Research and Analysis of Key Interaction Design Features of Virtual Space Co-Presence Based on XR Collaboration and Social VR Projects in Recent Five Years

Junlu Shen

Brunel University London School, North China University of Technology, Beijing, China
Shenjl0705@outlook.com

Abstract. In the past five years, the formation of co-presence in virtual space has gradually presented a clear path of interaction design. Based on the review and case comparison of 14 pieces of literature on XR collaboration and social VR from 2021 to 2026, this paper summarizes the relevant designs into five high-frequency features: shared spatial reference and co-existing context, social representation and embodied clues, communication channels and social signals, joint attention and state visibility, and action coupling and cooperation mechanism. The study found that the related programs generally showed structural tension between shared control and individual autonomy, cue saliency and interference burden, realism and mismatch/strangeness. Finally, it is concluded that the improvement of co-presence does not depend on the simple superposition of the number of clues, but on selecting the "minimum effective feature set" around the task goal, and maintaining a stable common reference, clear other perception and coordinated pace of action in different social scales and use situations through adjustable, hierarchical and degradable interaction mechanisms.

Keywords: Co-presence, Social presence, Virtual space, XR collaboration, Interaction design

1. Introduction

With the expansion of XR and social VR applications to remote collaboration, group communication and shared creation, the key to user experience is no longer whether to immerse, but whether to form a stable feeling of "being with others" in the virtual space. Relevant studies usually associate this experience with the concepts of social presence and co-presence. Recent reviews have pointed out that VR related assessments often need to distinguish different sub-concepts of presence, such as social, self and spatial dimensions. Otherwise, it is easy to confuse spatial immersion with social presence, thus affecting the comparability of conclusions [1]. At the same time, research on multi person collaboration and social interaction also shows that avatar presentation, facial and nonverbal cues, and communication organization mechanism will significantly affect users' sense of social presence, conversation satisfaction, and collaboration quality [2,3].

From the point of view of design practice, the reason why co-presence is difficult to obtain stably is often due to the fracture of various clues. In remote collaboration, spatial inconsistency will

destroy the common reference and deixis, so that the collaboration needs additional alignment mechanism to be established [4-6]. In co- located or mixed scenes, viewpoint differences and occlusion will magnify the uncertainty of "what do people mean", thus increasing clarification costs and weakening interaction coherence [7]. In social communication, if there is no appropriate guidance for the round and attention distribution of group dialogue, it is easy to cause difficulty in interrupting and disorder of interaction [2]. At the same time, when key non-verbal cues such as face and emotion are blocked by head-mounted displays, missing in tracking, or inconsistent in presentation, they may reduce social presence and introduce negative experiences such as uncanniness [3,8,9]. These problems jointly indicate that co-presence is not an atmosphere generated automatically, but is supported by a cue system that can be engineered and interactive clue system.

Based on this, this paper focuses on reusable interaction design features, rather than simply comparing experimental paradigms. This paper focuses on the research projects directly related to co-presence in collaboration/social XR from 2021 to 2026, which can provide clear description of interaction mechanism and evidence of effect, covering AR/MR remote collaboration, social VR dialogue, embodied avatar expression and multimodal collaborative feedback. In order to avoid focusing on the discussion of simple experimental methods, this paper emphasizes that the screening criteria is how to achieve interaction by presenting specific design features of works, and reports its impact on collaboration efficiency, joint attention, social presence/co-presence or related experience indicators. By summarizing how the existing systems realize the co-presence related clues as specific interaction mechanisms, and discussing their effects and boundary conditions, a set of transferable design feature framework is extracted to support the design and research of subsequent virtual space collaboration and social systems.

2. Constituent dimensions and design mechanisms of co-presence

Viewed as a whole, co-presence is not triggered by a single enhancement item, but is supported by a group of composable features. On the one hand, stable shared reference is needed to answer " whether people are in the same situation"; on the other hand, people should make the other party like a real other through social representation and social signals; at the same time, it relies on joint attention and state visibility to reduce collaboration uncertainty; when the task enters the joint operation stage, the action coupling mechanism is required to promote "seeing each other" to "doing things with each other synchronously". According to this, the following paper summarizes five dimensions, and refines the function logic and boundary conditions by comparing different implementation paths in each dimension.

