



## OPEN ACCESS

## EDITED BY

Giovanni Vecchiato,  
National Research Council of Italy, Italy

## REVIEWED BY

Ryoichi Nakashima,  
Kyoto University, Japan  
Ruggero Eugeni,  
Catholic University of the Sacred Heart, Italy  
Yuer Yang,  
The University of Hong Kong, Hong Kong SAR,  
China

## \*CORRESPONDENCE

Yuki Harada,  
✉ haradayuuki00@gmail.com

RECEIVED 20 January 2025

ACCEPTED 05 May 2025

PUBLISHED 16 May 2025

CORRECTED 07 July 2025

## CITATION

Harada Y, Sato Y, Kambara A, Oki K, Takahashi R,  
Utsumi N, Tsuji R and Sato Y (2025)  
Psychological evaluation of a virtual connected  
room from the perspective of we-mode.  
*Front. Virtual Real.* 6:1563866.  
doi: 10.3389/frvir.2025.1563866

## COPYRIGHT

© 2025 Harada, Sato, Kambara, Oki, Takahashi,  
Utsumi, Tsuji and Sato. 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.

# Psychological evaluation of a virtual connected room from the perspective of we-mode

Yuki Harada <sup>1,2\*</sup>, Yoshihiro Sato <sup>3</sup>, Ayumi Kambara <sup>1</sup>, Kazuo Oki <sup>3</sup>,  
Ryo Takahashi <sup>3</sup>, Nobuyuki Utsumi <sup>3,4</sup>, Ryuhei Tsuji <sup>5</sup> and  
Yoshimichi Sato <sup>1</sup>

<sup>1</sup>Faculty of Humanities, Kyoto University of Advanced Science, Kyoto, Japan, <sup>2</sup>Center for Social and Psychological Research of Metaverse, Kyoto University of Advanced Science, Kyoto, Japan, <sup>3</sup>Faculty of Engineering, Kyoto University of Advanced Science, Kyoto, Japan, <sup>4</sup>School of Environment and Society, Institute of Science Tokyo, Tokyo, Japan, <sup>5</sup>Faculty of Applied Sociology, Kindai University, Higashi-Osaka, Japan

A potential factor for video-meetings that negatively influences users' psychological aspects is the reduction of the we-mode. The we-mode refers to a cognitive state in which team members share mental states, allowing them to coordinate actions based on each other's conditions and enhance joint performance. To address this reduction, we developed a prototype video-meeting system (virtual connected room) designed to enhance the presence of users as if they were in the same room. This study evaluated the effectiveness of a virtual connected room from the perspective of the we-mode. In the experiment, a pair of participants performed three tasks in face-to-face, virtual connected, and display-based remote conditions. We used the referential communication task to evaluate the efficiency of communications and the joint Simon and number judgement tasks to evaluate the we-mode. The results showed that the virtual connected condition influenced the responses of joint Simon and number judgment, but did not influence the performance of the referential communication task. These results suggest that the virtual connected room promotes the we-mode.

## KEYWORDS

video-meeting, we-mode, co-representation, joint Simon effect, spontaneous visual taking

## 1 Introduction

Since the onset of the COVID-19 pandemic, the demand for video-meeting systems has increased across various fields. Despite their benefits, these systems have been reported to impair communication efficiency and produce negative emotional symptoms (Murphy, 2020). Research has shown that video-meetings cause delays (Boland et al., 2022) and decreases (Balters et al., 2023) in conversational turn taking. Furthermore, the long-term utility of video-meeting systems has been linked to depression (Montag et al., 2022), anxiety (Kaplan-Rakowski, 2021), and chronic tiredness (Zoom fatigue: Shoshan and Wehrt, 2021). To understand the mechanisms underlying these negative aspects, it is important to evaluate video-meetings from the viewpoints of cognitive and psychological science.

A potential cognitive factor that can negatively influence the efficiency of communication in video-meetings is the decrease in "we-mode," a cognitive state in which team members share task-relevant mental states such as task rules, perspectives, and intentions. Gallotti and Frith (2013) asserted that interacting with partners in joint tasks

