TYPE Original Research
PUBLISHED 15 September 2025
DOI 10.3389/frvir.2025.1623764

Classifying interpersonal interaction in virtual reality: sensor-based analysis of human interaction with pre-recorded avatars

Yoshiko Arima*, Yuki Harada and Mahiro Okada

Department of Psychology, Center for Social and Psychological Research of Metaverse, Kyoto University of Advanced Science, Kyoto, Japan

Check for updates

OPEN ACCESS

EDITED BY
Antonio Sarasa-Cabezuelo,
Complutense University of Madrid, Spain

REVIEWED BY
Asuka Takai,
Advanced Telecommunications Research Institute International (ATR), Japan
Mengting Gong,
Jinan University, China

*CORRESPONDENCE
Yoshiko Arima,
arima.yoshiko@kuas.ac.jp

RECEIVED 06 May 2025
ACCEPTED 19 August 2025
PUBLISHED 15 September 2025

CITATION
Arima Y, Harada Y and Okada M (2025) Classifying interpersonal interaction in virtual reality: sensor-based analysis of human interaction with pre-recorded avatars.
Front. Virtual Real. 6:1623764.
doi: 10.3389/frvir.2025.1623764

This study investigates human engagement with a non-responsive, pre-recorded avatar in VR environments. Rather than bidirectional collaboration, we focus on unidirectional synchrony from human participants to the avatar and evaluate its detectability using sensor-based machine learning. Using a random forest model, we classified interactions into cooperation, conformity, and competition, achieving an F1 score of 0.89. Feature importance analysis identified hand rotation and head position as key predictors of interaction states. We compared human-human and human interaction with a non-responsive avatar (pre-recorded motion replay) during a joint Simon task by covertly switching collaborators between humans and non-responsive avatars. Using the classification model, a synchrony index was derived from VR motion data to quantify behavioral coordination patterns during joint actions. The classification indexes were associated with higher cooperation in human-human interactions ($p = 0.0262$) and greater conformity in human interaction with a non-responsive avatar ($p = 0.0034$). The synchrony index was significantly lower in the non-responsive avatar condition ($p < 0.001$), indicating reduced interpersonal synchrony with non-responsive avatars. These findings demonstrate the feasibility of using VR sensor data and machine learning to quantify social interaction dynamics. This study aimed to explore the feasibility of a sensor-based machine learning model for classifying interpersonal interactions in VR, based on preliminary data from small-sample experiments.

KEYWORDS

virtual reality, human activity recognition, machine learning, joint simon effect, interpersonal synchrony

1 Introduction

We applied wearable sensor-based motion analysis and machine learning (decision trees and linear mixed models) to classify interpersonal synchrony patterns under human–human and human–avatar (non-responsive) conditions. This approach provides quantitative insight into how interaction context modulates synchrony. We propose a synchrony index derived from VR motion and gaze data, which captures behavioral coordination patterns during joint actions. To examine variations in synchrony, we designed an experimental setting in which participants engaged in a joint Simon task while their collaborator was covertly switched between a human and a non-responsive avatar.

2 Joint simon experiment

The Simon effect (Simon, 1969) is a spatial compatibility effect in which a match or mismatch between the spatial location of a stimulus and its response influences behavior. For example, suppose that red or green stimuli appear randomly on the left or right side of a screen as targets, the response is delayed if the button and stimulus positions do not match, whereas if the task is a Go/No-Go task, a delay does not occur. However, when two stimuli are assigned individually to a pair, the Simon effect reappears as if the pair represents a single person, even though each individual's task is identical to that of the Go/No-Go task (Sebanz et al., 2003). This is known as the joint Simon effect (JSE). Our research team has confirmed that the JSE occurs in VR environments (Harada et al., 2025). This study utilizes the same VR experimental setting to examine how human-human and human interaction with a non-responsive avatar influence social coordination and synchrony.

Explanatory theories suggest that the JSE reflects either (i) task co-representation or (ii) spatial coding relative to the position of the collaborator (Dolk et al., 2014). Experiments comparing these two explanatory theories indicate that the latter, i.e., the reference-coding hypothesis (Sellaro et al., 2015; Sangati et al., 2021), is supported by more studies. However, as will be discussed below, the co-representation hypothesis is not precluded because the JSE weakens when collaborators are taught that they are unconscious, non-living entities. These hypotheses relate to whether we recognize non-living collaborators merely as reactive entities or as agents capable of shared representations. This pilot study serves as a proof-of-concept for applying sensor-based machine learning models to classify social coordination patterns in immersive virtual environments.

2.1 JSE with bot

Previous research on the JSE with non-human collaborators has yielded conflicting results (Stenzel et al., 2016). reported that the JSE occurs even when the collaborator is a non-living entity. In contrast (Tsai et al., 2008), analyzed action indices and event-related potentials and found that the JSE emerged only when participants believed their partner to be human.

This discrepancy can be explained by the perception of intentionality (Tsai et al., 2007; Stenzel et al., 2012). demonstrated that the JSE is enhanced when the collaborator is perceived as having intentionality, suggesting that recognizing intentionality facilitates action simulation in the motor system. Furthermore (Stenzel and Liepelt, 2014), found that the perception of agency—i.e., seeing another person pressing a button—precedes the cognition of that person's intention. Agency can be inferred by observing simple moving figures, even in the absence of explicit social cues (Heider and Simmel, 1944).

Thus, not all JSE-related co-representation processes rely on higher-order cognition (Miss et al., 2022; Liepelt et al., 2016) demonstrated that the JSE was intensified when activity in the anterior cingulate cortex, which is associated with motor intentions, was suppressed. Their findings suggest that when the JSE occurs, the distinction between self and other motor intentions becomes less defined. While the perception of agency is processed automatically through perceptual cues, the intention of others is subsequently inferred.

To further investigate factors influencing automatic JSE processes, this study examines interpersonal synchrony as an indicator of self-other undifferentiated states (Paladino et al., 2010). Additionally, we explore whether participants recognize the bot as a human in interpersonal synchronization.

2.2 Interpersonal synchrony

Face-to-face communication evokes a subconscious process of spontaneous synchronization of attention, behavior, and brain waves. A meta-analysis of synchrony studies showed that sensory and interpersonal synchrony resulted in prosocial attitudes and behaviors (Rennung and Göritz, 2016). As a causal effect in the opposite direction, pro-sociality can promote synchrony. For example (Fronda and Balconi, 2022), demonstrated that the act of giving affected performance and brain-brain synchrony during cooperative tasks. Smykovskyi et al. (2024) revealed that negative emotions disrupted intentional synchrony during sensorimotor interactions. Furthermore (Hao et al., 2024), showed that group identity influenced brain-to-brain synchrony and cooperative decision-making behaviors.

Interpersonal synchrony is assumed to be an automatic process because it occurs within a short reaction time (RT) (Decety et al., 2011). Synchrony studies have primarily been conducted by measuring the cross-correlation coefficient (CCC) of physiological data. For example (Guastello et al., 2023), proposed a system in which each member's physiological data was obtained individually, and then cross-correlation was used to distinguish multiple influences on others.

In the present study, we classified interpersonal activities using sensor data related to pairwise units and applied them to human activity recognition (HAR) research. HAR has yielded numerous results through the use of smartphone sensor data and other machine-learning sources to classify activity types, particularly in exercise situations.

2.3 Social presence

Social presence refers to the perception of being attended to and understood by another entity during an interaction. Recent studies have shown that people can experience social presence even with artificial agents under certain conditions. For example, Chen et al. (2023) developed and validated a multidimensional scale for assessing robot social presence, expanding traditional dimensions such as physical presence and conscious awareness to include interactional aspects like dialog behavior and emotional understanding. Similarly, Sogemeier et al. (2024) reported that temporal cues, such as response latency, were more influential than visual realism in eliciting social presence with in-car voice assistants, suggesting that behavioral responsiveness may be more critical than appearance in creating a sense of connectedness. Munnukka et al. (2022) further demonstrated that perceived anthropomorphism increased social presence in web-based avatar interactions, which in turn fostered trust, although avatar appearance itself had no significant effect on perceived anthropomorphism. Based on these findings, the present study implemented a non-verbal VR bot avatar that replicated recorded human motion but did not engage in speech or dialog. Due to the small sample size, we used only two 7-point Likert-scale items assessing how realistic and human-like the avatar appeared. These items formed a minimal composite index of perceived social presence. At the end of the experiment, participants were also asked to identify in which session they believed they had interacted with a bot. Follow-up interviews were conducted to explore when and how they noticed the bot—or whether they failed to detect it at all. This combined qualitative and quantitative data was used to construct a binary variable ("Bot-Notice") indicating whether the bot was consciously recognized.