2.1. Shared spatial reference and co-existing context

In virtual space, "being together" is often first reflected as an operable spatial consensus, that is, whether both two sides organize their vision, deixis and operation around the same reference system, so as to form a stable co-existing context under the conditions of different places and different devices. The core of shared reference is not a globally consistent environment reproduction, but to transform the task related areas into a common coordinate system and allow the reference to be reconfigured with the task phase. As shown in Figure 1, Fink et al. established a local reference based on the work area, maintained order by entering/leaving and fading in and out, and strengthened the sense of space in the same work area [4]; in contrast, as shown in Figure 2, Grønbæk et al. directly put the co-existing context on the metaphor of face-to-face/side-by-side positioning and whiteboard, and used the mechanism of "returning to natural scale" to repair the

reference drift caused by scaled relocation [5]. Further, Huang et al. used portal as a spatial anchor to establish a shared surface, and allowed switching between "merge co-editing" and "separate parallel", showing that the shared reference needs reconfigurable rather than fixed layout [6]. Therefore, the establishment of stable and switchable shared reference around the task focus is the basis for maintaining the virtual co-existence context.

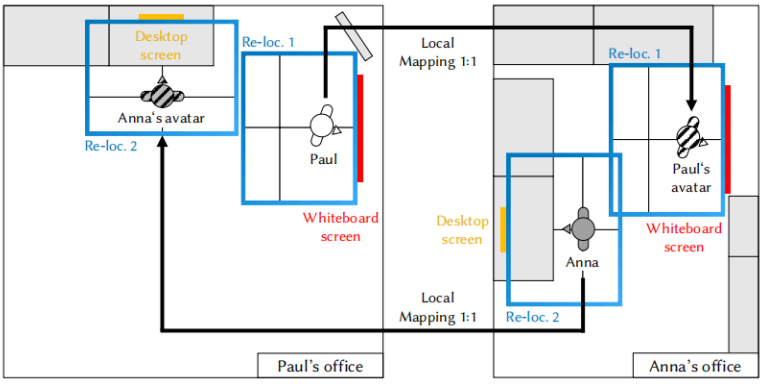

Figure 1. The relocation function deploys the local coordinate system to ensure that all collaboration areas maintain a unified spatial reference system [4]

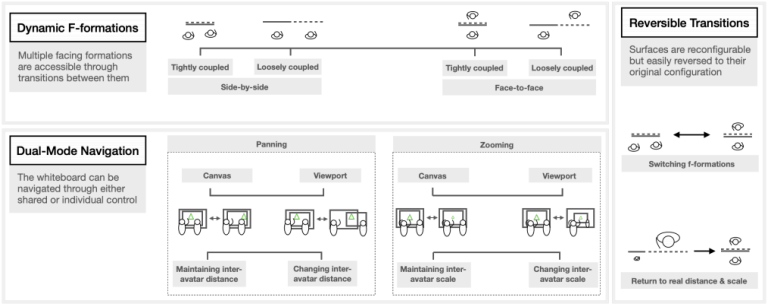

Figure 2. Overview of design principles of mixed whiteboard [5]

2.2. Social representation and embodied clues

Social representation and embodied clues determine "whether the other party is like a person who is present". Its role is not limited to emotional expression itself, but also to provide readable nonverbal signals for interaction, so as to support trust, closeness and tacit cooperation. The key is that these clues must be consistent and credible under the constraints of real tracking ability, rendering resources and privacy. Kimmel et al. showed that facial expression cues (especially eye + mouth combination) can enhance social existence, but also suggested that when the cues are too strong, they may compete with task attention [3]; Kullmann et al. further revealed that the consistency of cues is the key, that is, the change of facial animation coverage and consistency will affect self-recognition and sense of strangeness, which in turn will weaken the sense of reality [8]. Along the idea of minimum effective set, Kang et al. pointed out that the absence of key emotional parameters would have a negative effect, indicating that under constraints, it is preferable to give priority to ensuring the key emotional clues and consistent presentation, rather than pursuing the maximum coverage [9]. Therefore, the embodied presentation strategy with consistency as the priority and key clues as the core is more likely to continuously support the co-presence experience under a variety of devices and individual preferences.

