Computers and Artificial Intelligence; Vol.3, No.6; 2026
E-ISSN: 3006-3728 P-ISSN: 3006-371X
DOI: https://doi.org/10.70267/cai.v3n6.197210
Published by: Zeus Press

The Effects of Virtual Avatar Representation on Social Presence in Social Virtual Reality: A Systematic Review

Zhijing Xie*

School of Artificial Intelligence, Anhui University of Science and Technology, Huainan 232001, China

**Corresponding author: Zhijing Xie.*

Abstract

Social virtual reality (Social VR) relies on avatars to convey another person's appearance, actions, and voice. Yet more detailed representations do not always produce stronger social presence. This review examines experimental evidence on avatar appearance, behavior, self-related features, and multimodal cues, while distinguishing overall social presence from co-presence and other measured subdimensions. Using PRISMA 2020 as a reporting guide, we searched IEEE Xplore (20 August 2026) and the ACM Digital Library (28 August 2026) for empirical research published since 2021. The searches identified 382 records; 41 reports underwent full-text assessment, and 21 were included. Twelve core reports addressed synchronous human-to-human Social VR, while nine extended reports examined related immersive settings. Findings were synthesized qualitatively across four dimensions of avatar representation. Visible partners or hands, some facial cues, and identity-congruent avatar features were associated with higher social presence or co-presence under particular conditions. Comparisons involving additional motion capture, avatar customization, and spatial audio did not establish consistent advantages in the outcomes assessed. Multisensory and AI-avatar systems sometimes improved social experience, but their compound interventions do not isolate the effects of individual cues. Differences in task demands, interaction partners, and measurement approaches limit direct comparisons. Together, the findings suggest evaluating avatar features in relation to the social information an interaction requires, rather than assuming that greater representational fidelity is inherently beneficial.

Keywords

social virtual reality, avatar representation, social presence, co-presence, embodiment, nonverbal cues, multimodal interaction, systematic review

1. Introduction

In social virtual reality (Social VR), the other person is encountered through an avatar rather than directly. The avatar's appearance, movement, facial activity, gaze, and voice determine which social cues remain visible or audible during interaction. More elaborate representation can reproduce additional detail, but that does not necessarily make a partner's intentions easier to interpret.

Social presence refers to the perceived presence of another social actor in mediated interaction, not the sensation of occupying a virtual space [1,2]. Its measures include overall ratings and more specific dimensions such as co-presence, attention, and interdependence. These are related but non-equivalent outcomes; agency

© 2026 The Author(s). Authors retain copyright in their articles.

and body ownership concern the user's relationship to an avatar rather than the interpersonal experience measured here.

The experimental record illustrates why these distinctions matter. Aseeri and Interrante [3] found differences in co-presence, but not overall social presence, across avatar conditions. Kimmel et al. [4] reported no significant differences in the assessed social-presence dimensions across motion-capture conditions, although interpersonal distance changed. Neither finding supports a general rule that greater representational detail reliably improves social presence.

Similar care is needed when discussing identity. Kim et al. [5] compared avatar attributes that matched or did not match participants' appearance, voice, and name; Macey et al. [6] examined whether participants customized or were assigned an avatar. Their different outcomes should not be read as contradictory findings on a single variable. Matching one's identity and choosing an avatar are distinct manipulations.

Multisensory systems introduce another difficulty: effects may be reported for an entire configuration rather than its components. VR Whispering [7], for instance, combined several visual, auditory, and tactile cues. Positive results for such a configuration cannot establish which individual cue produced the difference.

Cross-study comparison is further complicated by differences in tasks, partner types, and outcome measures. A result from synchronous human collaboration need not transfer to a prerecorded or AI-powered partner. Even studies that use the term social presence may report different subdimensions. Such distinctions are necessary when interpreting both significant and non-significant comparisons.

This review examines the evidence across four categories: appearance representation, behavioral interaction, self-identification, and multimodal representation. It compares what was manipulated, what was measured, and which explanations the reported designs can support. Four research questions guide the analysis:

RQ1: Which dimensions of virtual avatar representation are primarily manipulated or varied in Social VR and related immersive social interaction studies?

RQ2: How do different types of avatar representation affect social presence, co-presence, and their related subdimensions?

RQ3: Is the effect of avatar representation on social presence influenced by research contexts such as interaction tasks, interaction formats, partner types, device configurations, and measurement approaches?

RQ4: What evidence gaps and methodological limitations remain in current research on the relationship between avatar representation and social presence, and which directions should future studies prioritize?

2. Methods

2.1 Review Design and Scope

This review examines empirical research on avatar representation and interpersonal social presence in immersive VR/XR interaction. The search and selection processes were reported with reference to PRISMA 2020 [8]. Synchronous human-to-human Social VR constitutes the core evidence; eligible studies of asymmetric or asynchronous avatar-mediated interaction were retained as extended evidence.

2.2 Databases and Search Strategy

The formal search was confined to IEEE Xplore, searched on 20 August 2026, and the ACM Digital Library, searched on 28 August 2026. Author websites, institutional repositories and other open-access sources were used only to retrieve full texts, not to identify additional records.

The original search string used in this review was: [[avatar OR embodiment OR "virtual character" OR "digital character" OR "virtual human" OR "digital human"] AND ["social presence" OR copresence OR "self-presence" OR "shared presence"] AND [VR OR "virtual reality" OR "social VR" OR XR OR "social virtual reality"]] AND [01/01/2021 TO 31/12/2026].