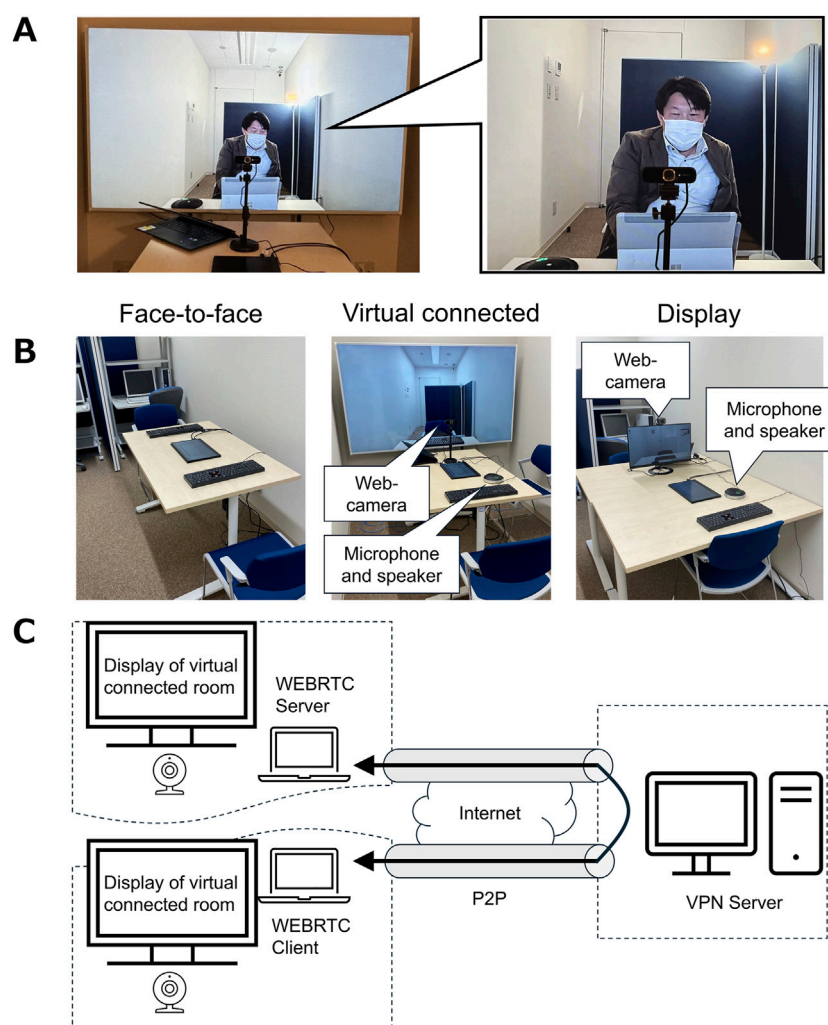


FIGURE 1

Illustration of experimental settings. (A) Virtual connected room. (B) Three experimental conditions: face-to-face, virtual connected, and display, from left to right. (C) A schematic illustration of network system in the virtual connected room.

promotes the sharing of minds among team members. Shared minds can coordinate actions based on a partner's condition and improve joint performance. For instance, during a volleyball match, when an opponent spikes the ball, each member seamlessly performs various actions such as blocking, tossing, and spiking, according to their roles. These coordinated actions are supported by the we-mode.

The we-mode has been considered to consist of cognitive states such as co-representation and spontaneous perspective-taking. Co-representation is a state in which the task rules for each member are shared among members (Sebanz et al., 2006). For example, in volleyball, blockers block the opposite ball, setters set the ball, and attackers spike the ball to get points. These representations are shared among team members to coordinate their actions. Spontaneous perspective-taking is a cognitive process in which individuals unintentionally perceive a partner's visual perspectives. In joint task situations, people automatically take a partner's visual perspective and use it to recognize information related to the task (Surtees et al., 2016). Recently, Sobel and Sims (2024) showed that interactions with a partner in a video-meeting do not produce co-representation of task rules. This suggests a

reduction in the we-mode during video-meeting situations, resulting in decreased communication efficiency.

A reason why video meetings do not effectively promote the we-mode may be that the presence of a partner is weakly perceived in this situation. Previous studies have suggested that co-representation requires the perception of a partner's presence (Sebanz et al., 2003), and that spontaneous perspective-taking requires the recognition of joint contexts (Surtees et al., 2016). During a typical video-meeting, users communicate and interact in a situation where their faces are displayed on a monitor or are not shown at all. Such settings may weaken the perception of social presence, thereby reducing recognition of the joint context. If this is the case, enhancing the perception of a partner's presence may facilitate the we-mode even in video-meeting situations.

To enhance the presence of others, we developed a prototype video-meeting system that covered the entire surface of the wall in a room (Figure 1A). The virtual connected room is designed to produce an illusion of the opposite person being in the same room. The system adjusts the visual scale of the opposite person to make them appear as if they were seated on the opposite side of

