The influence of perspective on VR job interview training

Fumitaka Ueda, Yuichiro Fujimoto*, Taishi Sawabe, Masayuki Kanbara and Hirokazu Kato

Graduate School of Science and Technology, Nara Institute of Science and Technology, Nara, Japan

Third-person perspectives in virtual reality (VR) based public speaking training enable trainees to objectively observe themselves through self-avatars, potentially enhancing their public speaking skills. Taking a job interview as a case study, this study investigates the influence of perspective on the training effects in VR public speaking training and explores the relationship between training effects and the sense of embodiment (SoE) and presence, as these concepts are central to virtual experiences. In the experiment, VR job interview training was conducted under three conditions: a first-person perspective (1PP), a typical third-person perspective from behind the avatar (Back), and a third-person perspective from the front of the avatar (Front). The results indicate that participants trained in the Front condition received higher evaluations from others in terms of verbal communication skills and the overall impression of the interview compared to those trained in the other conditions, highlighting the advantages of training while observing a self-avatar. Furthermore, it was confirmed that training effects correlated with the subcomponents of SoE and presence, suggesting that these trends may vary depending on perspective.

1 Introduction

Public speaking is delivering a speech before an audience, such as interviews and presentations (Nikitina, 2011). Public speaking skills are crucial for social evaluation and can significantly impact an individual's life. Interviews, especially those conducted during job hunting and higher education, can have a direct impact on an individual's life. To efficiently enhance public speaking skills, it is necessary to practice in situations that closely mirror real-life scenarios (Smith and Frymier, 2006). However, training in such conditions can be challenging due to the difficulty in securing an audience and an appropriate practice place. In recent years, virtual reality (VR) public speaking training has garnered substantial attention.

In VR, the first-person perspective (1PP) is commonly used, particularly in gameplay scenarios. However, the third-person perspective (3PP) is also employed. This perspective allows users to observe themselves from a third-party standpoint through an avatar, which is a virtual representation of the user. Thus, 3PP is effective in enhancing spatial awareness (Gorisse et al., 2017). Bodyswaps in United Kingdom provides training from the front 3PP in its VR-based interview training module[1]. Moreover, prior research shows that using 3PP

---

1 https://bodyswaps.co/

in VR public speaking reduces anxiety (Bellido Rivas et al., 2021) and facilitates objective self-evaluation when reflecting on one's presentation (Zhou et al., 2021). These results suggest the potential benefits of employing 3PP in VR public speaking training. On the other hand, although public speaking skills determine audience evaluations, no research has confirmed the effect of perspective on public speaking skills. In this study, we define changes in public speaking skills before and after training as training effects and investigate the influence of perspective on these effects.

Furthermore, the sense of embodiment (SoE) and presence are pivotal concepts in virtual experiences. The SoE refers to the sense that arises when certain body properties are processed as if they were one's own (Kilteni et al., 2012). SoE is intimately linked with human cognition, including emotional changes, environmental recognition, and bodily movements in VR (Osimo et al., 2015; Ogawa et al., 2019; Burin et al., 2019). On the other hand, presence, or the sense of presence, is the subjective experience of being in a virtual environment. It refers to the sensation of engagement that emerges between the individual and the virtual environment (Witmer and Singer, 1998). A greater sense of presence is anticipated to enhance learning and performance (Witmer and Singer, 1998; Kothgassner et al., 2012). Moreover, in the context of public speaking, where practice in a closely simulated real-life scenario is effective, increased presence may lead to more significant training effects. The question here is how SoE and presence relate to the training effects of VR public speaking training.

1.1 Virtual reality public speaking and job interview training

VR elicits responses in people similar to those in the physical world, making it a valuable tool for simulating complex and realistic situations and contexts (Diemer et al., 2015; Slater and Sanchez-Vives, 2016). In VR public speaking simulations, scenarios are highly customizable (Takac et al., 2019). This customization enables the creation of situations that would be challenging to replicate in the real world, providing flexibility in defining the complexity and context. Many studies on public speaking training have concentrated on cognitive aspects. For example, VR interventions have effectively reduced social anxiety and fear of public speaking (North et al., 1998; Anderson et al., 2013; Takac et al., 2019). VR interventions led to anxiety reduction comparable to traditional interventions, such as face-to-face therapy (Ebrahimi et al., 2019).