The query combined terms for avatars, social presence and VR/XR environments. The date filter ran from 1 January 2021 to 31 December 2026, but records were eligible only if available by the relevant August 2026

© 2026 The Author(s). Authors retain copyright in their articles.

search date. Self-presence and shared presence broadened the search; neither alone satisfied the review's interpersonal social-presence outcome requirement. Specific cues such as facial movement, gaze, hand tracking and spatial audio were identified during screening rather than searched separately.

2.3 Inclusion and Exclusion Criteria

Eligible reports had to present empirical data from human participants in a VR/XR social interaction setting, compare an identifiable aspect of avatar representation, and measure social presence, co-presence or a directly corresponding interpersonal social-presence outcome. Merely using an avatar in the interface was not sufficient. The publication also had to fall within the specified search period and be available by the relevant database cutoff.

Reports were excluded if they assessed only spatial presence, immersion, agency, body ownership, trust, likability or performance without an interpersonal social-presence outcome. Purely technical development papers, theoretical work, reviews, posters, conference abstracts and extended abstracts were also excluded. Eligible asymmetric XR and asynchronous avatar-mediated interactions were retained as extended evidence rather than treated as equivalent to synchronous human-to-human Social VR.

2.4 Literature Screening and Full-Text Assessment

A single reviewer screened titles and abstracts, retaining uncertain but potentially relevant records for full-text assessment. At the full-text stage, eligibility was checked against the interaction setting, actual avatar-related manipulation, social-presence measure and experimental comparison. If a PDF's title and contents differed, the decision followed the actual paper. Potentially overlapping publications were compared by participants, tasks, conditions and data sources.

IEEE Xplore returned 341 records and the ACM Digital Library 41, yielding 382 records with no duplicates recorded. Title and abstract screening excluded 341 records: 104 partially outside the review scope and 237 unrelated. Of the 41 reports assessed in full text, 20 were excluded: the VR system had not yet been applied in practice (n = 6), the environment was not immersive VR (n = 5), social presence or co-presence was not measured (n = 5), or no avatar-related variable was examined (n = 4). The final synthesis included 21 reports (Fig. 1).

The included reports comprised 12 core and nine extended reports. Core reports primarily concerned synchronous, immersive interaction between human participants. Extended reports addressed settings such as asymmetric XR, avatars controlled by experimenters or actors, spatial audio, asynchronous interaction and AI-powered embodied avatars. This distinction reflects proximity to conventional synchronous Social VR, not study quality. Reports, rather than individual experiments, are the counting unit: report [9] includes two independently recruited experiments (N = 36 each) but counts as one report.

2.5 Data Extraction

For each included report, the extracted information covered the study setting, participant sample, device or platform, type of interaction partner, avatar representation comparison, social-presence instrument, reported statistical findings and the direction of those findings. Where a publication contained separate experiments or conditions, their results were recorded separately without counting the publication more than once.

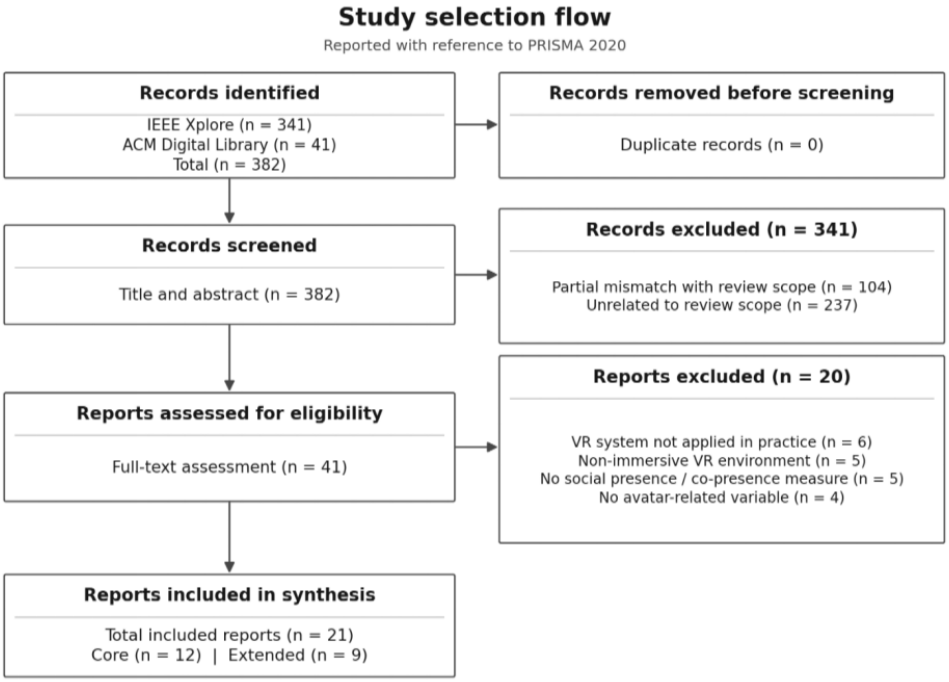
Overall social presence was distinguished from co-presence and other reported subdimensions, such as attentional allocation, perceived message understanding, affective understanding and behavioral interdependence [2]. Significant, nonsignificant and hypothesis-inconsistent outcomes were retained. Task-performance and other behavioral measures were considered alongside social-presence results, not substituted for them.

2.6 Thematic Coding and Evidence Synthesis

The studies were coded by the social information changed by the avatar representation: appearance, behavioral interaction, self-identification or multimodal representation. A report could contribute evidence to more than one theme, but it entered the overall count of 21 reports only once.

© 2026 The Author(s). Authors retain copyright in their articles.