2.3. Communication channels and social signals

Communication channels and social signals affect the co-presence of organization. Who to listen to, when to interrupt, and how to express the tone will directly shape the social order and sense of participation in the virtual space. Effective design usually does not make communication more dazzling, but turns communication into a controllable, understandable and politely organized interactive structure. As shown in Figure 3, Yan et al. made "speak to whom" a directional delivery mechanism to reduce bystanders' interference and improve the sense of control, making the multi person dialogue closer to the local private language structure in reality [10]; complementary to this, Lee et al. facing the turn-taking and attention alignment of group conversation, made the requests of new speakers easier to be captured through the combination of light effects and spatial audio, but overly frequent prompts may cause the opposite effect and interrupt the speaker [2]. On the other hand, as shown in Figure 4, Zhang externalized the emotion and rhythm of speech into playable subtitles to enhance the expression and interactive atmosphere, but also led to typical trade-offs, that is, the richer the expression, the more need to control the visual load and occlusion risk [11]. Therefore, the communication design to enhance co-presence should give consideration to expressiveness and organizability, and provide adjustable signal strength and intervention boundary for different social scales.

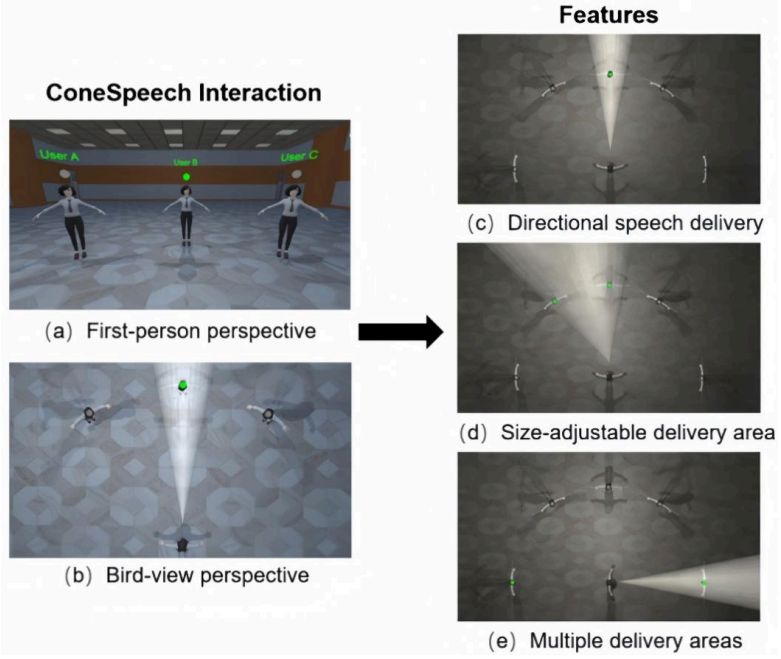

Figure 3. Audible only to users in the conical area directly opposite the speaker (a: First-person perspective, b: Bird-view perspective, c: Directional speech delivery, d: Size-adjustable delivery area, e: Multiple delivery areas) [10]

Figure 4. Interesting visual cue subtitles (A: the system generates dynamic subtitles with both semantic and emotional expressions based on real-time voice; B: Users can interact with subtitles like operating real objects; C: Subtitles will trigger exaggerated dynamic feedback effect after touching others.) [11]

2.4. Joint attention and state visibility

Joint attention and state visibility are used to reduce the uncertainty of "what people are looking at/doing", which is one of the most direct supports of co-presence in virtual collaboration. When the focus and action state are presented timely and appropriately, it is easier for collaboration to form a continuous rhythm and common understanding. It should be emphasized that visibility is not the more the better, but should serve for readability, representability and low interference. Darbar et al. enhanced their common awareness of each other's visual focus with two-way fixation cues, and improved the readability and stability of common attention information through the presentation strategy of "minimum clutter" (such as excluding thermal maps and fixation rays) [12]; in contrast, Wang et al. made the other party's perspective into a controllable parallel window to enhance the efficiency of information understanding and cooperation, but also showed that when sharing was suspended or cut off, the sense of co-presence might decline, suggesting that strong synchronization does not necessarily lead to a stronger presence [13]. Gottsacker et al. used a more conservative perspective alignment strategy to reduce occlusion and anaphora ambiguity, while avoiding the sense of instability caused by continuous updates, and emphasized that visibility should serve readability and anaphora [7]. Therefore, the design for common attention should be based on the principle of selective presentation of key states, which can improve the alignment efficiency and avoid information overload and the destruction of field continuity of presence.