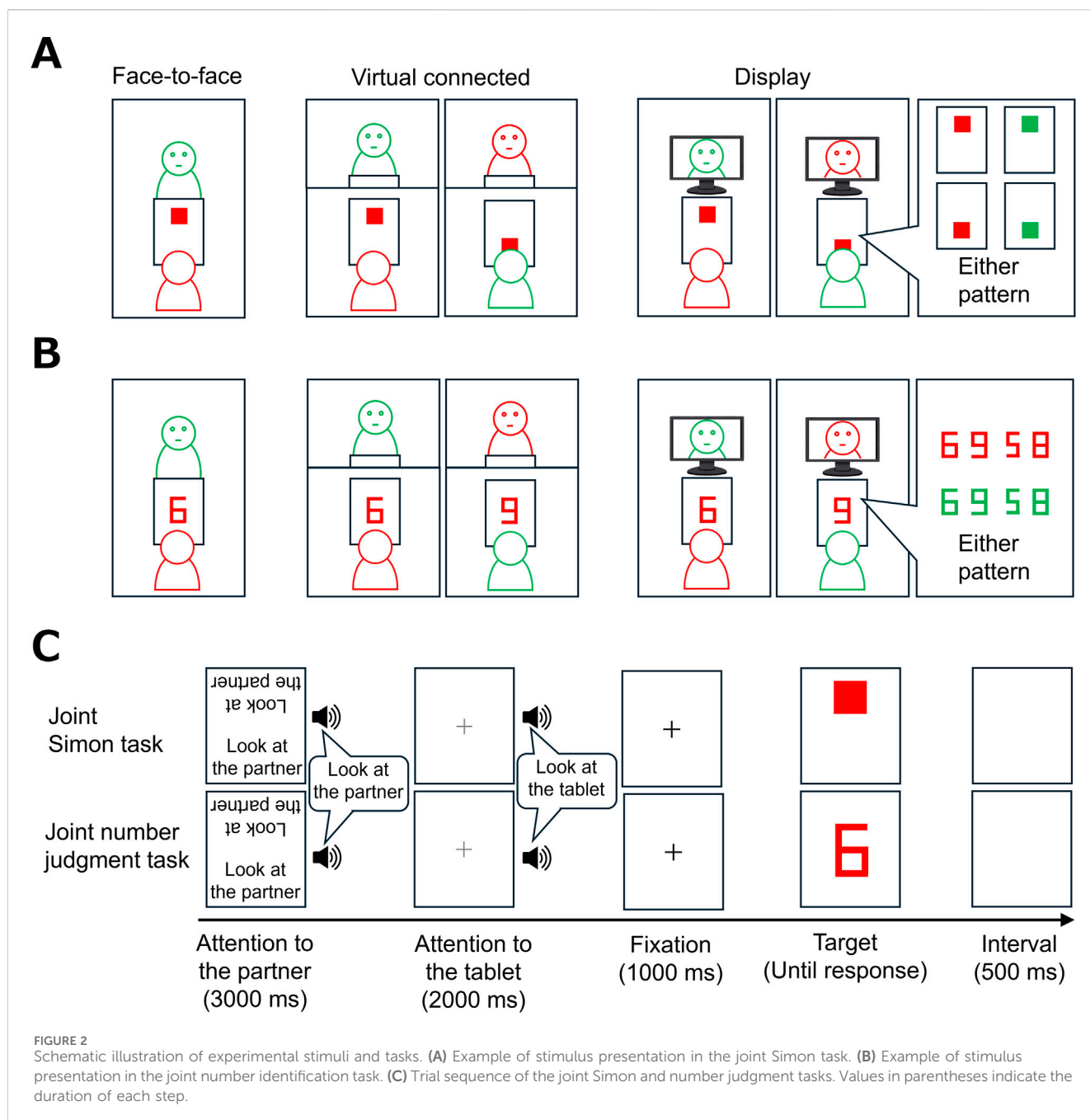


FIGURE 2

Schematic illustration of experimental stimuli and tasks. (A) Example of stimulus presentation in the joint Simon task. (B) Example of stimulus presentation in the joint number identification task. (C) Trial sequence of the joint Simon and number judgment tasks. Values in parentheses indicate the duration of each step.

the room. The edges of the display are designed with a texture similar to that of the wallpaper to reduce the visibility of the seam between the display and wall. Audio communication is facilitated by a built-in microphone that can prevent howling. The overall goal of this system is to enhance the sense of presence during remote interactions.

The present study psychologically evaluated whether the virtual connected room could promote the we-mode. In the experiment, a pair of participants performed three joint tasks under face-to-face, virtual connected, and display meeting conditions (Figure 1B). The first task was a referential communication task (Clark and Wilkes-Gibbs, 1986), which required participants to communicate geometric information to a partner. This task evaluates the

communication efficiency in a given situation (Arbuckle et al., 2000). Communication efficiency can be improved through the we-mode because of shared minds. We hypothesized that the correct responses to referential communication would be more frequent in the virtual connected condition than in the display condition. The second task was the joint Simon task (Figure 2A). In this task, two participants were asked to respond complementarily to a colored target presented near themselves or partner. Importantly, responses to targets presented near partner are suppressed when participants perceive the presence of a partner (joint Simon effect: Sebanz et al., 2003). This joint Simon effect is considered to be derived from the co-representation (Sebanz et al., 2006). Therefore, we hypothesized that a virtual connected room would produce a

joint Simon effect. The third task was a number judgment task (Figure 2B; Surtees et al., 2016). In this task, two participants were asked to judge complementarily whether the presented number (i.e., 5, 6, 8, or 9) was greater than 7. Visual perspectives between both participants were congruent for digits 5 and 8, but incongruent for digits 6 and 9. The we-mode interferes with recognition for incongruent digits more strongly than for congruent digits. Therefore, we hypothesized that the virtual connected would produce delayed responses in incongruent digits.

## 2 Materials and methods

To evaluate the effect of the virtual connected room, we conducted three behavioral tasks and analyzed participants' performance. Details regarding participants, experimental settings, and task procedures are provided below.