VR public speaking training can also enhance verbal and nonverbal skills (Chollet et al., 2015; Valls-Ratés et al., 2022). A meta-analysis revealed that VR training programs aimed at developing social skills may be more effective than alternative training programs, particularly for improving more complex social skills (Howard and Gutworth, 2020).

In addition to exposure therapy training, systems have been developed to encourage users to review and correct their performance using feedback based on physical and oral information. Hoque et al. (2013) proposed an interview training system that visualizes data on smile rate and prosody, allowing trainees to review non-verbal behavior data while watching

recordings. There is also research on providing real-time visual feedback to trainees (El-Yamri et al., 2019), as well as evaluating its acceptability and validity (Palmas et al., 2021; Tanaka et al., 2017).

Other studies have investigated the effects of modifying user perspective, self-avatars, and virtual audience characteristics in training. The appearance of self-avatars and the behavior of virtual audiences have been found to reduce anxiety and stress (Aymerich-Franch et al., 2014; Thakkar et al., 2022). Delivering a speech from the third-person perspective behind the avatar reduced state anxiety (Bellido Rivas et al., 2021). In contrast, the correlation between the realism of the virtual audience's appearance and anxiety was low (Kwon et al., 2013).

As mentioned earlier, many studies have primarily focused on anxiety. While reducing anxiety is a beneficial outcome of training, it does not necessarily lead to better audience evaluation (King and Finn, 2017). Therefore, it is necessary to investigate the impact on speaking skills, which are directly connected to audience evaluation. This study focuses on examining the relationship between perspective and changes in both verbal and nonverbal skills of job interview.

1.2 Influence of perspective on performance in VR

Numerous studies have explored the effects of different perspectives on dynamic task performance and motor accuracy in VR (Salamin et al., 2006; Bhandari and O'Neill, 2020). Previous research indicates that 1PP is suitable for situations requiring precise interaction (Gorisse et al., 2017; Medeiros et al., 2018). However, 1PP has the disadvantage of providing limited information due to its restricted field of view (Wang et al., 2022). By contrast, in 3PP, the avatar and camera are positioned farther apart, providing a wider field of view. This improves spatial awareness (Salamin et al., 2006; Gorisse et al., 2017; Cmentowski et al., 2019).

Another advantage of 3PP is that it provides an objective view of one's avatar. People tend to overestimate their own abilities and attributes, as seen in the illusory superiority (Hoorens, 1993), or underestimate them, as in the below-average effect (Kruger, 1999), making it difficult to evaluate themselves objectively. However, by using 3PP in VR, objectivity may be facilitated. Reflecting on one's own presentation from the perspective of an audience resulted in more objective self-evaluation, especially for those with low confidence in their speaking skills (Zhou et al., 2021). This underscores the importance of 3PP in VR training.

1.3 Sense of embodiment in VR

In VR, SoE refers to the sensation experienced within one's avatar, an alter ego in a virtual environment. Since avatars are a fundamental component of most VR applications, SoE is an essential aspect of the VR experience. SoE consists of three subcomponents: the sense of body ownership, the sense of agency, and the sense of self-location.

Sense of body ownership (SoBO) is the perception of being the subject of an action, experiencing movement, or feeling specific sensations (Tsakiris et al., 2007). It is also defined as the awareness of a body as one's own (Roth and Latoschik, 2020).

Sense of agency (SoA) refers to the feeling that one is the cause or generator of an action (Gallagher, 2000; Tsakiris et al., 2007). While SoBO occurs during both passive experiences and voluntary actions, SoA is particularly influenced by voluntary actions (Tsakiris et al., 2006). SoA arises when one's movements are accurately replicated in real-time (Jeunet et al., 2018).

Sense of self-location (SoSL) is the spatial experience of perceiving oneself as being located at the avatar's position (Kilteni et al., 2012). SoSL is strongly influenced by the perspective position.