2.5. Action coupling and cooperation mechanism

When interaction moves from "joint sight/joint understanding" to "joint operation", the core of co-presence will further turn to coordination at the action level, which is specifically reflected in whether the user can timely perceive the interaction status of the other party, predict the next step of the other party, and maintain a collaborative rhythm under conflict or occlusion conditions. At this point, mechanism design should not only present information, but also provide feedback structure that can be used to organize actions. As shown in Figure 5, Johnson et al. reduced the cost of expert understanding and instruction landing through the familiar surrounding space + real-time view of novice activity area, making the remote guidance closer to the cooperation rhythm of the same site [14]; Kim et al. decomposed collaborative manipulation into dual consciousness of self-state and others' state, and distinguished the two with multimodal feedback. They can still maintain instant awareness of others' actions under occlusion or high load, but also suggested that there is a comfort threshold for strong feedback, which needs to make a fine balance between perceptible and undisturbed [15]. Therefore, mechanism oriented toward action coupling should highlight the

coordinated state structure, and achieve a balance between usability and comfort through cross modal and strength control.

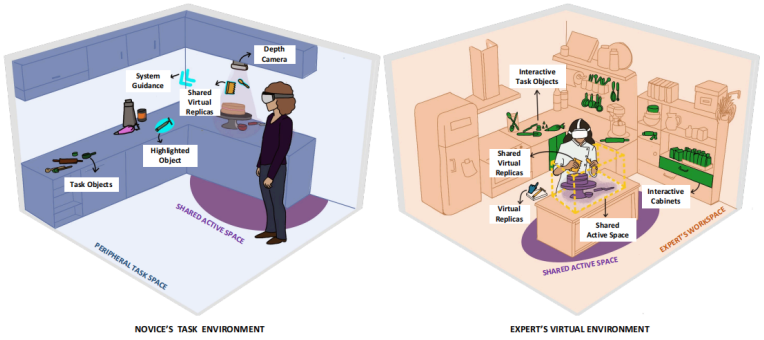

Figure 5. The familiar surrounding space + real-time view of novice activity area [14]

2.6. Comprehensive comparison and induction

To sum up, co-presence in virtual space can be understood as a progressive chain from "situation alignment" to "action synergy". Among them, shared space reference provides a common situation for interaction, social representation and embodied clues make each other a credible social subject, communication channels and social signals establish an organized interactive order, joint attention and state visibility reduce the uncertainty of collaboration, and action coupling and collaboration mechanism further promote the presence experience into a coordinated joint practice. The five dimensions are not independent of each other, but play a role with different weight combinations in different task stages and social scales; therefore, the key of the design is not to superimpose more clues, but to select the "minimum effective feature set" around the task goal, and handle the boundary conditions such as attention load, privacy constraint and comfort threshold through an adjustable and degradable mechanism, so as to maintain a stable and credible co-presence experience.

3. Challenges and prospects

Co-presence in virtual space is essentially an organized collaborative state. The key is not more clues, but whether the shared reference, social clues and communication and coordination mechanism can be properly configured. In other words, the challenge of current design is how to present key clues in a moderate and adjustable way, so as to maintain a coherent experience in complex tasks and multi person situations. At the same time, the existing research still has some limitations. On the one hand, the task types, equipment conditions and evaluation indicators relied on by different research institutes are quite different, leading to the difficulty of horizontal comparison and integration between different conclusions; on the other hand, more clues, higher realism or more complex feedback do not necessarily correspond to a better experience. On the contrary, over design may lead to problems such as increased cognitive load, enhanced information interference or imbalance of interaction rhythm. It can be seen that the design of co-presence in virtual space in the future should not be one-sided pursuit of technology stack, but should pay more attention to the adaptation relationship between different design features and specific use situations.

For the subsequent research and practice, there are four directions worth promoting. First, establish a method of "feature arrangement" around the task stage, so that the system can switch the combination and intensity of clues in different stages such as discussion, execution and replay;

second, take the lack of sensing, privacy preference and inconsistent presentation as the norm, and develop interpretable and controllable degradation strategies to avoid the experience backfire caused by being half real and half false; third, for multi person large-scale scenes, it systematically solves the visual clutter and attention load caused by the explosion of clues, so that co-presence can still remain readable and negotiable when it expands to a larger groups; fourth, the habit formation and social norms in long-term use are included in the assessment to test the continuous impact of co-presence mechanisms in the real cooperative relationship and team division of labor.

4. Conclusion

Focusing on the key interaction design features of co-presence in virtual space, this paper reviews and analyzes the research on XR collaboration and social VR in recent years, and summarizes the main factors affecting the generation and maintenance of co-presence experience from the aspects of shared space reference and co-existing context, social representation and embodied clues, communication channels and social signals, joint attention and state visibility, action coupling and cooperation mechanism. On the whole, co-presence is not a natural incidental experience in the virtual environment, nor can it be achieved by relying on a single technology or single function, but the result of multiple interaction cues coordinating with each other and jointly taking effect in specific tasks and social situations. Only when the system can provide users with stable common reference, clear other information, coherent communication order and coordinated cooperation mechanism, users are more likely to form a strong subjective experience of "being together with others".