Figure 1: Study selection flow with reference to PRISMA 2020 [8]



Within each theme, the synthesis compared the manipulated representation, the outcome construct and the direction of the reported effect. Differences were then considered against task demands, interaction format, partner type and measurement choices. Results from conditions combining several cues were interpreted at the level of the complete intervention unless the study independently tested the components. Nonsignificant findings were recorded as nonsignificant, not interpreted as evidence of equivalence. A qualitative synthesis was used instead of pooling effect sizes because the manipulations, study contexts and outcome measures were not sufficiently comparable.

Table 1A: Characteristics and Main Findings of Core Included Reports

Study / sample	Interaction context	Representation comparison	SP / CP measures and findings
[3] Aseeri & Interrante (2021) N = 36	Dyadic VR collaboration	No avatar vs scanned avatar vs real-body representation	Overall SP: not significant. CP: scanned and real representations > no avatar ($\chi^2(2) = 17.20$, $p < 0.01$).
[6] Macey et al. (2025) N = 50	VRChat collaborative puzzle	Custom vs assigned avatar	Six NMSPI social-presence dimensions: no statistically significant differences; high co-presence scores suggest a possible ceiling effect.
[10] Brag et al. (2026) N = 36	Dyadic communication	Live video vs 2D/3D avatar; display modality	Avatar conditions lower than video on CP, affective understanding, and behavioral interdependence; attentional allocation did not differ; HMD added no measurable benefit.
[9] Wang et al. (2025) Two studies: N = 36 + 36	Table-tennis and badminton collaboration	Body / upper-body / head-and-hands visibility × 1PP / 3PP	SP generally higher in 3PP; limited body visibility reduced SP in the badminton study. Findings varied by task.
[11] Lee et al. (2023) 17 dyads recruited (N = 34); 16 analyzed (N = 32)	Dyadic card-matching task	Partner hands visible vs hidden × reward condition	Total SPQ: higher with visible hands ($F(1,31) = 17.78$, $p < 0.001$). Reward: no significant main effect on overall SP.
[7] Wang et al. (2025), VR Whispering	Private conversation in Social VR	Bundled visual / spatial-audio / voice / airflow cues	CP, attention and emotional-contagion measures: higher in the combined

© 2026 The Author(s). Authors retain copyright in their articles.

Study / sample	Interaction context	Representation comparison	SP / CP measures and findings
12 dyads; N = 24			intervention. Individual cue effects cannot be separated.
[5] Kim et al. (2023) N = 48	Social VR interaction	Self vs non-self appearance, voice and name	SP: higher for self-congruent appearance, voice and name; voice showed the strongest effect. Appearance × voice interaction reported.
[12] Kimmel et al. (2023) N = 48	Collaborative VR tasks	Neutral vs eye vs mouth vs eye-and-mouth animation	CP: mouth > neutral; the facial-cue effect varied with task.
[13] Fidalgo et al. (2023) N = 12	Remote pointing collaboration	MAGIC gesture and viewpoint transformation	CP and pointing agreement: higher with MAGIC; AA and PMU: no statistically significant differences.
[14] Sasaki & Igarashi (2025) 14 dyads; N = 28	Dyadic haptic collaboration	Self-avatar and partner-avatar visibility (ALL / ALONE / OTHER / NONE)	SP and CP: higher when the partner avatar was visible; no corresponding clear performance effect.
[15] Adkins et al. (2024) N = 120	Collaborative VR	Hand tracking vs controllers × audio present vs absent	SP: higher with audio; no significant effect of input modality or input × audio interaction.
[16] Venkatraj et al. (2024) 20 dyads (N = 40)	Shared-avatar co-embodiment	Haptics × control allocation × task	CP varied mainly with task freedom. Changes in agency and CP did not follow the same pattern.

Notes: SP = social presence; CP = co-presence (a distinct social-presence dimension, not equivalent to overall SP); SPQ = Social Presence Questionnaire; NMSPI = Networked Minds Social Presence Inventory; HMD = head-mounted display; 1PP/3PP = first-/third-person perspective; AA = attentional allocation; PMU = perceived message understanding. N refers to participants unless otherwise stated. "Not significant" reports a statistical result, not proof of no effect.

Table 1B: Characteristics and Main Findings of Extended Included Reports

Study / sample	Interaction context	Representation comparison	SP / CP measures and findings
[4] Kimmel et al. (2024) N = 18	Controlled collaborative VR experiment	No extra motion capture vs hand tracking vs full-body MoCap	Three NMSPI dimensions (CP, affective understanding and interdependence): no significant differences; interpersonal distance changed.
[17] Hu et al. (2025) N = 40	Multi-user VR cinema	Movement realism × voice realism	SP: voice realism significant ($F = 7.574$, $p = 0.009$); movement realism and interaction not significant.
[18] Zhang & Steed (2025) N = 40 analyzed (42 recruited)	Multiple-avatar encounters	Real-time actor control vs prerecorded responses	CP: higher with live control ($p = 0.027$).
[19] Immohr et al. (2024) 16 dyads; N = 32	Dyadic VR conversation and spatial task	Binaural spatial vs diotic audio × scene arrangement	SP: no statistically significant differences across auralization conditions; direct preference for binaural sound did not yield a corresponding SP effect.
[20] Yoon et al. (2023) 27 dyads; N = 54	Asymmetric AR–VR collaboration	Opaque vs semitransparent vs near-transparent avatar	CP: near-transparent avatar < opaque or semitransparent avatar; effects were not uniform across other social-experience outcomes.
[21] Yu et al. (2021) N = 24	3D asymmetric telepresence	Point-cloud user representation vs personalized 3D virtual-character avatar	SP and CP: higher with point-cloud representation than with virtual-character avatars.
[22] Hart et al. (2021)	Desktop–VR communication	Expression × representation × placement	SP-related outcomes: no significant main effects; perceived expression–voice fit

© 2026 The Author(s). Authors retain copyright in their articles.