SoE is intimately linked to cognitive and behavioral changes in VR. For instance, SoBO and SoA affect body movements and motor performance (Newport et al., 2010; Zopf et al., 2011; Kilteni and Ehrsson, 2017; Matsumiya, 2021). Burin et al. (2019) report that when the avatar's body is perceived as one's own, in other words, when SoBO is experienced, there is a greater effect on body movements.

SoE is not limited to physical actions; it can also influence emotions. When avatar movements are synchronized with the user's movements, moods improve, and emotions are felt more positively compared to when avatar movements are not synchronized with the user's actions (Osimo et al., 2015; Jun et al., 2018). Since SoBO and SoA are more strongly elicited during synchronization, SoE may contribute to emotional changes.

Concerning VR training, although no studies have examined the relationship between SoE and training effects, it has been suggested that SoE may indirectly influence these effects. Koek and Chen (2023) found that participants who interacted with a virtual agent while embodied in an avatar resembling themselves exhibited positive changes in self-esteem. This result may be related to the fact that the closer the avatar's appearance matches one's own, the stronger the perception of SoBO (Waltemate et al., 2018; Suk and Laine, 2023). Although self-esteem and interview confidence are not identical, given that confidence affects interview success (Tay et al., 2006), there could be a relationship between SoE and training effects. The second objective of this study is to investigate the relationship between SoE and training effects.

1.4 Presence and VR public speaking and interview

Presence in VR is defined as the subjective experience of being in a virtual environment (Witmer and Singer, 1998). Presence, along with SoE, is among the most studied elements in VR applications (Poeschl, 2017). Since presence arises between the user and the virtual environment, it differs from SoSL, which is the sensation between the user and an avatar (Slater and Wilbur, 1997; Kilteni et al., 2012). It is hypothesized that presence and its influencing factors can enhance learning effectiveness and performance (Witmer and Singer, 1998; Kothgassner et al., 2012).

In virtual environments characterized by high visual realism, subjective presence tends to be higher, potentially inducing more stress (Slater et al., 2009). Several studies have explored the relationship between anxiety and presence in VR training. Girondini et al. (2023) found a positive correlation between presence and anxiety during VR speech. In contrast, research by Kwon et al. (2013) indicated that anxiety during VR interviews was unaffected by presence.

The correlation between presence and anxiety in VR systems has been studied, but the findings remain inconsistent. By contrast, the relationship between presence and training effects has not been investigated. Given that practice under realistic conditions is generally more effective (Smith and Frymier, 2006), a stronger presence may result in greater training effects. Therefore, this study examines the relationship between presence and training effects of job interview training.

1.5 Research questions

The research questions of the present study are as follows:

RQ1 How does perspective during VR job interview training influence training effects?
RQ2 What is the relationship between SoE and training effects, and between presence and training effects?

To address these questions, we developed a VR job interview training system simulating a job interview and conducted a between-subjects experiment. We selected the job interview as the public speaking task because public speaking skills closely influence interview results and are a major concern for many students. In the experiment, VR job interview training was conducted over 5 days under three perspective conditions.

1.6 Hypotheses

1.6.1 Training effects by perspective condition

In the Back and Front conditions, participants can observe themselves from a third party's perspective, which may lead to a more relaxed training and potentially higher scores for Prosody, Response, and Overall (See Section 2.5.1 in detail). Furthermore, in the Front condition, participants can observe their body and facial movements in more detail through the avatar, which is expected to enhance Behavior scores. Consequently, we anticipate the following hypotheses regarding the influence of perspective on training effects:

H1 Improvements on *Prosody*, *Behavior* and *Response* scores will be higher in the Back and Front conditions compared to the 1PP condition.
H2 Improvements on *Behavior* score will be highest in the Front condition.

1.6.2 Relationship of SoE and presence to training effects

As discussed in Section 1.3, previous studies indicate that SoBO and SoA contribute to emotional changes and body movements. In the context of public speaking training, a strong SoBO and SoA may positively affect training effects. Meanwhile, it is expected that presence enhances learning and performance (Witmer and Singer, 1998; Kothgassner et al., 2012). Considering that practice in situations that closely resemble real-life scenarios is more effective (Smith and Frymier, 2006), we hypothesize that a stronger presence will lead to greater training effects. Consequently, we propose the following hypotheses:

FIGURE 1
A virtual office environment simulating a job interview. The user's avatar was positioned in a chair on one side of a table, while three interviewer agents were positioned on the opposite side.