### 2.1 Participants

Twenty undergraduate students (four males and 16 females) participated in this experiment. They were recruited as friends via web advertisements. All participants had normal or corrected-to-normal visual acuity. Their ages ranged from 19 to 25 years ( $M = 20.85$ ;  $SD = 1.56$ ). The sample size was determined based on previous studies investigating cognitive performance in joint Simon tasks (Liepelt et al., 2011; Tsai et al., 2008).

### 2.2 Experimental setting

To evaluate the virtual connected room, we used three experimental settings: face-to-face, virtual connected, and display (Figure 1B). In the face-to-face condition, participants physically interacted with each other to perform joint tasks. In the virtual connected and display conditions, the participants interacted with their partners in different rooms using remote techniques. In the two conditions, visual and auditory information was received by web cameras and microphones and presented to each other by displays and speakers.

The virtual connected room comprises pairs of large displays (SONY KJ-7580WK: 75 inch), laptop PCs (HP OMEN16 with Core i7 and RTX3070Ti), audio input/output devices (YAMAHA YVC-200), and web cameras. The video images were set to a 115-degree horizontal visual angle using an angle adjustment function (with a resolution of  $1,920 \times 1,080$ ). The width of the display was 167.5 cm, and it entirely covered the wall of the experimental room (190 cm wide). The web camera (Anker PowerConf C300) was positioned approximately 55 cm from the edge of the table based on the visual angles of the camera and display. In this setting, the visual size of the partners was manually adjusted but was not quantified using physical measurement tools. As shown in Figure 1C, the network between the two virtual connected rooms was connected using VPN software. This ensures fairness in network settings. The OBS corresponding to the WHIP was used to communicate video images and voices between the two systems using WebRTC with a P2P connection. Server functions were performed by Go2RTC,

which produces a latency ranging from approximately 10–30 ms under the best efforts.

The display condition consisted of pairs of displays (ASUS VZ249: 24 inch), laptop PCs (THIRDWAVE F-14RP5 with Core i5 and Iris Xe Graphics), audio input/output devices (YAMAHA YVC-200), and web cameras (Logicool C980GR). The video images were set at a 25-degree horizontal angle (the resolution was  $640 \times 360$ ). Zoom (Zoom Video Communications) was used to communicate audio-visual information. The latency ranged from approximately 10–15 ms under the best efforts.

For the joint Simon and number judgment tasks, visual stimuli were presented on a 15.6-inch tablet (cocopar B07X32HD76). Participants' responses were received by a keyboard. The tasks were created and run by MATLAB (Mathworks) with PsychToolBox (Brainard, 1997).

### 2.3 Task

We conducted three tasks to evaluate the virtual connected room from a psychological perspective.

#### 2.3.1 Referential communication task

Communication efficiency was evaluated using the referential communication task. In this task, participants communicated with each other to select a target tangram from an alternative. A set of eighteen tangram figures was taken from Micklos et al. (2020). Before the task, one participant received a target tangram and the other received a list of tangram alternatives. The former participant was required to explain the details of the target tangram verbally and non-verbally to the latter participant, who did not know which tangram was target. The latter participant was allowed to ask questions about the details of the target tangram until he or she was ready to select a thought-to-be plausible option from the alternatives. This task was repeated for 4 minutes under each experimental condition.

#### 2.3.2 Joint Simon task

In the joint Simon task (Figure 2A), two participants sat on chairs in front of a tablet in the same room (face-to-face condition) or in different rooms (virtual connected and display conditions). Before this task, one participant received the target color (i.e., red or green), and the other participant received the remaining color. The trial sequence for the task consisted of several steps (Figure 2C). First, the participants looked at their partner's face physically or remotely according to the auditory and visual instructions (3,000 ms). Second, they looked at the tablet according to another auditory instruction (2,000 ms). Thirdly, a fixation cross ("+") was presented for 1,000 ms to direct participants' attention to the center of the tablet. Subsequently, a square of red or green color was presented near own (compatible targets) or partner (incompatible targets) on the tablet until responses. The task of the participants involved pressing a key as accurately and quickly as possible when a square of a given color was presented, irrespective of the location. After 500 ms, the next trial began. No feedback was provided. The target location and color were randomized across the trials. There were 40 trials for each condition as follows: target location (own and partner's sides)  $\times$  target color (red and green)  $\times$  repetition (10).

### 2.3.3 Joint number judgment task

The trial sequence was the same as that of the joint Simon task, except for the following details (Figure 2B). After the fixation cross, a red or green number (5, 6, 8, or 9) was presented at the center of the tablet until responses. When the number was a target color, the participants pressed the upper key when the number was larger than seven or the lower key when it was smaller than seven as accurately and quickly as possible. After 500 ms, the next trial began. The target number and color were randomized across the trials. The total number of trials was 40: target number (5, 6, 8, and 9)  $\times$  target color (red and green)  $\times$  repetition (5).

## 2.4 Procedure

A pair of participants conducted the referential communication, joint Simon, and number judgment tasks under three experimental conditions. The order of the conditions was pseudo-randomized across the pairs of participants. A minimum of a 5-min break was provided between tasks to minimize fatigue and maintain optimal cognitive performance. To evaluate the balance of the condition order, we calculated the frequency of each of the three conditions and orders. A chi-square test showed that the order was not significantly unbalanced [ $\chi^2(4) = 0.187, p = 0.996$ ].