Study / sample	Interaction context	Representation comparison	SP / CP measures and findings
12 dyads (N = 24)			improved when facial expressions were enabled.
[23] Nam et al. (2025) N = 30	Asynchronous virtual meetings	Avagent vs recording vs voice-based AI	SP, belonging and emotional intimacy: higher in the Avagent condition; multiple features changed together.
[24] Wang et al. (2024) N = 128	Asynchronous Social VR	Spatial accommodation and added gaze / proximity cues	SP: higher with spatial accommodation plus gaze; spatial accommodation alone was not significant. Gaze was not isolated as an independent effect.

Notes: SP = social presence; CP = co-presence (reported separately from overall SP); NMSPI = Networked Minds Social Presence Inventory; MoCap = motion capture; AR = augmented reality; VR = virtual reality. Sample sizes are from the included reports; no independent experiment has been counted as a separate report.

3. Results

Tables 1A and 1B summarize the 21 included reports. The findings are grouped by the representation feature under examination: appearance, behavior, self-identification, and multimodal cues. Some reports contribute to more than one theme, but each is counted only once in the review. Comparisons below distinguish overall social presence from co-presence and other measured subdimensions.

3.1 Appearance Representation

Appearance studies varied in what was shown: whether a partner was visible at all, how much of the body could be seen, and how that body was rendered. These are different manipulations rather than successive degrees of visual realism.

3.1.1 Avatar Visibility and Visual Representation

Studies that changed avatar visibility found selective benefits for perceiving the other person. Aseeri and Interrante [3] compared no avatar with scanned and real-body representations. Co-presence was higher in both visible-avatar conditions ($\chi^2(2) = 17.20$, $p < 0.01$), although overall social presence did not differ significantly. Other outcomes were selective: the real avatar increased trustworthiness and attention to facial expressions relative to no avatar, while the scanned avatar increased attention to body posture; mutual-gaze differences depended on task. In a separate comparison, Sasaki and Igarashi [14] found higher social presence and co-presence when the partner avatar was visible; task performance did not show a corresponding improvement. Together, these results concern the presence of socially informative visual cues, not the incremental benefit of photorealistic detail.

The amount of visible body mattered differently across tasks. Wang et al. [9] studied full-body, upper-body, and head-and-hands representations together with first- and third-person perspectives in two separately recruited experiments. Reduced visibility was less consequential in the table-tennis task than in the spatially more demanding badminton task, where visibility conditions produced clearer differences in social presence and performance. In asymmetric AR/VR collaboration, Yoon et al. [20] found that increased avatar transparency could lower some social-experience ratings. Neither experiment supports treating maximum visibility as the best setting for every interaction.

Changes in rendering format produced another set of contrasts. Yu et al. [21] reported higher social presence and co-presence for point-cloud avatars than for virtual-character representations in asymmetric telepresence. Brag et al. [10], comparing avatar communication with live video across display conditions, reported lower co-presence and several related social-presence outcomes in the photorealistic-avatar condition. Because these studies differed in communication baseline, display arrangement, and social task, their results cannot be placed on a common realism ranking.

3.1.2 Facial Detail and Emotional Readability

Facial animation was associated with particular social-presence outcomes rather than uniform improvements. In Kimmel et al. [12], mouth movement increased co-presence relative to a neutral expression,

© 2026 The Author(s). Authors retain copyright in their articles.

while eye movement had the highest co-presence under the verbal-explanation task; the pattern depended on task context. Hart et al. [22] found that enabling facial expressions improved participants' judgments of how well the avatar matched the remote speaker's voice and tone, but did not produce significant changes in the reported affective-understanding or emotional-interdependence measures. The distinction matters: perceived communicative fit is not interchangeable with a change in social presence.

3.2 Behavioral Interaction

Behavioral studies examined the availability and responsiveness of gestures, gaze, body movement, and voice. Effects differed according to the manipulated channel and the social-presence measure used.

3.2.1 Fidelity of Nonverbal Social Cues

Two collaboration experiments showed benefits from making action-related cues available. Lee et al. [11] found higher social presence when a partner's hands were visible during a card-matching task ($F(1,31) = 17.78$, $p < 0.001$); the reward manipulation did not significantly affect the overall score. Fidalgo et al. [13] used gesture and viewpoint remapping to improve remote pointing. Pointing agreement and co-presence increased, but attentional allocation and perceived message understanding did not differ significantly. Thus, an improvement in task-relevant action information did not necessarily extend to every measured social-presence dimension.

Responsiveness and gaze were tested under less conventional interaction formats. Zhang and Steed [18] reported higher co-presence for actor-controlled, real-time virtual characters than for prerecorded responses. In asynchronous Social VR, Wang et al. [24] reported higher social presence for a condition combining spatial accommodation with added gaze, whereas spatial accommodation alone was not significant. Because gaze was examined within that combined configuration, its independent effect cannot be inferred from the reported contrast. These studies compare response arrangements, not isolated network-delay manipulations.

3.2.2 Motion Tracking Dimensions and Fidelity

More extensive capture of body motion did not consistently change questionnaire-based social presence in the included comparisons. Kimmel et al. [4] tested conditions with no additional motion capture, hand tracking, and full-body motion capture combined with hand tracking. None of the three reported NMSPI dimensions differed significantly, although interpersonal distance did. Adkins et al. [15] likewise found no significant social-presence difference between hand tracking and controllers. These studies address different aspects of movement representation: the extent of visible tracking in one case and the input device in the other.