H3 Training effects will have a positive correlation with SoBO and SoA.

H4 A positive correlation will be found between presence and training effects.

2 Methods

This section presents the experiment conducted using the VR job interview training system. The experiment was approved by the Ethics Committee of Nara Institute of Science and Technology and was conducted in accordance with the institutional ethical guidelines.

2.1 Overview

The experiment had two primary objectives: the first was to investigate how different perspectives during VR job interview training influences training effects, specifically changes in public speaking skills for job interviews. The second was to explore how SoE and presence relate to training effects. The experiment spanned 5 days: VR job interview training was conducted once on day 1, twice each on days 2 through 4, and once on day 5, totaling eight sessions. Additionally, face-to-face mock interviews, without the VR system, were conducted at the beginning of day 1 and the end of day 5. The difference in evaluations between the two mock interviews, conducted before and after the training, was defined as the training effect. The experimental results were analyzed based on these training effects.

2.2 Job interview training system design

2.2.1 Virtual environment

A virtual office environment simulating a job interview was created using Unity (see Figure 1). The user's avatar was seated in a chair on one side of a table, while three interviewer agents were positioned on the opposite side, engaging in a job interview simulation. In this training,

the middle agent asked questions to the user, and the user responded, mimicking a typical job interview scenario.

2.2.1.1 Interviewer agents

The interviewer agents were represented by virtual humans sourced from Greta (Pelachaud, 2017) and Microsoft Rocketbox (Gonzalez-Franco et al., 2020). These agents followed a 40-s idle animation, allowing for natural breathing movements and occasional subtle posture adjustments. The agent's spoken voice was created using VOICEVOX[2]. Furthermore, the agent's mouth movements were synchronized with his speech using uLipSync[3], a Unity asset designed for lip-syncing characters. The questions posed by the agent and their sequence remained consistent throughout the interaction. After the user responded to a question, the agent provided a brief response (e.g., "Thank you", "I understand.") and after a few seconds, proceeded to the next question. The timing of the user's response completion and the agent's speech was judged and controlled by the experimenter. The agent delivered only simple replies and did not adapt to the user's utterances. Based on the study by Mostajeran et al. (2020), which demonstrated that users felt sufficient realism with three agents, the number of interviewers was set to three. Only the middle agent speaks, as this is intended to mimic the practice in many Japanese job interviews where the lead interviewer primarily asks the questions.