3 Research question

In this study, we aimed to develop a machine learning-based classification model for interpersonal interactions in VR using sensor data. To validate this model, we examined differences between human-human and human-bot interactions in a joint Simon task.

Research Question 1: To what extent can interpersonal interactions in cooperative tasks be effectively classified using VR sensor data?

As part of our initial exploratory analysis, we conducted a preliminary experiment to classify the activities of pairs in the Simon task based on two basic types of interpersonal interaction in the social sciences: competition and cooperation. We expected that interpersonal synchrony would be more prominent in behaviors classified as cooperative.

The purpose of the preliminary experiment was to discover important features for classifying interpersonal behavior from various sensor data, while determining which phase of the Simon task could be more accurately classified by dividing the task into smaller phases. For this purpose, we used a random forest model, which makes it easy to judge the importance of features, and adopted the most important feature as an indicator of synchrony. Random forest is a machine learning technique widely used in machine learning competitions due to its high prediction accuracy and robustness against overfitting. Sekitani and Murakami (2022) compared 30 statistical and machine learning models, including their combinations, using symmetric mean absolute percentage error (sMAPE) and mean absolute scaled error (MASE). Their results demonstrated that among individual machine learning methods, Random Forest achieved the highest accuracy. Unlike single decision trees, random forest mitigates overfitting by aggregating multiple trees, enhancing generalization performance. Additionally, it provides an intuitive method for evaluating feature importance, making it a valuable tool for understanding the contribution of each variable in predictive modeling.

Research Question 2: What are the key differences between human-human and human interaction with a non-responsive avatar?

In the main experiment, we established a bot condition in which a bot avatar was introduced as a collaborator and compared the bot condition with a human condition, where participants performed the joint Simon task with a human partner. The bot avatar was created by tracing the sensor data of a human in a preliminary experiment. In the bot condition, synchronization from human to bot is expected, but synchronization from bot to human does not occur. Therefore, it is expected that synchrony in the bot condition will be reduced compared to the human condition. We hypothesized that synchrony would be the key difference between the bot and human conditions, while also exploring other potential differences.

4 Methods

4.1 Preliminary experiment

4.1.1 Participants

Eight participants (six men and two women; college students aged 19–21 years) enrolled in the study. The participants were segregated into four groups, with each pair referred to as collaborators.

4.1.2 Participation-agreement procedures

Recruitment was open for 1 week from 17 February 2023. Participants were given an explanation of the consent document in the laboratory, and the informed consent procedure was carried out. The participants were handed a paper that outlined the experiment and data-handling procedures, which were explained by the experimenter. All eight participants agreed to participate in the study. The experimental data were obtained using anonymized ID numbers. This ensured that the data were not linked to the participants' names.

4.1.3 Devices

The VR systems were established in two separate rooms. Each system comprised a VIVE Pro Eye (HMD), two controllers (VIVE Controller 2018), two base stations (SteamVR Base Station 2.0), and a computer. The VR environment was created using Unity (2021.3.1f1) in a server-client network using "Netcode for Game Objects." In this environment, paired participants entered the same virtual space and interacted via physical actions. No audio communication was available, and the VR environment featured two avatars, buttons, a display, and a mirror (Figure 1). All experimental configurations and spatial arrangements shown in Figure 1 represent the layout within the virtual reality environment, not the physical laboratory setup. The avatars were able to move based on six-coordinate data (three positions and three rotations) obtained from the HMD and two controllers. These avatars were boxy and lacked personality traits, and their movements were executed using the "Final IK (Inverse Kinematics)" asset. Red- and green-labeled reaction buttons were placed in front of the avatar in the VR space. The RTs were acquired via collision detection when the avatar touched a button.

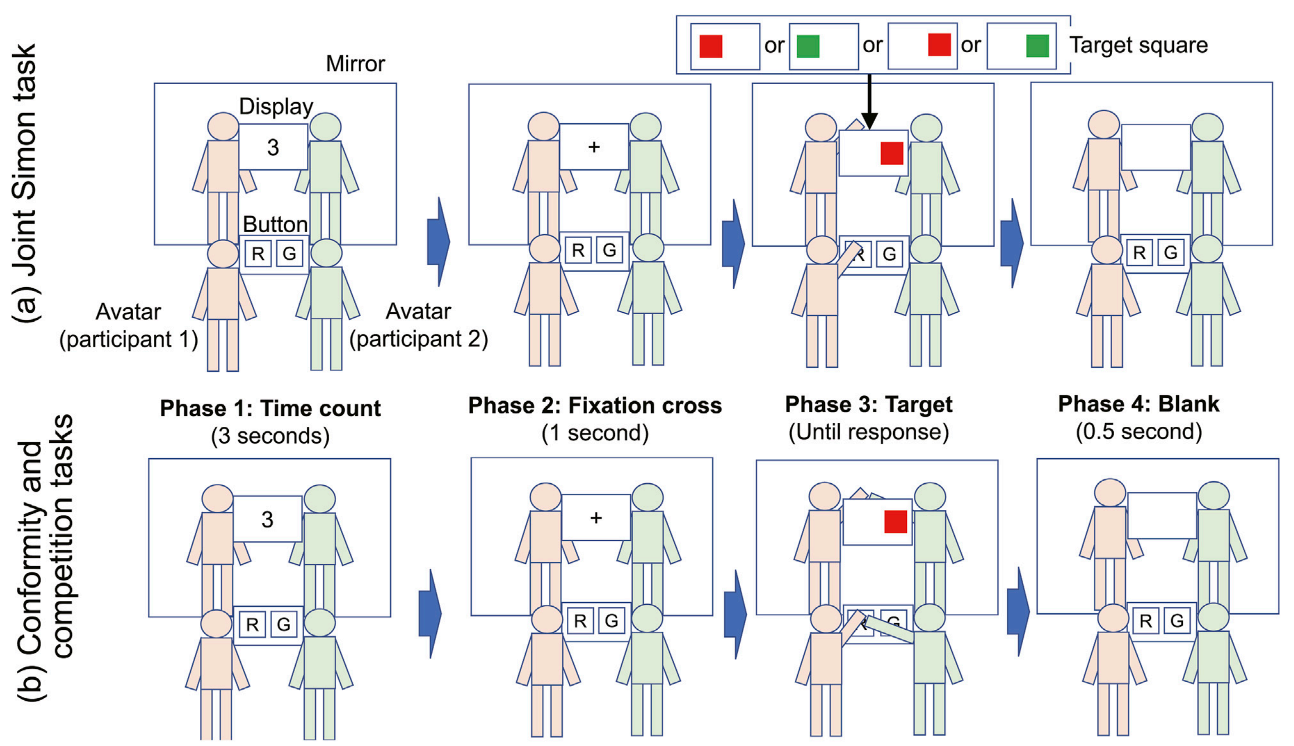


FIGURE 1
Experimental setup for the joint Simon task and conformity/competition tasks. Note: The upper section shows the joint Simon task procedure, while the lower section illustrates the conformity and competition tasks.

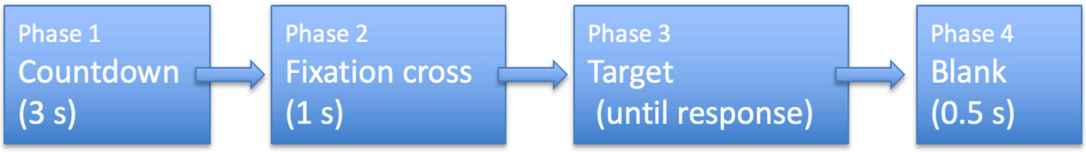


FIGURE 2
Task phases in the joint Simon task. Note: Phase 1: A 3-s countdown display. Phase 2: Presentation of a black fixation cross, "+", at the center of the display for 1 s. Phase 3: Presentation of targets (red or green) on either the left or right side of the display until a response was obtained. Phase 4: A blank interval of 0.5 s before the next countdown began.