Hu et al. [17] directly contrasted movement and voice realism in a multi-user VR cinema. Movement realism was not significant for social presence, whereas voice realism was ($F = 7.574$, $p = 0.009$), with no significant movement-by-voice interaction. The three studies therefore cannot establish equivalence among tracking conditions, but they do show that adding or refining body movement did not yield a detectable social-presence advantage in the measures and settings examined.

3.2.3 Latency, Jitter, and Synchrony

No included report directly manipulated network latency, avatar-motion delay, jitter, or audiovisual desynchronization while measuring interpersonal social presence or co-presence. The live-versus-prerecorded comparison [18] and asynchronous interaction study [24] vary the format of a social response, not a controlled delay interval. Neither provides a direct estimate of the effect of temporal misalignment.

3.2.4 Spatial Audio and Directional Perception

Auditory availability, voice realism, and sound spatialization were not interchangeable interventions. Adkins et al. [15] found higher social presence with audio present than absent, but did not compare spatial with non-spatial sound. Immohr et al. [19] made that more specific comparison, testing binaural auralization against diotic audio, and found no consistent social-presence advantage for spatialization. The positive findings from the multisensory VR Whispering system [7] cannot resolve the spatial-audio question because sound was changed alongside visual and tactile cues.

© 2026 The Author(s). Authors retain copyright in their articles.

3.3 Self-Identification

Self-related manipulations addressed control over an avatar, the choice of its appearance, and its correspondence with the user's identity. These conditions measured different aspects of self-representation.

3.3.1 Embodiment and Agency

Agency did not track social presence in a uniform way. Lee et al. [11] observed higher joint agency and social presence when the partner's hands were visible, while associations between these constructs did not establish a causal pathway. In the shared-avatar study by Venkatraj et al. [16], control allocation and haptics affected agency and perceived synchrony, but co-presence varied mainly with the degree of task freedom. The two outcomes could therefore change independently within an experimental design.

3.3.2 Avatar Customization

Macey et al. [6] compared customized and assigned avatars during dyadic collaboration and found no significant differences across the six measured social-presence dimensions. Co-presence was high in both conditions, which may have constrained sensitivity to small differences. The study did not measure psychological ownership, so the null customization finding cannot be interpreted as a test of whether feeling ownership of an avatar affects social presence.

3.3.3 Self–Avatar Congruence

Identity congruence produced a different result. Kim et al. [5] compared self-congruent and non-self-congruent appearance, voice, and name conditions. Social presence was higher with self-appearance, self-voice, and legal-name cues; voice produced the largest reported effect, and appearance interacted with voice. This was a comparison of identity correspondence, not of whether participants had been allowed to customize an avatar. The two manipulations should therefore not be combined into a single result about personalization.

3.4 Multimodal Representation

Multimodal studies combined or crossed several representation channels. Their experimental contrasts differ in an important respect: some allow individual channels to be compared, whereas others test an entire interaction system.

3.4.1 Combined Visual, Auditory, and Movement Cues

VR Whispering [7] integrated avatar proximity and gestures with voice processing, spatial sound, lighting, and tactile airflow. Compared with its control condition, the complete intervention increased co-presence and several other social-presence-related outcomes. The contribution of any one component was not isolated. Nam et al. [23] likewise reported higher social presence, belonging, and emotional intimacy for an AI-powered embodied avatar than for recordings or voice-based AI in asynchronous meetings. That contrast evaluated the Avagent arrangement as a whole, including its generated responses and audiovisual embodiment.

Factorial comparisons provide narrower evidence about how channels combine. Kim et al. [5] observed an interaction between self-appearance and self-voice, in addition to effects of the identity-congruence factors. Hu et al. [17], however, found a significant effect of voice realism without a significant movement-by-voice interaction. The interaction results concern different variables and settings; they do not establish a general benefit from combining more modalities.

3.4.2 Cross-Modal Consistency and Unexamined Conflicts

None of the included reports experimentally introduced a conflict such as delayed lip movement relative to speech or a sound source displaced from the visible speaker. Hart et al. [22] measured whether facial expressions seemed compatible with the remote speaker's voice and tone, but did not manipulate audiovisual contradiction. Likewise, interactions between appearance and voice congruence [5] or movement and voice realism [17] should not be treated as tests of cross-modal conflict. The included evidence therefore does not determine how deliberately inconsistent cues affect interpersonal social presence.

© 2026 The Author(s). Authors retain copyright in their articles.

4. Discussion

4.1 Technical Fidelity and Social Cue Effectiveness

A more precise interpretation of avatar fidelity requires separating the information reproduced by a system from the information that an interaction partner can actually use. High-resolution appearance and extensive tracking can improve the technical representation of a person, but neither property ensures that the resulting cues will be meaningful in a particular social task. The studies reviewed here offer comparisons relevant to this distinction; they do not directly test it as a causal mechanism.

Visibility studies show what can happen when an otherwise missing social cue becomes available. Aseeri and Interrante [3] observed higher co-presence with scanned and real-body avatars than with no avatar, even though overall social presence did not change significantly. Sasaki and Igarashi [14] found that seeing a partner avatar mattered more for social presence than seeing one's own avatar. Lee et al. [11] reported higher social presence when a collaborator's hands were visible during card matching. These experiments differ in design, but each changes the availability of information about another person's presence or actions; none establishes that increasingly detailed rendering, by itself, would produce the same benefit.

More elaborate motion representation yielded a different pattern. Full-body motion capture did not improve the three social-presence subdimensions assessed by Kimmel et al. [4], although interpersonal distance varied. Adkins et al. [15] found no significant difference between controllers and hand tracking in social presence, while the presence of audio did make a difference. In the VR cinema examined by Hu et al. [17], voice realism showed a significant effect, unlike movement realism. These are not interchangeable tests of 'fidelity': they alter different channels, in different settings, and sometimes compare an input device rather than the visible quality of an avatar's movement.