In each condition, the participants performed the referential communication task, joint Simon task, and joint number judgment tasks, in that order. After joint number judgment, the participants went to the next experimental condition and repeatedly conducted three tasks. The experiment ended when the final condition was achieved.

## 2.5 Data analysis

The sample size of data corresponded to the number of pairs of participants (i.e., 10) in the referential communication task but to that of participants (i.e., 20) in the joint Simon and number judgment tasks. This is because each response of the referential communication was contributed by both participants but that of joint Simon and number judgment can be calculated by each participant.

For the joint Simon and number judgment tasks, response time data obtained from incorrect and outlier responses were excluded, according to Franconeri and Simons (2003). Outlier values were operationally defined as the mean + 2 standard deviations. Consequently, we analyzed 96.00% of total data in the joint Simon task and 93.08 of that in the joint number judgement task.

## 3 Results

To examine our hypotheses, we analyzed data from three perspectives: the efficiency of communication, co-representation of task rules, and spontaneous perspective-taking. Communication efficiency was assessed based on the number of correct responses for the referential communication task. Co-representation was evaluated using response times for the joint Simon task. The spontaneous perspective-taking was evaluated using response times for the joint number judgment task.

## 3.1 Efficiency of communication

The mean number of correct responses for the referential communication task are shown in Figure 3A. To evaluate the effect of experimental condition (face-to-face, virtual connected, display) on the efficiency of communication, an analysis of variance was conducted on the correct responses. The main effect was not significant [ $F(2, 18) = 0.613, p = 0.553$ ].

## 3.2 Co-representation of task rules

Mean response times for the joint Simon task are shown in Figure 3B. To estimate the joint Simon effect, we standardized the response times based on the following formula: (response time for the incompatible target – response time for the compatible target)/response time for the compatible target (Figure 3C). These values represent the strength of the joint Simon effect, which is used as an indicator of co-representations. To examine whether the joint Simon effect was larger in the virtual connected condition than in the display condition, an analysis of variance was conducted on the standardized values with the factor of experimental condition (face-to-face, virtual connected, and display). The main effect was significant [ $F(2, 38) = 5.279, p = 0.009$ ]. A multiple comparison test revealed that the joint Simon effect was significantly greater in the face-to-face condition than in the display condition [ $t(19) = 2.748, p = 0.038$ ]. The differences between the face-to-face and virtual connected conditions and between the virtual connected and display conditions were not significant ( $ts < 2.340, ps > 0.091$ ).

Furthermore, to evaluate whether the virtual connected room produced co-representations, one sample t-tests were conducted between the joint Simon effect in each condition and zero (i.e., no effect). The results revealed that the face-to-face and virtual connected conditions produced significant joint Simon effects [ $t(19) = 5.098, p < 0.001$ ;  $t(19) = 2.310, p = 0.032$ ], whereas the display condition did not [ $t(19) = 0.846, p = 0.408$ ]. The face-to-face and virtual connected conditions produced significant joint Simon effects, whereas the display conditions did not.

These results are inconsistent with our hypothesis that co-representation would be stronger in the virtual connected room than in the display-based video meeting. Instead, they suggest that a virtual connected room promotes the co-representation of task rules, whereas a display-based video meeting does not.

## 3.3 Spontaneous taking of partner's perspective

Mean response times for the joint number judgment task are shown in Figure 3D. To estimate the spontaneous acquisition of the partner's perspective, we standardized the response times based on the following formula: (response time for the incongruent target (6, 9) – response time for the congruent target (5, 8))/response time for the congruent target (Figure 3E). These values represent the strength of spontaneous perspective acquisition. To examine whether spontaneous perspective-taking was more strongly