2.2.1.2 Avatar creation

The user's avatar head, including the hair, was automatically generated from a photograph of the user's face using the Headshot Auto function in Character Creator 4. The avatar's body shape remained at the default setting, which represents a standard body shape.
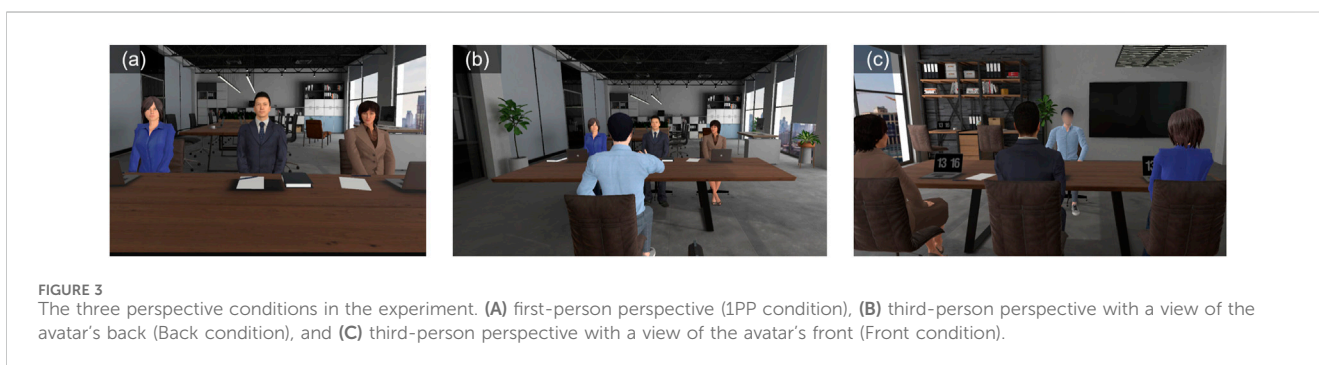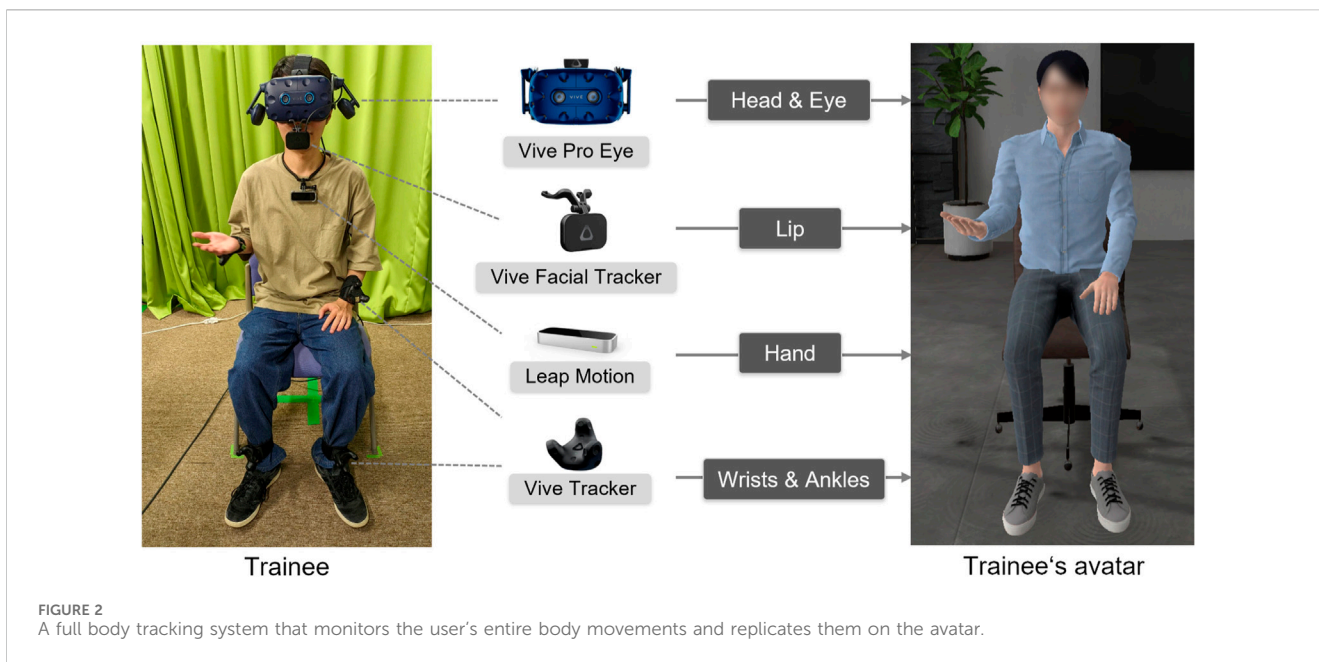
2.2.2 Full body tracking system

To induce SoE, particularly SoBO and SoA, synchronizing the movements of the avatar with those of the user's entire body is effective (Maister et al., 2015; Pyasik et al., 2022). Consequently, we implemented a full body tracking system that monitors the user's complete body movements and replicates them onto the avatar (see Figure 2). The HTC Vive Pro Eye, utilized as a head-mounted display (HMD), not only displays visuals but also supports eye tracking, allowing us to mirror the user's gaze and eyelid movements onto the avatar. The Vive Facial Tracker captures the user's facial expressions, specifically lip movements, while the Leap Motion detects hand movements. Additionally, the user wears four Vive Trackers, with two on each wrist and ankle. The SteamVR Base Stations track the position and orientation of the HMD and trackers, and the acquired data being used to control the avatar's body movements through Final IK, a Unity asset that supports the inverse kinematics system.