The main challenge of current research is that there is still a lack of unified standards among different task scenarios, equipment conditions and evaluation dimensions, and there is always a trade-off between clue enhancement and cognitive load, authenticity and comfort. In the future, it is still necessary to further explore the combination mode and function boundary of key design features under different social scales, task stages and media conditions on the basis of the same concept definition and evaluation framework, so as to promote the formation of a more universal and transferable design framework.

References

[1] Bareišytė, L., Slatman, S., Austin, J., Rosema, M., van Sintemaartensdijk, I., Watson, S., & Bode, C. (2024). Questionnaires for evaluating virtual reality: A systematic scoping review. Computers in Human Behavior Reports, 16, 100505.

[2] Lee, G., Lee, D. Y., Su, G. M., & Manocha, D. (2024). "may i speak?": Multi-modal attention guidance in social vr group conversations. IEEE Transactions on Visualization and Computer Graphics, 30(5), 2287-2297.

[3] Kimmel, S., Jung, F., Matviienko, A., Heuten, W., & Boll, S. (2023, April). Let's face it: Influence of facial expressions on social presence in collaborative virtual reality. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-16).

[4] Fink, D. I., Zagermann, J., Reiterer, H., & Jetter, H. C. (2022). Re-locations: Augmenting personal and shared workspaces to support remote collaboration in incongruent spaces. Proceedings of the ACM on Human-Computer Interaction, 6(ISS), 1-30.

[5] Grønbæk, J. E. S., Sánchez Esquivel, J., Leiva, G., Velloso, E., Gellersen, H., & Pfeuffer, K. (2024, May). Blended whiteboard: Physicality and reconfigurability in remote mixed reality collaboration. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1-16).

[6] Huang, X., & Xiao, R. (2024). Surfshare: Lightweight spatially consistent physical surface and virtual replica sharing with head-mounted mixed-reality. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 7(4), 1-24.

[7] Gottsacker, M., Numan, N., Steed, A., Bruder, G., Welch, G. F., & Feiner, S. (2025, April). Decoupled hands: An approach for aligning perspectives in collaborative mixed reality. In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (pp. 1-7).
[8] Kullmann, P., Schell, T., Menzel, T., Botsch, M., & Latoschik, M. E. (2025). Coverage of facial expressions and its effects on avatar embodiment, self-identification, and uncanniness. IEEE Transactions on Visualization and Computer Graphics.
[9] Kang, S., Song, H., Yoon, B., Kim, K., & Woo, W. (2023, October). Effects of different facial blendshape combinations on social presence for avatar-mediated mixed reality remote communication. In 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) (pp. 439-440). IEEE.
[10] Yan, Y., Liu, H., Shi, Y., Wang, J., Guo, R., Li, Z., ... & Shi, Y. (2023). Conespeech: Exploring directional speech interaction for multi-person remote communication in virtual reality. IEEE Transactions on Visualization and Computer Graphics, 29(5), 2647-2657.
[11] Zhang, Y., Wen, Y., Hu, S., & Lu, Z. (2025). SpeechCap: Leveraging Playful Impact Captions to Facilitate Interpersonal Communication in Social Virtual Reality. Proceedings of the ACM on Human-Computer Interaction, 9(7), 1-33.
[12] Darbar, R., Santuz, H., Taly, A., & Baaden, M. (2024, December). GazeMolVR: Sharing eye-gaze cues in a collaborative VR environment for molecular visualization. In Proceedings of the International Conference on Mobile and Ubiquitous Multimedia (pp. 7-23).
[13] Wang, X., Shen, L., Chen, L., Fan, M., & Lee, L. H. (2025). Teamportal: Exploring virtual reality collaboration through shared and manipulating parallel views. IEEE Transactions on Visualization and Computer Graphics.
[14] Johnson, J. G., Sharkey, T., Butarbutar, I. C., Xiong, D., Huang, R., Sy, L., & Weibel, N. (2023, April). Unmapped: Leveraging experts' situated experiences to ease remote guidance in collaborative mixed reality. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-20).
[15] Kim, H., & Park, J. M. (2024). Providing dual awareness using multimodal cues for collaborative manipulation in virtual environments. International Journal of Human–Computer Interaction, 40(12), 3298-3312.