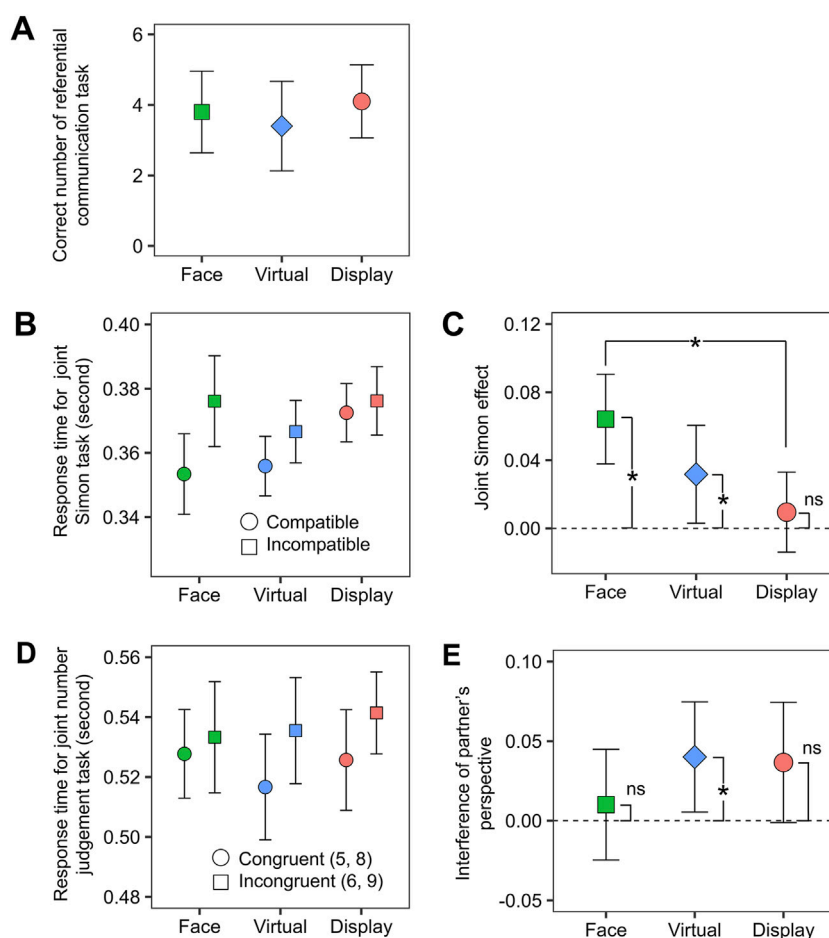


FIGURE 3

The results obtained from the three tasks. (A) The mean correct number of referential communication task. (B) Mean correct response time for the joint Simon effect. (C) Mean joint Simon effect. (D) Mean correct response times in the joint number judgment task. (E) Mean effect of interference from the partner's visual perspective. Error bars represent 95% confidence intervals. Asterisks represent significant differences ( $p < 0.05$ ). Nonsignificant differences are denoted by ns.

facilitated in the virtual connected condition than in the display condition, an analysis of variance was conducted on the values with the factor of experimental conditions. The main effect was not significant [ $F(2, 38) = 1.356, p = 0.270$ ]. The results are inconsistent with our hypothesis.

To evaluate whether the virtual connected room promoted the spontaneous acquisition, one sample t-tests were conducted between the values in each condition and zero. The results revealed that the virtual connected condition promoted spontaneous acquisition [ $t(19) = 2.424, p = 0.0255$ ], whereas the face-to-face and display conditions did not [ $t(19) = 0.608, p = 0.550$ ;  $t(19) = 2.028, p = 0.057$ ].

These results are inconsistent with our hypothesis that spontaneous perspective-taking would be stronger in the virtual connected room than in the display condition. Instead, they suggest that the virtual connected room facilitates spontaneous taking of the partner's perspective, whereas the display-based video meeting does not. By contrast, the face-to-face setting did not promote spontaneous visual acquisition, which is inconsistent with the results obtained by Surtees et al. (2016).

## 4 Discussion

This study evaluates the effectiveness of a virtual connected room from a we-mode perspective. We observed reliable effects of the joint Simon and spontaneous visual acquisition in the virtual connected condition, although communication efficiency did not improve. Based on the results, we discuss the virtual connected room with respect to the we-mode.

We observed the joint Simon effect in the virtual connected condition, but not in the display condition. This suggests that interactions through the virtual connected room promote the co-representation of task rule, a type of we-mode (Arima and Harada, 2023; Arima et al., 2025; Harada et al., 2025). Regarding the underlying mechanisms, two factors can be considered to facilitate the co-representation: perceiving the same space as a partner, and easily contacting the eyes of a partner. For the former factor, the virtual connected room is designed to perceive the presence of a partner as if they are in the same space because of the large display covering the walls of the room. Task representation has been reported to become shared as the distance between two

people decreases (Guagnano et al., 2010). Given this, same-space perception contributes to sharing task representations between the two individuals. For the latter, the cameras were positioned near the eyes of the participants in the virtual connected condition. This setting would produce eye contact between two people more frequently than in a display-remote setting. Others' eyes have been known to convey intentionality (Baron-Cohen et al., 2001), and perceiving the intentionality of a partner promotes the co-representation (Stenzel et al., 2012; Tsai et al., 2008). Consequently, the virtual connected room may share the representation between users through frequent eye contact.

The results of the joint number judgment task showed that spontaneous taking of the partner's visual perspective was produced in the virtual connected condition but not in the display condition. This suggests that the virtual connected room promotes the we-mode. However, we also observed that spontaneous taking did not occur under the face-to-face condition. Although this result is difficult to interpret, there are two possible accounts. One is that facing an intimate other in the same room might decrease the recognition of joint tasks. Spontaneous taking is not produced in a situation where two people exist but do not engage in joint tasks (Surtees et al., 2012). In our experiment, nine pairs of participants were friends, and face-to-face interactions in the same room were similar to their daily situations. This may decrease the recognition of the situation as a joint task, which may weaken the spontaneous taking of the partner's perspective. Another account of the null result is that the order of tasks might influence the participants' behaviors. In this study, the number judgment task was conducted after the joint Simon task, suggesting that participants may be experiencing more fatigue from the judgment task than from the Simon task. Previous studies have reported that performing cognitive tasks cause mental fatigue (Smith et al., 2019) and this fatigue may mediate higher cognitive processing (Guo et al., 2016). Whereas our study gave participants rest time among tasks, potential fatigue might reduce the we-mode in the joint number judgment task.