The task comprised four phases: Phases 1-4 (time count, fixation cross-presentation, target presentation, and blanks, respectively). Phases 1 and 3 were the countdown and motion phases, respectively (See Figure 2).

The participants were instructed to touch a button corresponding to the target color, regardless of its location. Each session comprised 16 or 32 consecutive trials, with the target color and position randomized between the trials.

4.1.4 Procedure

The participants were allowed to select either the client or host experimental booths. The terms "host" and "client" were designated because paired data were transmitted as streamed data from the client to the host PC. Following the instructions of the experimenter stationed at each booth, the participants were instructed to wear the HMDs and operate the controllers with both hands. Before each session, the participants were briefed on the colors of the stimuli for which they were responsible. Before commencing the joint task, the participants were instructed to view their collaborators directly and then confirm their avatars in the mirror set in the VR space.

The host stood on the right, whereas the client stood on the left. Figure 1 illustrates the experimental setup for the joint Simon task and conformity/competition tasks within the virtual reality environment. Participants interacted with color targets displayed on screen, responding by pressing corresponding buttons in the virtual environment according to task instructions. The host operated the buttons in the VR space using the left hand, whereas the client used the right hand. Thus, the right hand was not used on the host side, and the left hand was not utilized on the client side. The task involved pressing a button labeled with the

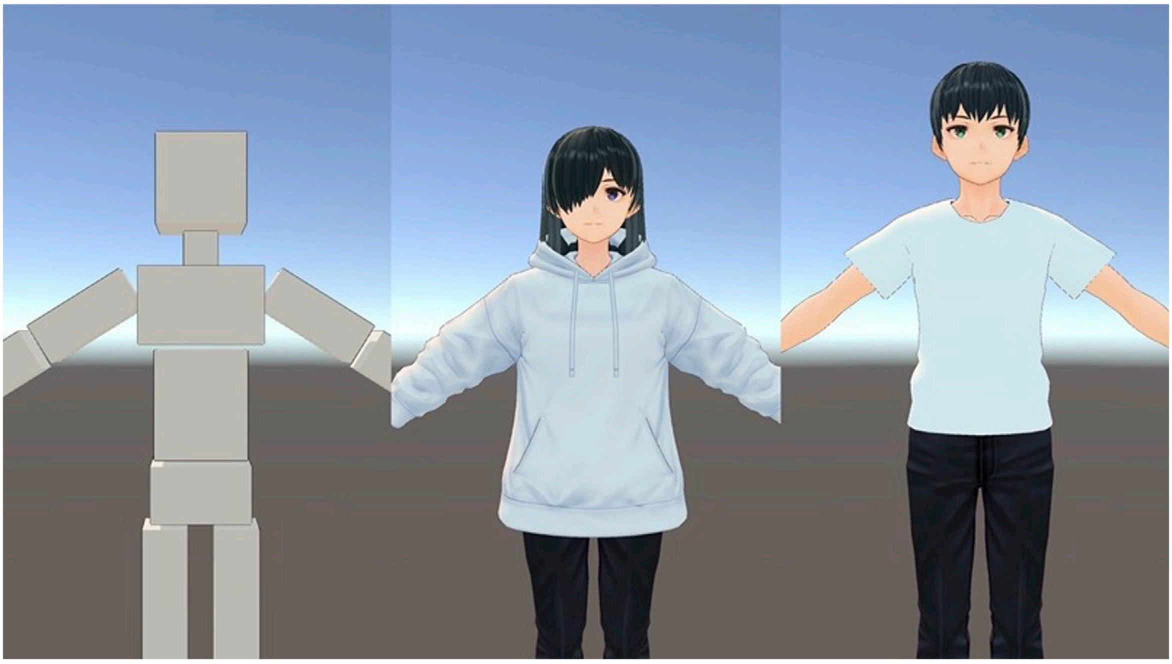


FIGURE 3
Avatars used in the main experiment. Note: For box-shaped avatars, the leftmost avatar was used regardless of participant gender. For human-shaped avatars, female participants used the middle avatar, while male participants used the left avatar. The avatars shown were generated using VRoid Studio (© pixiv Inc.), which permits research use under its license.

corresponding color name when the assigned color appeared. The participants were instructed to halt if they felt uncomfortable, lift their HMDs at the end of each session, and take breaks as required.

4.1.5 Sessions

The participants entered the space individually and completed eight practice trials for the Go/No-Go task. During the practice session, the correct answer was indicated when the correct button was touched. An incorrect answer was revealed when another button was touched or when a certain amount of time had elapsed without a touch being detected. If a participant failed in all eight trials, then the practice session was repeated. After the practice session was completed, the following sessions were conducted.

Session 2 involved the procedure shown in the upper section of Figure 1. The target colors in Session 3 were swapped to minimize the learning effects. Sessions 4 and 5 involved the procedure shown in the lower section of Figure 1. The detailed session structure for the preliminary experiment is presented in Table 1.

Sensor data from Sessions 4 (conformity) and 5 (competition) were used for machine learning and testing, respectively. In the conformity task session, the participants were instructed that “whichever target appears, touch the correct button at the same time as your companion.” During the competition task session, the participants were instructed that “whichever target appears, touch the correct button before your companion.”

4.1.6 Data processing

In the preliminary experiment, we investigated the features necessary for distinguishing between interpersonal behaviors in VR environments. To identify subtle differences in subconscious movements, we used the sensor data during Phase 1, i.e., the time at which the participants were staring at the countdown, as shown in Figure 1. The phases and procedures illustrated in Figure 1 were all conducted within the virtual reality space.

TABLE 1 Session structure for the preliminary experiment.

Session	Task	Target assignment	Trials
1	Go/No-Go task	Individual sessions for assigned target	32
2	Joint Simon task	Host: green; Client: red	32
3	Joint Simon task	Host: red; Client: green	16
4	Conformity task	Simultaneous touching	16
5	Competition task	Touch before opponent	16

The sensing data included gaze direction, eye position, pupil size (left and right), head position and rotation, and the position and rotation of the left and right controllers. For each of these, XYZ three-axis data were recorded where applicable. The transmission latency from the client to the host was approximately 0.01 s. Signals were sampled at a variable rate (80 Hz average), and after missing values were removed, the client and host data were linked at intervals of approximately 0.02 s and then used for machine learning. The features used for machine learning were: distance, which was obtained as the root sum of squares of the XYZ (Euler angle) of the position and gyro sensor at each sampling point; the velocity from the time difference; and the acceleration obtained from the time difference in velocity, which

TABLE 2 Session structure for the main experiment. Conformity task: participants were instructed to touch the correct target simultaneously. Competition task: participants were instructed to touch the correct target faster than their partner. Joint Simon task: participants were instructed to touch the button only when the target color assigned to them appeared.

Session	Task description	Avatar type	Trials
1	Go/No-Go task (individual)	Box avatar	32
2	Joint Simon task (human-human pair)	Box avatar	32
3	Joint Simon task (human-human pair, target colors swapped)	Human avatar	32
4	Conformity task (human-human pair)	Human avatar	16
5	Competition task (human-human pair)	Human avatar	16
6	Joint Simon task (human vs. Bot pair)	Human avatar	32
7	Joint Simon task (human vs. Bot pair, target colors swapped)	Box avatar	16

was used as the analysis data. After deleting samples with missing values, we used the Python sklearn Random Forest Classifier (n_ estimators = 250, random_state = 42) as the random forest model. A set of decision trees was constructed for a subset of randomly sampled training data, and predictions based on a subset of these features were aggregated to obtain the final prediction. After testing various feature types, we found that distance features yielded relatively high classification accuracy, whereas models using velocity and acceleration performed poorly. Therefore, we decided to use only distance features, except for triaxial gaze data, which were retained as they are considered essential for synchronization. To evaluate classification accuracy, cross-validation was conducted by iteratively designating data from three out of eight participants as the test set, while data from the remaining five were used for training. This process was repeated to ensure that each participant appeared in the test set at least once. The final model was trained on the full dataset after cross-validation.