2.3 Conditions

There were three conditions, each with the following descriptions. Figure 3 illustrates an example of the perspective for

---

2   https://voicevox.hiroshiba.jp/

3   https://github.com/hecomi/uLipSync

FIGURE 2
A full body tracking system that monitors the user's entire body movements and replicates them on the avatar.

FIGURE 3
The three perspective conditions in the experiment. **(A)** first-person perspective (1PP condition), **(B)** third-person perspective with a view of the avatar's back (Back condition), and **(C)** third-person perspective with a view of the avatar's front (Front condition).

each condition. We included the Front condition as the third condition because public speaking, which involves audience evaluation, requires consideration of the audience's viewpoint.

1PP Condition: In this condition, the viewpoint is set to 1PP, which corresponds to the self-avatar's viewpoint.
Back Condition: This perspective is viewed from behind the avatar, slightly to the right. The camera moves in sync with the user's actual head movements.
Front Condition: This perspective is from the front view of the avatar and represents another type of third-person perspective. In contrast with the Back condition, the camera is positioned behind the interviewers, and its movements correspond to the user's movements.

2.4 Participants

22 graduate students (13 males, 9 females) participated in the experiment, with a mean age of 23.3 (SD = 1.43), ranging from 22 to 27 years. All interview training sessions and mock interviews were

conducted in Japanese. Among the participants, 21 were native Japanese speakers, and one non-native Japanese speaker communicated fluently in Japanese. Participants self-reported their experiences with job interviews and their confidence in job interview performance when applying for the experiment. They were assigned to three conditions: 1PP (n = 7), Back (n = 8), and Front (n = 7). Of the participants, 11 were in the process of job hunting, eight had already completed their job hunting, and three had no previous job hunting experience but intended to seek employment in the future. We have made adjustments to approximately equalize across groups, particularly with respect to job search experience, which is likely to have a significant impact on this study. Table 1 shows the characteristics of the participants in each condition.

2.5 Evaluation methods

2.5.1 Evaluation of public speaking skills by others
To assess public speaking skills for job interviews, we developed a questionnaire comprising 16 items (see Table 2). As far as we

TABLE 1 Participant Characteristics. In Experience of Job Hunting, A: in the process of job hunting. B: had already completed their job hunting. C: had no job hunting experience (but intended to seek employment in the future).

	1PP	Back	Front
Age	M = 24.4, SD = 1.4	M = 24.8, SD = 1.7	M = 24.6, SD = 1.6
Biological Sex	Male: 5, Female: 2	Male: 4, Female: 4	Male: 4, Female: 3
Experience of Job Hunting	A: 3, B: 3, C: 1	A: 4, B: 3, C: 1	A: 4, B: 2, C: 1

TABLE 2 The list of 16 items of the questionnaire used to measure public speaking skills required for job interviews.

Category	Item
Prosody	• Speed: The interviewee spoke at an appropriate speed that was easy to understand • Fluency: The interviewee spoke fluently • Volume: The interviewee spoke at an adequate volume of voice • Intonation: The interviewee's speech sounded enthusiastic and confident • Politeness: The interviewee's speech sounded polite
Behavior	• Smile: The interviewee maintained a smile throughout the interview • Posture: The interviewee always sat with a straight back and well postured • Gesture: The interviewee used effective gestures relevant to the content of the speech • Attitude: The interviewee was always focused and calm during the interview • Eye contact: The interviewee maintained appropriate eye contact with the interviewers
Response	• Appropriateness: The interviewee was able to answer the questions appropriately • Expression: The interviewee expressed oneself in one's own words, not in common words • Honorific: The interviewee used appropriate honorifics • Logicalness: The interviewee used logical and easily understandable expressions • Conciseness: The flow of the speech was natural and concise
Overall	• I think the interview was successful, from an overall viewpoint

know, some studies have used overall impressions to evaluate interview performance, but there is currently no widely accepted scale for assessing non-verbal and verbal public speaking skills for job interviews. Consequently, we constructed this questionnaire based on previous literature (Fydrich et al., 1998; Naim et al., 2015; Nikitina, 2011). The questionnaire is categorized into three skill sets: Prosody, Behavior, and Response. Prosody and Behavior pertain to non-verbal skills, whereas Response focuses on verbal skills. Prosody pertains to tone of voice, Behavior to observable behavior during the interview, and Response to the content of answers provided. All three skill sets include five items each. In addition, there is an Overall scale for evaluating the general impression of the interview. A 7-point Likert scale (1: Strongly disagree, 7: Strongly agree) was employed for all items.