The facilitation of the we-mode in the virtual connected room may be interpreted from the perspective of enactivism and co-enactivism. Enactivist accounts suggest that cognition is not passively received but is actively constituted through self-organizing and recurrent sensorimotor contingencies (Read and Szokolszky, 2020). Co-enactivist perspectives further propose that such cognitive processes can be interpersonally co-constructed through coordinated interaction (De Jaegher and Di Paolo, 2007). From these perspectives, the virtual connected room may contribute to the co-constitution of peripersonal space between users. That is, it may enable users to perceive their partner's presence within an extended bodily space, thereby facilitating the we-mode. In contrast, the display-based video meeting may represent a non-constitution of peripersonal space due to diminished embodied interaction.

The virtual connected room did not improve communication efficiency. These results can be interpreted as two accounts. One possible explanation is that this might stem from an instability in the latency of audio communications in the virtual connected room. The range of latency was larger in the virtual connected condition (15–30 ms) than in the display condition (10–15 ms). Indeed, some participants reported that they experienced wide latency in audio communication in the virtual connected condition. Delays in audio communication reduce the efficiency and performance of communication tasks (Krauss and Bricker, 1967). Another

explanation is that this result might be attributed to limited sample sizes. While performance data for the referential communication task were obtained from only ten pairs of participants, data for the joint Simon and number judgment tasks were collected from 20 participants. This discrepancy suggests lower test power for the referential communications task. Since the present study aimed to primarily investigate the effect of a virtual connected room on the we-mode, we determined the sample size based on the requirements of the joint Simon and number judgment tasks. Future studies should address this limitation by recruiting sufficient sample sizes.

A potential limitation of this study is that the present results cannot necessarily predict the effectiveness of virtual connected room between individuals unfamiliar with each other. This study recruited pairs of participants familiar with each other because the virtual connected room is expected to be used in remote meetings for people with closer relationships. However, this study did not examine whether the virtual connected room enhances we-mode between unfamiliar people. Since there are demands in several fields for enhancing we-mode between unfamiliar people, this issue should be investigated in future studies.

## 5 Conclusion

In summary, our results showed that the virtual connected room promotes the co-representation of task rule and spontaneous perspective-taking between pairs of people, but does not increase communication efficiency. These results suggest that the virtual connected room provides an advantage towards improving the we-mode.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://osf.io/n9hef/>.

## Ethics statement

The studies involving humans were approved by the local ethics committee of the Faculty of Humanities at Kyoto University of Advanced Science. 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. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## Author contributions

YH: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing, Validation. YhS: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Resources, Software, Writing – original draft, Writing – review

and editing. AK: Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing – review and editing. KO: Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Writing – review and editing. RoT: Conceptualization, Funding acquisition, Investigation, Methodology, Writing – review and editing. NU: Conceptualization, Funding acquisition, Investigation, Methodology, Writing – review and editing. RuT: Conceptualization, Funding acquisition, Investigation, Methodology, Writing – review and editing. YmS: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review and editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by FY2024 Kyoto University of Advanced Science Research Grant (Cross-disciplinary activities to create new fields).

## Acknowledgments

We thank Nana Kamei and Mayuko Ohtsuki affiliated at Kyoto University of Advanced Science for their assistance with data collection.

## References

- Arbuckle, T. Y., Nohara-LeClair, M., and Puchkar, D. (2000). Effect of off-target verbosity on communication efficiency in a referential communications task. *Psychol. Aging* 15, 65–77. doi:10.1037//0882-7974.15.1.65
- Arima, Y., and Harada, Y. (2023). Competition or cooperation: classification in a VR environment based on sensor data. In *Proc. Int. Conf. Collab. Technol. Soc. Comput.*, Cham: Springer, 144–151. doi:10.1007/978-3-031-42141-9\_10
- Arima, Y., Harada, Y., and Okada, M. (2025). Interpersonal synchrony in human-robot interaction: Sensor analysis of interpersonal interaction using VR data [Preprint]. *Research Square*. doi:10.21203/rs.3.rs-5846894/v1
- Balters, S., Miller, J. G., Li, R., Hawthorne, G., and Reiss, A. L. (2023). Virtual (Zoom) interactions alter conversational behavior and interbrain coherence. *J. Neurosci.* 43, 2568–2578. doi:10.1523/JNEUROSCI.1401-22.2023
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001). The “reading the mind in the eyes” test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. Child Psychol. Psychiatry* 42, 241–251. doi:10.1111/1469-7610.00715
- Boland, J. E., Fonseca, P., Mermelstein, I., and Williamson, M. (2022). Zoom disrupts the rhythm of conversation. *J. Exp. Psychol. Gen.* 151, 1272–1282. doi:10.1037/xge0001150
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi:10.1163/156856897X00357
- Clark, H. H., and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition* 22, 1–39. doi:10.1016/0010-0277(86)90010-7
- De Jaegher, H., and Di Paolo, E. (2007). Participatory sense-making: an enactive approach to social cognition. *Phenom. Cogn. Sci.* 6, 485–507. doi:10.1007/s11097-007-9076-9
- Franconeri, S. L., and Simons, D. J. (2003). Moving and looming stimuli capture attention. *Percept. Psychophys.* 65, 999–1010. doi:10.3758/bf03194829
- Gallotti, M., and Frith, C. D. (2013). Social cognition in the we-mode. *TiCS* 17, 160–165. doi:10.1016/j.tics.2013.02.002
- Guagnano, D., Rusconi, E., and Umiltà, C. A. (2010). Sharing a task or sharing space? On the effect of the confederate in action coding in a detection task. *Cognition* 114, 348–355. doi:10.1016/j.cognition.2009.10.008
- Guo, Z., Chen, R., Zhang, K., Pan, Y., and Wu, J. (2016). The impairing effect of mental fatigue on visual sustained attention under monotonous multi-object visual