4.2 Main experiment

4.2.1 Participants

The participants of this experiment were recruited through a university website, and assigned to each experimental day. Recruitment was open for 2 months from 14 May 2023, and the informed consent procedure was the same as the preliminary experiment, but was obtained in advance via a web-based questionnaire in order to avoid coercion in obtaining consent due to face-to-face situations in the laboratory. Owing to the absence of one participant, the final number of participants was 18, which comprised seven men, nine women, and two other genders. The average age of the participants was 19.83 years (standard deviation [SD] = 1.04). All participants were assessed for handedness using a self-report questionnaire. Of the 18 participants, 16 were right-handed, 1 was left-handed, and 1 reported being ambidextrous or having no hand preference. After the experiment, the participants were instructed to complete a questionnaire survey and interview, for which they received an honorarium of approximately $10 (1,500 yen) after completion. The device and experimental procedures were identical to those used in the preliminary experiments. Two types of avatars, i.e., a box and a human, were designed for other research purposes (Figure 3).

4.2.2 Sessions

All participants completed the sessions in the same fixed order shown in Table 2.

4.2.3 Dependent variable

Correct response rate: Correct responses were counted when participants touched the correct color target in trials where they were required to respond and refrained from touching in trials where they were not. Subsequently, the correct response rate was divided by the number of trials.

JSE: The mean RT delay (RTs for incompatible targets minus RTs for compatible targets) during the joint Simon task (Sessions 2, 3, 6, and 7) minus that of the Go/No-Go task (Session 1) was calculated. The RTs for correct responses with more than two standard deviations from the mean RT were excluded as outliers. Additionally, pairwise data from participants whose RT could not be measured because of equipment failure were excluded.

Bot cognition: After the experiment, a structured assessment combining questionnaires and follow-up face-to-face interviews was conducted to systematically evaluate participants' awareness of the bot condition. The assessment protocol was designed to minimize leading questions and retrospective bias. A participant who perceived the human collaborator to be a bot was assumed to be unaware of the discrimination between humans and bots. In the data analysis, binary values of 1 and 0 were used to indicate the awareness and unawareness of bots, respectively. As a proxy for perceived social presence, we also included a two-item measure rated on 7-point Likert scales. The items assessed how realistic and how human-like the avatar appeared. The sum of these two items was used as an index of social presence (Mean = 8.78, standard deviation [SD] = 2.29).

Sensor data: By performing the procedures of the preliminary experiment, the distance was obtained as the root sum of squares of the XYZ (Euler angle) of the position and gyro sensor at each sampling point; the velocity was the time difference between the two; and the acceleration was the time difference between the two. Because the accuracy of the classification model using velocity and acceleration data was low during the machine learning process, we adopted a model that used only distance features. The resulting features were the HMD position, HMD rotation, and 12 variables of position information for the left- and right-controller positions and rotations (Figure 2). The eye-gaze and pupil size features used in the preliminary experiments were not used in the main experiment because no sensing data corresponded to the bot. Under the bot condition, only trace data from Phase 3 were used; thus, data from Phase 3 were used to formulate the classification model. Samples with missing values on the host or client side were deleted. After removing missing values, the number of observations obtained from Sessions 3, 4, and 5 was 16,759 for training and testing, out of a total of 124,193 observations (approximately 86.9% of the original data were retained). Sensor data from Sessions 2, 6, and 7 were prepared as files on the host and client sides to compare the human and bot conditions.

Pair Activity Probability: Details of machine learning, angular transformations, and statistical models are described in Section 4.3.

Synchrony Index: As in the preliminary experiment, the most important feature for classifying paired activities was the rotation of the unused hand (host side right-hand rotation). Therefore, the cross-correlation (CCC) of the sensor data for the host side right-hand rotation and client-side left-hand rotation was calculated for each trial and then used as the interpair synchrony index using MATLAB's XCORR function. After normalizing the sensor data for each trial, the maximum value obtained at lag0 was used as the CCC index.

4.3 Data analysis and modeling procedures

4.3.1 Machine learning classifier

We used the MATLAB Classification Layer application for the machine learning model to compare the decision trees, random forests, support vector machines (SVMs), and neural nets. The results showed that even a single decision tree provided a correct answer rate exceeding 90%, which is comparable to the performance of other methods. Thus, we adopted a decision tree model to identify the most important features. The Gini diversity index was used as the splitting criterion.

Each of the sensor datasets in the three training sessions was subjected to machine learning, with target variables for classification (10% was used for data verification and cross-validation).

The Receiver Operating Characteristic (ROC) value calculated from the true-positive and false-positive rates exceeded 0.98, which was sufficient for classification accuracy. The ROC curves are presented in the Supplementary Appendix. The final number of branching nodes was 191.

The most important features for classifying joint activities were the right-hand rotation unused by the host and the left-hand rotation unused by the client. The features of the model were similar to those of the preliminary experiments, which classified conformity and competition in the countdown phase, thus suggesting that joint action in the subconscious movement can be classified using the categorization model based on the motion phase.

Using MATLAB's trained predict function, we applied the classification model to Sessions 2, 6, and 7 as test sessions using the 12 feature variables. The classification probability results for each observation were the activity indices of cooperation, conformity, and competition.

4.3.2 Angular transformation and index calculation

These probability values were angularly transformed using $\arcsin(\sqrt{\text{probability}}) \times 180/\pi$. We adjusted for a probability of 0 by setting $\arcsin(\sqrt{0.0833}) \times 180/\pi$ and a probability of 1 by setting $\arcsin(\sqrt{1-0.0833}) \times 180/\pi$. These corrections were performed based on the usual adjustment $(1/4N)$ for angular transformations. Thus, the minimum and maximum possible values were 16.54 and 78.69, respectively. These values were averaged for each trial and used as cooperation, conformity, and competition indices for the pair activity.

The term $1/4N$ used in the angular transformation refers to the usual adjustment for proportions, where N is the number of response options. After applying this correction, the angular transformation of the minimum and maximum possible proportions (0 and 1) results in values of 16.54 and 78.69, respectively. These are dimensionless values resulting from the arcsine transformation of relative proportions; therefore, no physical units are associated with them.

4.3.3 Linear mixed models

Because the human condition comprised data that switched from the client to the host for comparison with the bot condition, we analyzed the condition effects via multilevel analysis. For the linear mixed model, paired groups were specified as random-effect factors after centralization was performed, in which the mean value of each paired group was subtracted from each indicator.

As the Akaike Information Criterion (AIC)s of each indicator's random intercept and random slope models were similar or lower for the random slope model, we report the results for the random slope model here. Considering the few people in the random variable and a p-value that is likely to be high, we report the results of the robust model obtained via the log-likelihood ratio test. Owing to the low overall variance, we report the fixed-factor effects of the mixed model, as well as the results of the test using marginal mean estimation (in contrast to the human condition set to 1 and the bot condition set to 0).

5 Results

5.1 Preliminary experiment

A random forest model was applied to 21 features selected during the training sessions. The results showed that the confusion matrix between the model predictions and observed data was 88%, and the F1 score was 0.8925 (precision = 0.8066; recall = 0.9988).

Table 3 shows the top features selected by the decision tree classifier in the preliminary experiment. Features with importance ≥ 0.10 are listed individually, and all others are summarized in a single row.

The most important features for classification were the position and rotation of the left and right controllers, followed by the position and rotation of the head-mounted display (HMD), and finally, the gaze and pupillary reflexes. The higher importance of host-side

TABLE 3 Feature importance scores from the decision tree classifier (preliminary experiment). Features with importance < 0.10 are summarized in one row.

Importance	Feature name(s)
0.16	Host right-hand rotation
0.15	Client left-hand rotation
0.14	Host left-hand position
0.10	Host right-hand position
<0.10	Host left-hand rotation, Client head position, Host head position
	Host head rotation, Client left-eye pupil size, Client right-eye
	pupil size, Client right-hand position, Client gaze direction
	Host right-eye pupil size, Host gaze (x, y, z), Client left-eye
	pupil size, Client head rotation, Client gaze (x, y, z)

features was probably due to a slight delay in data transmission from the client side. This finding suggests that conformity or competition can be predicted by the twisting motion of the hands of a person who does not touch the button. Therefore, using the normalized variables of the host's right-hand rotation and its counterpart, i.e., the client's left-hand rotation, we calculated the measure of synchrony for each of the four pairs, and the cross-correlation coefficient (CCC) was calculated for each of the four pairs as a measure of synchrony.