The findings for visual realism also resist a single ranking. Yu et al. [21] reported higher social presence and co-presence with point-cloud avatars than with virtual-character representations in asymmetric telepresence. Brag et al. [10], using a different communication baseline, found lower scores on several social-presence dimensions for photorealistic avatar communication than for live video. The comparisons establish effects within their respective systems, not a general ordering of point clouds, realistic avatars, and video. For HCI design, the defensible question is which social information each representation exposes or obscures under its actual interaction conditions.

4.2 Interaction Contexts and Measurement Differences

The relationship between a representation choice and social presence becomes clearer when task demands are examined alongside the manipulation. Wang et al. [9] reported two independent experiments, one using table tennis and the other badminton, with different visibility and viewpoint conditions. Reduced body visibility was more consequential in the task with greater spatial demands. This within-publication contrast is stronger evidence of task dependence than a comparison between unrelated studies, although it should not be turned into a universal threshold for how much of the body an avatar must display.

Facial and gestural findings are similarly specific. In Kimmel et al. [12], mouth movement increased co-presence relative to a neutral face, while the role of eye movement differed with the communication task. Fidalgo et al. [13] reported better pointing agreement and higher co-presence with gesture and viewpoint remapping, without corresponding significant changes in attentional allocation or perceived message understanding. In these cases, the relevant improvement concerned a particular communicative activity or social-presence dimension, not every aspect of the interaction.

Response mode and partner type introduce further differences, but they should not be mistaken for conventional tracking fidelity. Zhang and Steed [18] found higher co-presence when virtual characters were controlled live by an actor rather than played back from recordings. Wang et al. [24] reported a social-presence benefit for an asynchronous interaction condition combining spatial accommodation with added gaze; spatial accommodation alone was not significant. The latter result does not isolate gaze as an independent cause. Nam et al. [23] compared an AI-powered embodied avatar with recordings and voice-based AI in asynchronous meetings. Its advantage concerns the whole Avagent interaction condition, including the form of response,

© 2026 The Author(s). Authors retain copyright in their articles.

rather than an independently manipulated facial or body cue. Neither the live-versus-recorded comparison nor the asynchronous studies directly measures the effect of network latency.

Interpretation also depends on what the questionnaire asks. The null overall social-presence result in Aseeri and Interrante [3] coexisted with a significant co-presence difference. In Hart et al. [22], more appropriate-looking facial expressions improved participants' judgments about communication fit without significant changes in the reported affective dimensions. Macey et al. [6] found no significant difference between customized and assigned avatars across six social-presence dimensions, with co-presence high in both conditions. A non-significant overall score, a significant subdimension, and a communication judgment are three different findings; combining them under a single 'positive' or 'negative' label would discard the distinction the measures were designed to capture.

The apparent contrast between customization and self-similarity illustrates the same point at the level of the independent variable. Kim et al. [5] manipulated correspondence between users and their avatars in appearance, voice, and name; social presence was higher in the self-congruent conditions, with the strongest effect associated with voice and an appearance-by-voice interaction. Macey et al. [6], however, tested whether the avatar was customized or assigned. Choosing an avatar is not the same manipulation as making it resemble the user. Neither comparison, on its own, demonstrates an effect of psychological ownership, which was not directly measured in the customization study.

4.3 Evidence Gaps and Research Priorities

The most useful next experiments would isolate relationships that current system-level comparisons leave unresolved. None of the included reports independently varied network latency, motion jitter, or audiovisual timing while directly measuring interpersonal social presence. A controlled design could change avatar response delay while holding appearance, task, and partner behavior constant, then examine whether co-presence and other social-presence dimensions respond differently. This would be more informative than treating live-versus-recorded interaction as a substitute for a latency experiment.

Cross-modal consistency poses a separate, testable problem. Existing comparisons of appearance and voice congruence [5], or movement and voice realism [17], examine combinations of properties, not necessarily contradictions between cues. Experiments that deliberately offset lip movement from speech or displace the apparent sound source from the speaking avatar could identify whether such mismatches affect interpersonal judgments. The delay or spatial offset would need to be manipulated independently of the realism of each channel.

A related design question is whether cues can compensate for a channel that is unavailable. Audio presence was influential in Adkins et al. [15], and voice realism mattered in Hu et al. [17], whereas spatial rather than diotic audio showed no consistent social-presence advantage in Immohr et al. [19]. VR Whispering [7] combined sound with visual and tactile components and therefore cannot settle this question. Factorial comparisons of, for example, hand visibility and voice cues could test substitution or interaction directly. The purpose would not be to identify a universally dominant modality, but to determine which information remains useful when another channel is restricted.

Longer and less scripted encounters deserve attention for a different reason: most included experiments used short tasks with defined goals. The current sample provides limited evidence about sustained group interaction, changing familiarity between partners, and repeated use of the same avatar. Studies of these settings could also measure self-similarity, customization, and psychological ownership separately, rather than assuming that one is a proxy for another.

4.4 Methodological Limitations of the Included Evidence

Statistical results from the included experiments must be read in light of their designs and sample sizes. Fidalgo et al. [13] studied 12 participants, Kimmel et al. [4] studied 18, and Adkins et al. [15] studied 120. These figures do not establish a quality ranking, but they make it especially inappropriate to interpret every non-significant comparison as evidence of no effect. Effect estimates, uncertainty intervals, and replication would help distinguish small effects from designs that had limited ability to detect them.

© 2026 The Author(s). Authors retain copyright in their articles.

