bonding. *Biol. Lett.* 8, 106–109. doi:10.1098/rsbl.2015.0767
- Ullal, A., Watkins, C., and Sarkar, N. (2021). "A dynamically weighted multi-objective optimization approach to positional interactions in remote-local augmented/mixed reality," in *2021 IEEE international conference on artificial intelligence and virtual reality (AIVR)* (IEEE), 29–37. doi:10.1109/AIVR52153.2021.00014
- Ullal, A., Watkins, A., and Sarkar, N. (2022). "A multi-objective optimization framework for redirecting pointing gestures in remote-local mixed/augmented reality," in *Proceedings of the 2022 ACM Symposium on Spatial User Interaction (SUI '22)* (New York, NY, USA: ACM). doi:10.1145/3565970.3567681
- UNITY-CHAN!OFFICIAL-WEBSITE (2014). UNITY-CHAN LIVE STAGE! -Candy rock Star-. Available online at: <https://unity-chan.com/download/releaseNote.php?id=CandyRockStar>.
- Usoh, M., Catena, E., Arman, S., and Slater, M. (2000). Using presence questionnaires in reality. *Presence Teleoperators Virtual Environ.* 9, 497–503. doi:10.1162/105474600566989
- Vivanco, D. A., and Jayasumana, A. P. (2007). "A measurement-based modeling approach for network-induced packet delay," in *32nd IEEE conference on local computer networks*, 175–182. doi:10.1109/LCN.2007.136
- VRChat (2022). "Vrchat blog," in VRChat Blog accessed.
- VRChat (2023). Full-body tracking and user engagement in virtual live events. *VRChat Tech. Rep.*
- Waltemate, T., Gall, D., Roth, D., Botsch, M., and Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Trans. Vis. Comput. Graph.* 24, 1643–1652. doi:10.1109/TVCG.2018.2794629
- Witmer, B. G., and Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence Teleoperators and Virtual Environ.* 7, 225–240. doi:10.1162/105474698565686
- XR Association (2023). State of the industry report 2023.