Using this conformity or competition classification model, we analyzed how the ratio of conformity or competition status changed during the countdown phase in the joint Simon session. We discovered that the occurrence probability of a category classified as competition increased every second in Groups 3 and 4, whereas it decreased in Groups 1 and 2. The CCCs for each pair of groups in Groups 1–4 were 0.8872, 0.9998, 0.8581, and 0.8436, respectively. These results indicate that the synchrony index tended to be higher in Groups 3 and 4, whose competitive activity was higher than that of Groups 1 and 2.

6 Discussion

The preliminary experiments showed that sensor data from the countdown phase, which had less motion, can be used to identify the differences between conformity and competition training sessions. Hand and head rotations contributed more significantly than position and gaze direction. The activity during the countdown phase in the joint Simon session showed two patterns: one in which the ratio of competitive activities increased during the countdown phase, and another in which it decreased, with the former characterized by greater synchrony. This result contradicts the prediction that synchrony occurs in conformity activities. Therefore, in the main experiment, we added a joint Simon task as a "cooperation" target for training and created three categories: cooperation, conformity, and competition.

The bot conditions used in the main experiment were created by monitoring the behavior during the motion phase. Therefore, although the countdown phase was involved in the preliminary experiment, a classification category was created in the main experiment using the motion phase. The preliminary experiments showed that hands that were not used for button touching were more important for classification and that they gradually increased or decreased during the countdown up to the motion phase. Based on these results, we expect the features of the classification model using the countdown phase to appear in the classification model using the motion phase.

6.1 Main experiment

6.1.1 Decision tree

Details of machine learning, angular transformations, and statistical models are described in Section 4.3.

The ROC value calculated from the true-positive and false-negative rates exceeded 0.98, which was sufficient for the classification accuracy. The confusion matrix and ROC curves are presented in the Supplementary Appendix. The final number of branching nodes was 191. The classification criteria for the top six branches of the decision tree are shown in Figure 4.

The confusion matrix is presented in Figure 5.

6.1.2 Effects of human or bot conditions

Avatar behavior varied systematically between conditions. In the human condition, avatars reflected participants' real-time movements via the VR tracking system, providing natural responsiveness to each participant's actions. In the bot condition, avatars replayed pre-recorded human movements from earlier experimental sessions, with no adaptive responses to participant behavior. All bot avatar motions were pre-recorded in all bot sessions (Sessions 6 and 7), and there was no dynamic adjustment to participant behavior.

Details of machine learning, angular transformations, and statistical models are described in Section 4.3.

These three activity indices (cooperation, conformity, and competition) exhibit mutually constrained relationships. The correlation between concordance and competition was uncorrelated in both conditions, whereas the correlation between cooperation and competition indicators was $r = -.5432$ ($p = 0.0198$) in the human condition and $-.5078$ ($p = 0.0314$) in the bot condition. The correlation between the cooperation and conformity indices was $r = -.7821$ ($p < 0.001$) in the human

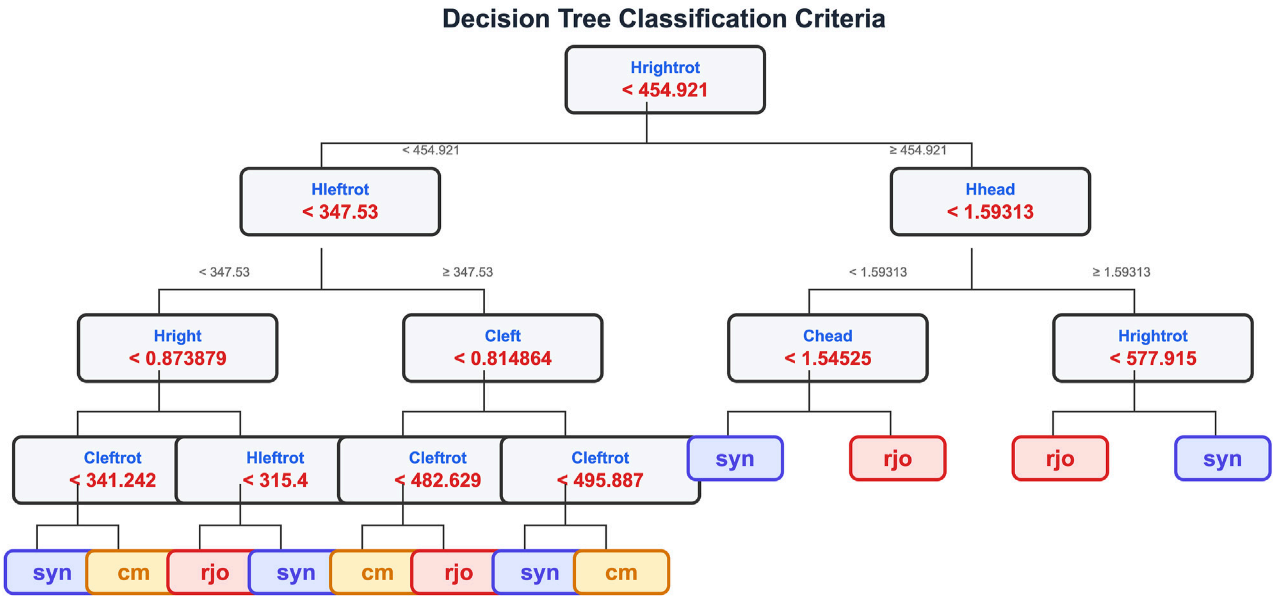


FIGURE 4
The classification criteria for the decision tree. Note: Branching conditions up to the third level are shown in the Supplementary Appendix. The decision tree shows branching conditions based on VR sensor measurements. Numerical values represent threshold values for splitting nodes. Each condition evaluates whether the sensor measurement is below (<) the specified threshold, with left branches representing true conditions and right branches representing false conditions. Terminal nodes indicate the final classification: cm (competition), rjo (joint simon), syn (conformity). Variable abbreviations: Hrightrot (Host right-hand rotation), Hleftrot (Host left-hand rotation), Hhead (Host head position), Cleftrot (Client left-hand rotation), Chead (Client head position), Hright (Host right-hand position), Cleft (Client left-hand position).

condition and $r = -.6061$ ($p = 0.0077$) in the bot condition, both of which were high. Therefore, to examine the effect of the conditions on the three activities, we examined each dependent variable individually. Figure 6 shows the angular-transformed mean values for each condition before centralization by the group mean.

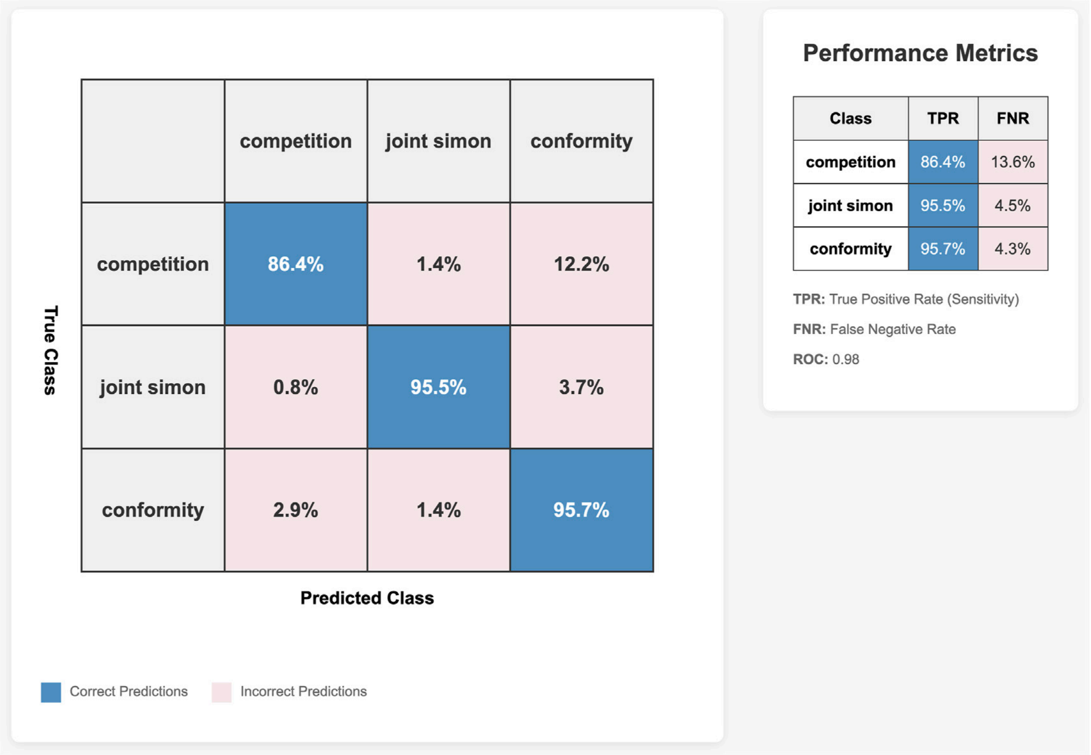
Figure 6 shows the angular-transformed mean values for cooperation, conformity, and competition activity indices under human and bot conditions. Cooperation activity was significantly higher in the human condition, whereas conformity activity was significantly higher in the bot condition.