Measurement practices further restrict direct comparison. Macey et al. [6] reported all six NMSPI subdimensions, whereas Kimmel et al. [4] analyzed three. Both studies also reported high co-presence ratings, leaving less room for differences to appear on that scale. Behavioral observations complicate rather than invalidate the questionnaire findings: interpersonal distance changed across motion conditions in Kimmel et al. [4] despite non-significant changes in the reported social-presence subdimensions; Fidalgo et al. [13] recorded improvements in pointing agreement alongside a more selective pattern across questionnaire dimensions. Behavioral performance and perceived social presence should be retained as separate outcomes.

Finally, the level at which an intervention is manipulated limits the level at which it can be explained. VR Whispering [7] changed several visual, auditory, and tactile cues together. The Avagent condition in Nam et al. [23] likewise bundled generated responses with an embodied audiovisual presentation. Those studies evaluate complete interaction arrangements, which is useful for system design, but do not identify the individual contribution of each component. Treating the package as the unit of inference avoids claiming an effect that the experimental contrast never isolated.

4.5 Limitations of the Present Review

This review is bounded by its search and screening process. IEEE Xplore and the ACM Digital Library were the only formal search databases, with searches completed on 20 and 28 August 2026, respectively, for the specified 2021–2026 publication period. Research indexed primarily in psychology or communication databases may therefore be absent. Screening and eligibility assessment were carried out by one reviewer, which leaves room for selection and classification error. Background sources cited to establish theory or context were not counted as part of the 21-report synthesis.

The final sample contains 12 core and nine extended reports. This distinction makes the difference between synchronous human-to-human Social VR and adjacent contexts visible, but it does not make those contexts experimentally equivalent. The unit counted is a report: Wang et al. [9], for example, contains two separately recruited experiments and still contributes one report to the total. A report may inform more than one thematic category without becoming an additional item of evidence. Because manipulations, participants, and outcome measures vary substantially, the present synthesis compares specific patterns and their limits rather than calculating a pooled effect size.

5. Conclusion and Future Directions

Across the 21 research reports included in this review, there was no consistent benefit from making avatars more visually or behaviorally elaborate. Some comparisons showed higher social presence or co-presence when partner bodies, hands, or identity-related cues were available; others found no statistically significant differences when motion-tracking coverage or movement realism increased. The studies tested different interventions and outcomes, so these findings do not establish that cue relevance is more important than technical fidelity. They do challenge the assumption that greater representational detail is a reliable proxy for stronger social presence.

What participants were doing, and what researchers measured, changed how these findings could be interpreted. A difference in co-presence did not necessarily appear in an overall social-presence score or in other subdimensions. The two collaborative tasks included in this review also showed different consequences of reduced avatar visibility. Self-similarity and customization addressed distinct design decisions: one concerns an avatar's correspondence to its user, while the other concerns how that avatar is chosen. Results from AI-mediated, asynchronous, and asymmetric XR contexts extend the scope of the evidence, but they should not be read as direct replications of synchronous human-to-human Social VR.

For HCI design, the practical question is what an avatar needs to communicate in a given setting. Visible hands may help partners coordinate a shared manipulation, while facial behavior or voice may be more pertinent to a conversational exchange. A more detailed representation is not necessarily more informative for either task. This is a task-sensitive design implication supported by comparisons across the included reports, not a universal ranking of avatar features.

Future experiments should test that implication by separating cue availability from rendering or tracking fidelity and varying the demands of the interaction. Factorial designs could also isolate components that

© 2026 The Author(s). Authors retain copyright in their articles.

existing multisensory and AI-mediated systems manipulate together. Direct tests of avatar response delays and audiovisual mismatches remain absent from the included evidence. Comparisons would be more informative if they reported the social-presence instrument and its subdimensions, effect estimates, and relevant behavioral outcomes rather than significance alone. Finally, repeated encounters and sustained group interaction would help establish whether patterns observed in short, structured tasks persist in everyday Social VR use.

References

[1] OH C S, BAILENSON J N, WELCH G F. A systematic review of social presence: definition, antecedents, and implications[J]. Frontiers in Robotics and AI, 2018, 5: 114. DOI: 10.3389/frobt.2018.00114.

[2] BIOCCA F, HARMS C, BURGOON J K. Toward a more robust theory and measure of social presence: review and suggested criteria[J]. Presence: Teleoperators and Virtual Environments, 2003, 12(5): 456–480. DOI: 10.1162/105474603322761270.

[3] ASEERI S, INTERRANTE V. The influence of avatar representation on interpersonal communication in virtual social environments[J]. IEEE Transactions on Visualization and Computer Graphics, 2021, 27(5): 2608–2617. DOI: 10.1109/TVCG.2021.3067783.

[4] KIMMEL S, LANDWEHR E, HEUTEN W. Kinetic connections: exploring the impact of realistic body movements on social presence in collaborative virtual reality[J]. Proceedings of the ACM on Human-Computer Interaction, 2024, 8(CSCW2): Article 371, 30 pages. DOI: 10.1145/3686910.

[5] KIM H, PARK J, LEE I-K. “To be or not to be me?”: exploration of self-similar effects of avatars on social virtual reality experiences[J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 29(11): 4794–4804. DOI: 10.1109/TVCG.2023.3320240.

[6] MACEY A-L, BUJIĆ M, BELOUSOV A, KEROUS B, JÄRVELÄ S, HAMARI J. Are we in this together? Investigating the relationship of avatar customisation, social presence, and communication outcomes in collaborative embodied virtual reality[J]. Proceedings of the ACM on Human-Computer Interaction, 2025, 9(7): Article CSCW371, 20 pages. DOI: 10.1145/3757552.