## 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.

## Correction note

A correction has been made to this article. Details can be found at: [10.3389/frvir.2025.1649096](https://doi.org/10.3389/frvir.2025.1649096).

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## 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.

attention task in long durations: an event-related potential based study. *PLoS ONE* 11, e0163360. doi:10.1371/journal.pone.0163360

Harada, Y., Arima, Y., and Okada, M. (2025). Effect of virtual interactions through avatar agents on the joint Simon effect. *PLOS ONE*, 20, e0317091. doi:10.1371/journal.pone.0317091

Kaplan-Rakowski, R. (2021). Addressing students' emotional needs during the COVID-19 pandemic: a perspective on text versus video feedback in online environments. *Educ. Technol. Res. Dev.* 69, 133–136. doi:10.1007/s11423-020-09897-9

Krauss, R. M., and Bricker, P. D. (1967). Effects of transmission delay and access delay on the efficiency of verbal communication. *J. Acoust. Soc. Am.* 41, 286–292. doi:10.1121/1.1910338

Liepert, R., Wenke, D., Fischer, R., and Prinz, W. (2011). Trial-to-trial sequential dependencies in a social and non-social Simon task. *Psychol. Res.* 75, 366–375. doi:10.1007/s00426-010-0314-3

Micklos, A., Walker, B., and Fay, N. (2020). Are people sensitive to problems in communication? *Cog. Sci.* 44, e12816. doi:10.1111/cogs.12816

Montag, C., Rozgonjuk, D., Riedl, R., and Sindermann, C. (2022). On the associations between videoconference fatigue, burnout and depression including personality associations. *J. Affect. Disord. Rep.* 10, 100409. doi:10.1016/j.jadr.2022.100409

Murphy, K. (2020). *Why Zoom is terrible*. New York Times, Vol. 23. Available online at: <https://www.nytimes.com/2020/04/29/sunday-review/zoom-video-conference.html> (Accessed April 29, 2020).

Read, C., and Szokolszky, A. (2020). Ecological psychology and enactivism: perceptually-guided action vs. sensation-based enaction. *Front. Psychol.* 11, 1270. doi:10.3389/fpsyg.2020.01270

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *TiCS* 10, 70–76. doi:10.1016/j.tics.2005.12.009

Sebanz, N., Knoblich, G., and Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition* 88, B11–B21. doi:10.1016/s0010-0277(03)00043-x

Shoshan, H. N., and Wehrt, W. (2021). Understanding “Zoom fatigue”: a mixed-method approach. *Appl. Psychol.* 71, 827–852. doi:10.1111/apps.12360



- Smith, M. R., Chai, R., Nguyen, H. T., Marcora, S. M., and Coutts, A. J. (2019). Comparing the effects of three cognitive tasks on indicators of mental fatigue. *J. Psychol.* 153, 759–783. doi:10.1080/00223980.2019.1611530
- Sobel, B. M., and Sims, V. K. (2024). Joint action over Zoom: assessing corepresentation with remote partners using the joint Simon task. *Research Square* doi:10.21203/rs.3.rs-4509905/v1
- Stenzel, A., Chinellato, E., Tirado Bou, M. A., del Pobil, Á. P., Lappe, M., and Liepelt, R. (2012). When humanoid robots become human-like interaction partners: corepresentation of robotic actions. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 1073–1077. doi:10.1037/a0029493
- Surtees, A., Apperly, I., and Samson, D. (2016). I've got your number: spontaneous perspective-taking in an interactive task. *Cognition* 150, 43–52. doi:10.1016/j.cognition.2016.01.014
- Surtees, A. D. R., Butterfill, S. A., and Apperly, I. A. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. *Br. J. Dev. Psychol.* 30, 75–86. doi:10.1111/j.2044-835X.2011.02063.x
- Tsai, C., Kuo, W., Hung, D. L., and Tzeng, O. J. L. (2008). Action co-representation is tuned to other humans. *J. Cogn. Neurosci.* 20, 2015–2024. doi:10.1162/jocn.2008.20144