An analysis of the human or bot-condition effect with the cooperation indicator as the dependent variable revealed that the estimated value of 7.37 ($SE = 3.25$) was significant ($t(9) = 2.27$, $p = 0.0495$). Similarly, a comparison of the mean estimate of the neighborhood with the human and bot conditions of 1 and 0, respectively, was significant ($z = 2.22$; Holm's $p = 0.0262$). As shown in Figure 6, the ratio of cooperation activity in the human condition was higher than that in the bot condition. An examination of the human- or bot-condition effect with the conformity activity as the dependent variable showed that the estimated value of -6.56 ($SE = 2.24$) was significant ($t(9) = 2.93$, $p = 0.0167$) and that the difference in the marginal mean estimate was substantial ($z = 2.93$; Holm's $p = 0.0034$), i.e., statistically significant. As shown in Figure 6, the conformity activity in the bot condition was higher than that in the human condition. An examination of the human-/bot-condition effect with competition activity as the dependent variable showed that the estimated value for the condition effect was insignificant, i.e., -1.43 ($SE = 2.28$). No difference in competitive activity was observed; however, cooperation activity was significantly greater in the human condition, whereas conformity activity was significantly greater in the bot condition.

An analysis of the human- and bot-condition effects with the JSE as the dependent variable revealed that the estimated value of 0.0033 ($SE = 0.009$) was insignificant. Meanwhile, an examination of the human- and bot-condition effects with the percentage of correct responses to the joint Simon task as the dependent variable found that the estimated value of -0.0111 ($SE = 0.0043$) was significant ($t(8.55) = 2.61$, $p = 0.0294$). Although a trend toward a higher percentage of correct responses was observed in the bot condition, a test of the difference between the marginal estimates showed $z = 1.03$ ($p = 0.0535$), which was not significant. The correct response rates are presented in Figure 7. The variance in the percentage of correct responses was higher in the human condition, whereas that in the bot condition was minimal. This is presumably because the bots consistently provided correct answers to the questions. In the bot condition, bots replayed pre-recorded human movements taken from correct trials, resulting in consistently correct responses for every trial.

Figure 7 presents the percent correct responses for the joint Simon task in each condition using a raincloud plot, which shows the distributions and box plots for each condition.

An analysis of the effect of the human or bot conditions on the synchrony index as the dependent variable showed the estimated value was 0.274 ($SE = 0.0035$), which was significant ($t(7.73) = 3.65$, $p < 0.001$). Additionally, the difference in the marginal estimates was significant ($z = 6.68$, $p < 0.001$). The results for each group are illustrated in Figure 8, where lower means and higher variances for synchrony were indicated under the bot condition. The lack of synchrony with the bot may have



True Class	competition	joint simon	conformity
competition	86.4%	1.4%	12.2%
joint simon	0.8%	95.5%	3.7%
conformity	2.9%	1.4%	95.7%

Predicted Class

🟦 Correct Predictions 🟥 Incorrect Predictions

Performance Metrics

Class	TPR	FNR
competition	86.4%	13.6%
joint simon	95.5%	4.5%
conformity	95.7%	4.3%

TPR: True Positive Rate (Sensitivity)

FNR: False Negative Rate

ROC: 0.98

FIGURE 5
Confusion matrix of predicted target and actual values from cross-validation. Note: Cross-validation was performed using 10% of the observed values of Session 3, 4, and 5 as test data. True-positive and true-negative rates are shown on the right. The ROC value calculated from these values was 0.98.

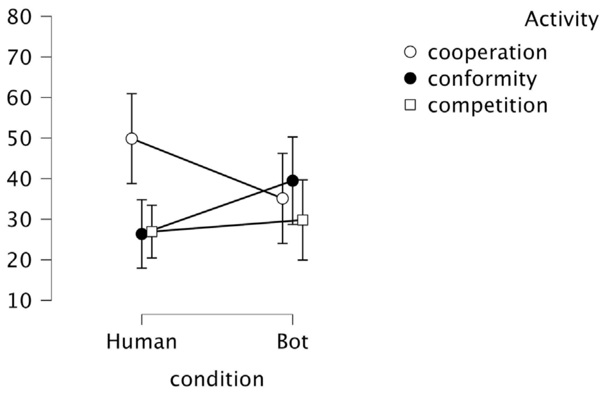


FIGURE 6
Effect of the human and bot condition on paired activity. Note: The vertical axis represents the mean classification probability after angular transformation, with values ranging from a minimum of 16.54 to a maximum of 78.69. The error bars indicate 95% confidence intervals.

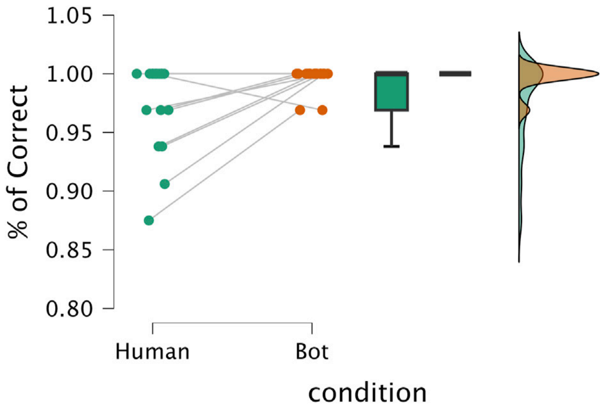


FIGURE 7
Correct response rates for each condition. Note: The vertical axis represents the correct response rate, and the error bars indicate 95% confidence intervals. The vertical axis represents the average percentage of correct responses during the human and bot sessions. Box plots and distributions are shown on the right. Green and red indicate human and bot conditions, respectively.

caused this difference, depending on whether the pair was aware or unaware of the bot. However, the difference in the mean bot awareness (0,1) was insignificant. Next, we performed mediation analysis based on the bot condition to determine whether bot awareness was a mediating variable.

6.1.3 Mediation analysis

Mediation analysis was conducted separately to investigate the effects of activity on the joint Simon task performance under the human and bot conditions. The variables considered were the three

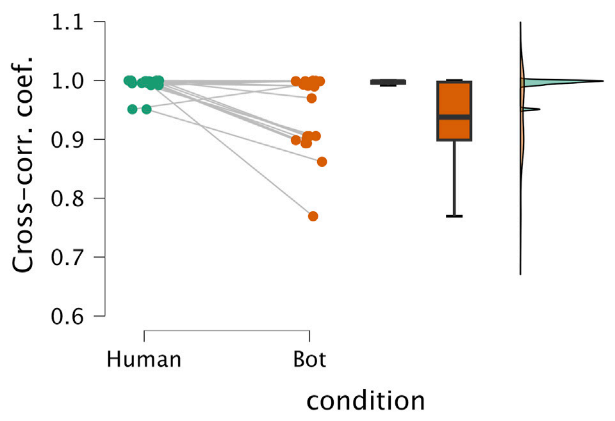


FIGURE 8
Raincloud plot of synchrony. Note: Green and red indicate human and bot conditions, respectively. The vertical axis represents the synchrony index for human and bot sessions. The maximum value was set to 1. Box plots and distributions are shown on the right. Green and red indicate human and bot conditions, respectively.

activities and synchrony indices as predictor variables, bot cognition as a mediating variable, and the JSE and correct response rate as the outcome variables. A 32-trial average was considered for the activity and synchrony indices to align with the correct response rate and sample size.