[7] WANG X, PENG K, HAO C, YU W, YI X, LI H. VR whispering: a multisensory approach for private conversations in social virtual reality[J]. IEEE Transactions on Visualization and Computer Graphics, 2025, 31(5): 3459–3469. DOI: 10.1109/TVCG.2025.3549579.

[8] PAGE M J, MCKENZIE J E, BOSSUYT P M, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews[J]. BMJ, 2021, 372: n71. DOI: 10.1136/bmj.n71.

[9] WANG Z, LI X, HUANG J, JIN Y, CHEN Y, ZHANG D. Effects of avatar visibility and perspective on social presence and performance in dynamic VR collaboration tasks[J]. IEEE Transactions on Visualization and Computer Graphics, 2025, 31(10): 9179–9194. DOI: 10.1109/TVCG.2025.3591830.

[10] BRAG A B, ANKEL L, GEORGE C, KOJIC T, MÖLLER S, SPANG R P. The QoE of photorealistic avatar communication: a benchmark against video conferencing[C]//2026 18th International Conference on Quality of Multimedia Experience (QoMEX), 2026: 1–7. DOI: 10.1109/QoMEX69967.2026.11618367.

[11] LEE S U, KIM J, LEE J. Effects of reward schedule and avatar visibility on joint agency during VR collaboration[J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 29(11): 4372–4382. DOI: 10.1109/TVCG.2023.3320221.

[12] KIMMEL S, JUNG F, MATVIIENKO A, HEUTEN W, BOLL S. Let’s face it: influence of facial expressions on social presence in collaborative virtual reality[C]//Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 2023: Article 429, 16 pages. DOI: 10.1145/3544548.3580707.

[13] FIDALGO C G, SOUSA M, MENDES D, DOS ANJOS R K, MEDEIROS D, SINGH K, JORGE J. MAGIC: manipulating avatars and gestures to improve remote collaboration[C]//2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR), 2023: 438–448. DOI: 10.1109/VR55154.2023.00059.

© 2026 The Author(s). Authors retain copyright in their articles.

[14] SASAKI G, IGARASHI H. Exploring remote collaborative tasks: the impact of avatar representation on dyadic haptic interactions in shared virtual environments[J]. IEEE Transactions on Visualization and Computer Graphics, 2025, 31(10): 8846–8858. DOI: 10.1109/TVCG.2025.3580546.

[15] ADKINS A, CANALES R, JÖRG S. Hands or controllers? How input devices and audio impact collaborative virtual reality[C]//30th ACM Symposium on Virtual Reality Software and Technology, 2024: Article 31, 12 pages. DOI: 10.1145/3641825.3687718.

[16] VENKATRAJ K P, MEIJER W, PERUSQUÍA-HERNÁNDEZ M, HUISMAN G, EL ALI A. ShareYourReality: investigating haptic feedback and agency in virtual avatar co-embodiment[C]//Proceedings of the CHI Conference on Human Factors in Computing Systems, 2024: Article 100, 15 pages. DOI: 10.1145/3613904.3642425.

[17] HU J, DING D, XU X, ZHANG J, JIN J, DONG F. Bilateral virtual companions: the impact of virtual humans' movement and voice realism on user perception and experience in multi-user VR cinemas[C]//2025 28th International Conference on Computer Supported Cooperative Work in Design (CSCWD), 2025: 2635–2641. DOI: 10.1109/CSCWD64889.2025.11033578.

[18] ZHANG J, STEED A. Single actor controlling multiple avatars for social virtual realities[C]//2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR), 2025: 146–155. DOI: 10.1109/VR59515.2025.00039.

[19] IMMOHR F, RENDLE G, LAMMERT A, NEIDHARDT A, MEYER ZUR HEYDE V, FRÖHLICH B, RAAKE A. Evaluating the effect of binaural auralization on audiovisual plausibility and communication behavior in virtual reality[C]//2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR), 2024: 849–858. DOI: 10.1109/VR58804.2024.00104.

[20] YOON B, SHIN J-E, KIM H-I, OH S Y, KIM D, WOO W. Effects of avatar transparency on social presence in task-centric mixed reality remote collaboration[J]. IEEE Transactions on Visualization and Computer Graphics, 2023, 29(11): 4578–4588. DOI: 10.1109/TVCG.2023.3320258.

[21] YU K, GORBACHEV G, ECK U, PANKRATZ F, NAVAB N, ROTH D. Avatars for teleconsultation: effects of avatar embodiment techniques on user perception in 3D asymmetric telepresence[J]. IEEE Transactions on Visualization and Computer Graphics, 2021, 27(11): 4129–4139. DOI: 10.1109/TVCG.2021.3106480.

[22] HART J D, PIUMSOMBOON T, LEE G A, SMITH R T, BILLINGHURST M. Manipulating avatars for enhanced communication in extended reality[C]//2021 IEEE International Conference on Intelligent Reality (ICIR), 2021: 9–16. DOI: 10.1109/ICIR51845.2021.00011.

[23] NAM H, SARVESH M, KANG S, WOO W, KIM K. Effects of AI-powered embodied avatars on communication quality and social connection in asynchronous virtual meetings[J]. IEEE Transactions on Visualization and Computer Graphics, 2025, 31(11): 10152–10162. DOI: 10.1109/TVCG.2025.3616761.

[24] WANG P, MILLER M R, QUEIROZ A C M, BAILENSON J N. Socially late, virtually present: the effects of transforming asynchronous social interactions in virtual reality[C]//Proceedings of the CHI Conference on Human Factors in Computing Systems, 2024: Article 794, 19 pages. DOI: 10.1145/3613904.3642244.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

© 2026 The Author(s). Authors retain copyright in their articles.

Copyrights

Copyright for this article is retained by the author (s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

© 2026 The Author(s). Authors retain copyright in their articles.