Owing to the high correlation between the three activity indicators that served as predictor variables, we performed principal component analysis as a standard procedure to avoid multiple linearities. Two principal components were extracted when eigenvalues greater than 1 were specified. The factor-loading matrix without rotation is presented in the Supplementary Appendix. Because the first principal component separated cooperation from other activities, we named the factor score of the first principal component the collaboration factor, i.e., collabH (as shown in Figure 9) and as collabB (as shown in Figure 10) for the bot condition. As the second principal component distinguished between competition and conformity, the score for the second principal component factor was named the competition factor, as indicated by competeH and competeB in Figures 9, 10, respectively. The first principal component was converted from negative to positive values, with higher values indicating greater cooperation. The outcome variables, JSE, and correct response rate were standardized and entered, and the regression coefficients were reported as standardized coefficients.

The path coefficients from the independent variables (two activity factors and synchrony) to the dependent variable (correct response rate) under the human condition are shown in Figure 9. The significant paths revealed that the collaboration factor increased the correct response rate, whereas the competition factor decreased the correct response rate. The statistics for each path are presented in Table 2 of the Supplementary Appendix. No effect on the JSE was observed, and bot cognition was not shown to be a mediating variable. The total R^2 values for the paths to the JSE, correct response rate, and bot cognition were 0.06, 0.28, and 0.11, respectively.

The path coefficients for the same variables under the bot condition are shown in Figure 10. The total R^2 values for the paths to the JSE, correct response rate, and bot cognition were 0.37, 0.10, and 0.35, respectively. As a key pathway, collaboration factors had a substantial total effect on enhancing bot cognition and reducing the JSE ($z = -3.17$, $p = 0.0015$).

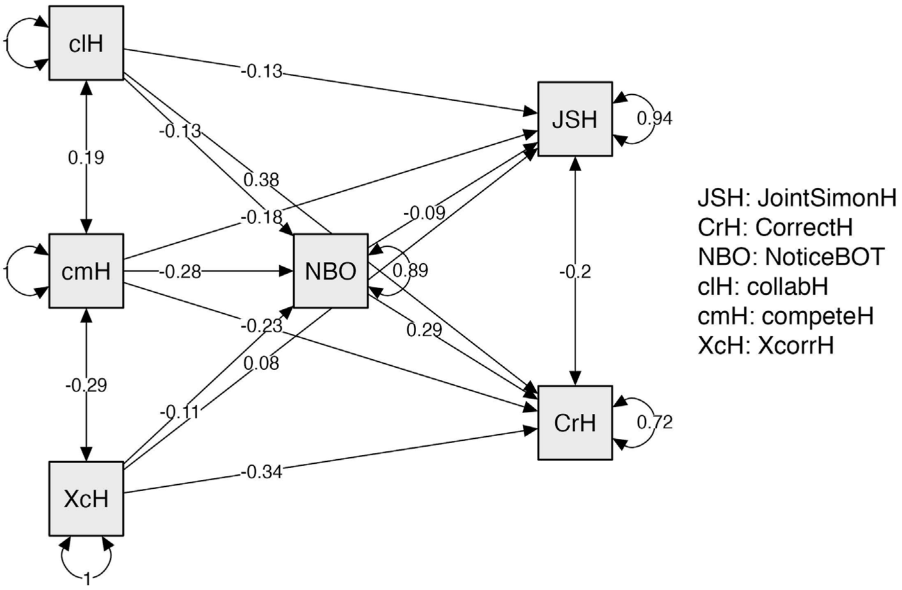


FIGURE 9
Path plot for the human condition. Note: The vertical axis represents standardized values of the dependent variables (JSE and correct response rate), while the horizontal axis represents the predictor variables (collaboration and competition factors). The numbers presented for each path are standardized path coefficients, where collabH and competeH denote the first and second principal components, respectively. Bot is a dummy variable with yes = 1 and no = 0. JSE and correct response rates are standardized.

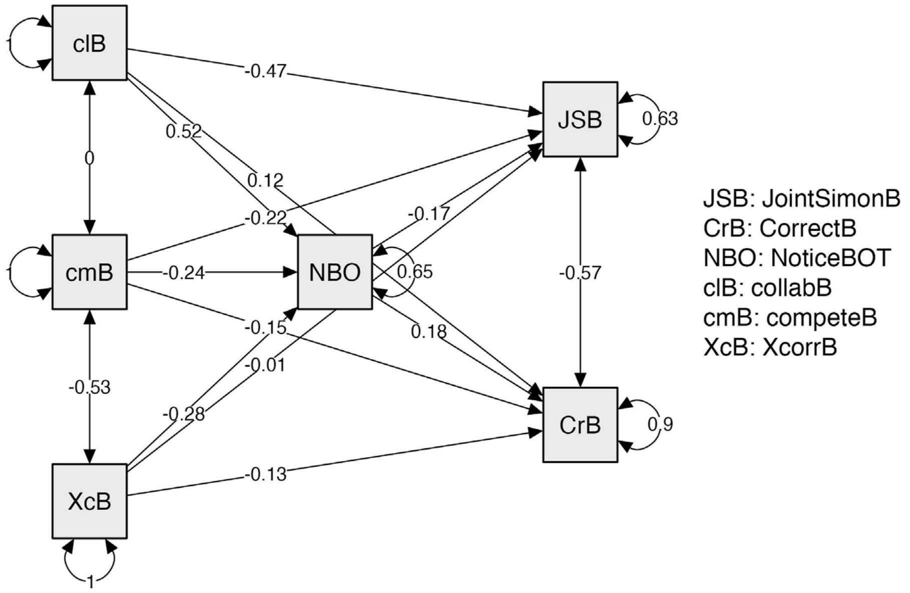


FIGURE 10
Path plot for the bot condition. Note: The vertical axis represents standardized values of the dependent variables (JSE and correct response rate), while the horizontal axis represents the predictor variables (collaboration and competition factors). collabX and competeX denote PCA components; bot cognition is a binary dummy; all dependent variables are standardized. The numbers shown for each path are standardized path coefficients; collabB and competeB indicate the first and second principal components, respectively.

To further examine the robustness of the mediation effect, we replaced the binary "Bot-Notice" variable with the continuous Social Presence index described in the methods section. However, when using Social Presence as a mediator, we did not observe any significant indirect effects on either the JSE or correct response rate in either condition.

The path analysis results for both conditions are presented in Figures 9, 10 (see Figures 9, 10). These figures illustrate the standardized path coefficients and demonstrate the differential effects of collaboration and competition factors on task performance under human and bot conditions.

The total R^2 values for the paths to the JSE, correct response rate, and bot cognition were 0.06, 0.28, and 0.11, respectively, for the human condition. For the bot condition, the total R^2 values were 0.37, 0.10, and 0.35, respectively. As a significant path, the collaboration factor substantially increased bot cognition and weakened the JSE as a total effect ($z = -3.17$, $p = 0.0015$).

We confirmed that an avatar's appearance did not affect the JSE or bot cognition. Specifically, avatar differences under the avatar condition (Sessions 6 and 7) were compared based on the classification probability as repeated factors. The results of an ANOVA indicated that the main effect of the avatar condition and the interaction effect between the avatar condition and classification were insignificant. Moreover, no significant interaction effects involving the bot avatar appearance were indicated.

In the preliminary experiment, a two-category model comprising competition and conformity was established initially. However, because the participants showed higher synchrony in competitive activities, the model was refined into a three-category classification comprising cooperation, conformity, and competition in the main experiment. The results of the main experiment revealed the feasibility of classifying joint activities based on subtle movements during Phase 3, i.e., when the participants were in motion, and in Phase 1, i.e., when the participants remained still. These results corroborate the predictions of the preliminary experiment, which identified the potential for preparing joint activities during the countdown phase. Furthermore, the results above suggest that the classification model and synchrony index used in this study were valid. A notable finding was the consistent selection of similar features for the synchrony index, which emerged as the most crucial feature for classification in both the preliminary and main experiments. This feature, which is a rotation of the unused hand, would not be readily observed by oneself or others, which suggests that behavioral synchronization phenomena appear as unconscious responses.

In the main experiment, we hypothesized that synchrony and cooperative activity under the bot condition would decrease compared with the human condition. A linear mixed model was used to analyze the effects of human and bot conditions on joint activities and synchrony indices. The results revealed a higher ratio of cooperative activity in the human condition and a high ratio of conformity in the bot condition. This finding is consistent with previous research indicating that humans tend to conform more to non-human agents when the agent's behavior is predictable or lacks social cues. The synchrony index was significantly lower in the bot condition, indicating reduced interpersonal synchrony with bot avatars. This suggests that while bots can elicit conformity, they may not facilitate the same level of behavioral coordination as human partners.

The mediation analysis further revealed that cooperation factors had a substantial total effect on enhancing bot cognition and reducing the JSE. However, bot cognition was not shown to be a mediating variable for the correct response rate or JSE. These findings highlight the complexity of human interaction with a non-responsive avatar and suggest that while participants may recognize non-responsive avatars as joint-action task partners, this recognition does not necessarily translate into improved task performance or increased synchrony.

These findings contribute to discussions on the mechanisms underlying the Joint Simon Effect (JSE). Rather than supporting one theoretical account over the other, our results indicate that bodily synchrony engages both processes simultaneously: participants appeared to use the partner as a spatial reference point while also sharing aspects of task representation. This suggests that interpersonal synchrony can serve as a behavioral signature of these intertwined mechanisms, highlighting how spatial coding and co-representation may co-occur rather than operate in isolation.

7 Limitations and future directions

This study has several limitations. First, the sample size was relatively small, which may limit the generalizability of the findings. Future studies should include larger and more diverse participant groups to validate the classification model and synchrony index. Second, the non-responsive avatars in this study were based on replayed human motion data and did not exhibit adaptive or interactive behaviors. Incorporating more sophisticated AI-driven avatars that can respond dynamically to human actions may yield different results and provide deeper insights into human interaction with responsive avatars. Third, the experimental tasks were limited to a specific joint Simon paradigm in a controlled VR environment. Expanding the range of tasks and exploring real-world applications will be important for understanding the broader applicability of these methods.

In terms of social presence, our findings replicated those of Munnukka et al. (2022), in that the visual appearance of the avatar did not significantly affect perceived anthropomorphism or social presence. However, the social presence index derived from questionnaire items did not significantly mediate any of the observed effects. This may be due to the limited sample size, which constrained both the statistical power and the number of items included in the questionnaire. Specifically, we used only two items to assess perceived realism and human-likeness of the avatar. In ongoing studies, we are addressing this limitation by incorporating a more comprehensive set of items to better capture the multidimensional nature of social presence.

Future research should also investigate the neural and psychological mechanisms underlying interpersonal synchrony and joint task performance in VR, as well as the impact of different types of avatars on social interaction. By addressing these limitations and exploring new directions, future studies can further advance our understanding of human-machine interaction in virtual environments.

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: The data behind this analysis has been made publicly available at OPENICPSR and can be accessed at (https://www.openicpsr.org/openicpsr/project/178601/version/V1/view). The VR sensor log data are available from the corresponding author upon request.

Ethics statement

The studies involving humans were approved by the Ethics Committee of Kyoto University of Advanced Science (Project No. 22H07). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

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

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by JSPS KAKENHI Grant Number 21K02988 and 23K22343, 23K25727.

Acknowledgments

We thank all participants for their time and effort in this study. We also thank the members of the Center for Social and Psychological Research of Metaverse for their valuable discussions and feedback.

Conflict of interest

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

Generative AI statement

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

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2025.1623764/full#supplementary-material

References

Chen, N., Liu, X., Zhai, Y., and Hu, X. (2023). Development and validation of a robot social presence measurement dimension scale. *Sci. Rep.* 13, 1502. doi:10.1038/s41598-023-28561-8

Decety, J., Michalska, K. J., and Kinzler, K. D. (2011). The contribution of emotion and cognition to moral sensitivity: a neurodevelopmental study. *Cereb. Cortex* 22, 209–220. doi:10.1093/cercor/bhr111

Dolk, T., Hommel, B., Prinz, W., and Liepelt, R. (2014). The (not so) social simon effect: a referential coding account. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 1248–1260. doi:10.1037/a0031031

Fronda, G., and Balconi, M. (2022). The effects of prosocial and antisocial behaviors on brain-to-brain synchrony during cooperative tasks. *Soc. Neurosci.* 17, 1–12. doi:10.1080/17470919.2021.1970012

Guastello, S. J., Reiter, K., Malon, M., Timm, P., Shircel, A., and Shaline, J. (2023). Catastrophe theory for dynamical systems in psychology. *Nonlinear Dyn. Psychol. Life Sci.* 27, 1–24. doi:10.1891/NDP-2023-0001

Hao, Y., Li, X., and Zhang, Y. (2024). Group identity modulates brain-to-brain synchrony and cooperative decision-making. *Soc. Cognitive Affect. Neurosci.* 19, 1–12. doi:10.1093/scan/nsad123

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

Heider, F., and Simmel, M. (1944). An experimental study of apparent behavior. *Am. J. Psychol.* 57, 243–259. doi:10.2307/1416950

Liepelt, R., Stenzel, A., and Lappe, M. (2016). The role of the anterior cingulate cortex in the joint simon effect. *Front. Psychol.* 7, 1862. doi:10.3389/fpsyg.2016.01862

Miss, J., Pfeuffer, C. U., and Kunde, W. (2022). The joint simon effect depends on perceived agency, but not intentionality, of the alternative action. *Psychol. Res.* 86, 1–15. doi:10.1007/s00426-020-01460-8

Munnukka, J., Talvitie-Lamberg, K., and Maity, D. (2022). Anthropomorphism and social presence in human-virtual service assistant interactions: the role of dialog length and attitudes. *Comput. Hum. Behav.* 135, 107343. doi:10.1016/j.chb.2022.107343

Paladino, M. P., Mazzurega, M., Pavani, F., and Schubert, T. W. (2010). Synchronous multisensory stimulation blurs self-other boundaries. *Psychol. Sci.* 21, 1202–1207. doi:10.1177/0956797610379234

Rennung, M., and Göritz, A. S. (2016). Taking turns or not? Children's prosocial responsiveness in dyadic and triadic interactions. *J. Exp. Child Psychol.* 141, 299–309. doi:10.1016/j.jecp.2015.07.009

Sangati, E., Willemse, C., and Hunnius, S. (2021). The role of action prediction and inhibitory control in the joint simon effect. *Psychol. Res.* 85, 1001–1013. doi:10.1007/s00426-020-01305-4

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

Sekitani, J., and Murakami, H. (2022). Framework for comparing accuracy of time-series forecasting methods. In: *2022 International Congress on Advanced Applied Informatics (IIAI-AAI)*. Kanazawa, Japan

Sellaro, R., Dolk, T., Colzato, L. S., Liepelt, R., and Hommel, B. (2015). Referential coding does not rely on location features: evidence for a non-spatial joint simon effect. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 186–195. doi:10.1037/a0038548

Simon, J. R. (1969). Reactions toward the source of stimulation. *J. Exp. Psychol.* 81, 174–176. doi:10.1037/h0027448

Smykovskyi, O., Koval, V., and Kostiuk, T. (2024). Emotional contagion and interpersonal synchrony in virtual reality. *Front. Psychol.* 15, 1234567. doi:10.3389/fpsyg.2024.1234567

Sogemeier, D., Naujoks, F., Forster, Y., Krems, J. F., and Keinath, A. (2024). Exploring the interaction between anthropomorphism and performance on trust and acceptance in in-vehicle voice assistants. *Preprint. SSRN.* doi:10.2139/ssrn.4637565

Stenzel, A., and Liepelt, R. (2014). The moving rubber hand illusion revisited: comparing movements and visuotactile stimulation to induce illusory ownership. *Conscious. Cognition* 26, 117–132. doi:10.1016/j.concog.2014.02.003

Stenzel, A., Chinellato, E., Bou, M. A. T., del Pobil, A. 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

Stenzel, A., Dolk, T., Hommel, B., and Liepelt, R. (2016). The joint simon effect: a review and theoretical integration. *Front. Psychol.* 7, 975. doi:10.3389/fpsyg.2016.00975

Tsai, C. C., Kuo, W. J., Jing, J. T., Hung, D. L., and Tzeng, O. J. L. (2007). A common coding framework in self-other interaction: evidence from joint action task. *Exp. Brain Res.* 182, 41–50. doi:10.1007/s00221-007-0972-6

Tsai, C. C., Kuo, W. J., Hung, D. L., and Tzeng, O. J. L. (2008). Action co-representation is tuned to other humans. *J. Cognitive Neurosci.* 20, 2015–2024. doi:10.1162/jocn.2008.20144