Two graduate students (one male and one female) used this questionnaire to evaluate public speaking skills of the participants, and they were not involved in the experiment. The evaluators watched recordings of mock interviews and rated them. Specifically, for each participant, each item was rated for each of the videos of the interview before the training (one video) and the interview after the training (one video). The study conditions assigned to participants were concealed from the evaluators. Additionally, the video presentation

order was randomized, ensuring that the evaluators were unaware of whether the videos were recorded before or after the training. The average ratings provided by the two evaluators were used as to assess each mock interview. If there was a discrepancy of 3 or more points between the ratings of the two evaluators for a specific item in the same interview, an additional evaluator rated the item. An additional evaluator, who was one of the authors of the paper, evaluated it under similar conditions to the other evaluators. Ultimately, the differences in evaluation before and after training were compared across the three conditions. The difference in the scores for prosody, behavior, response, and overall (i.e., the value obtained by subtracting the pre-training score from the post-training score) is defined as $Prosody_d$, $Behavior_d$, $Response_d$, and $Overall_d$, respectively. Before the evaluation tasks, evaluators were instructed to thoroughly review the evaluation criteria and standards for each score and to remember key focus areas before reviewing the videos.

2.5.2 Sense of embodiment

We employed the Virtual Embodiment Questionnaire (VEQ) (Roth and Latoschik, 2020) to measure SoBO and SoA. The VEQ consists of four questions for each aspect. The scores for SoBO and SoA were calculated by averaging the responses to these four items. The SoSL score was determined by averaging the reseponses to the two questions as described in a previous study (Piryankova et al., 2014). Participants completed these questionnaires using a 7-point Likert scale immediately after each VR training session. We collected eight sets of responses per participant and used the average of these responses for analysis.

2.5.3 Presence

To measure presence, we used the Igroup Presence Questionnaire (IPQ) (Schubert et al., 2001). The IPQ was administered concurrently with SoE assessment, and the average score from eight sessions was analyzed. The IPQ consists of three subscales and one additional general item that is not part of any subscale, as outlined below:

Spatial Presence (SP) Reflects the sense of being physically present in the virtual environment.
Involvement (Inv) Measures the level of attention devoted to the virtual environment and the degree of involvement experienced.
Experienced Realism (Real) Evaluates the subjective experience of realism within the virtual environment.
General Presence (GP) Gauges the general "sense of being there".

2.6 Procedure

Before the experiment, participants were asked to prepare their responses to the interview questions in approximately 1 hour. This

TABLE 3 The list of interview questions used in the experiment.

| Category	Sentence
Self introduction	• Please introduce yourself
School activities	Can you give us an overview of what you focused on during your student days? • What motivated you to engage in the activity? • What difficulties did you encounter in the activity? • What did you learn from the activity?
Research topic	• Please give us an overview of your current or past research • What difficulties did you encounter in your research? • Why did you choose the research topic? • How will your research impact society?
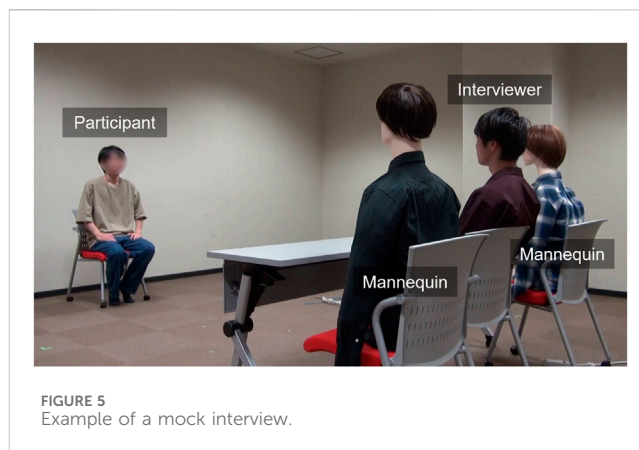
FIGURE 5
Example of a mock interview.
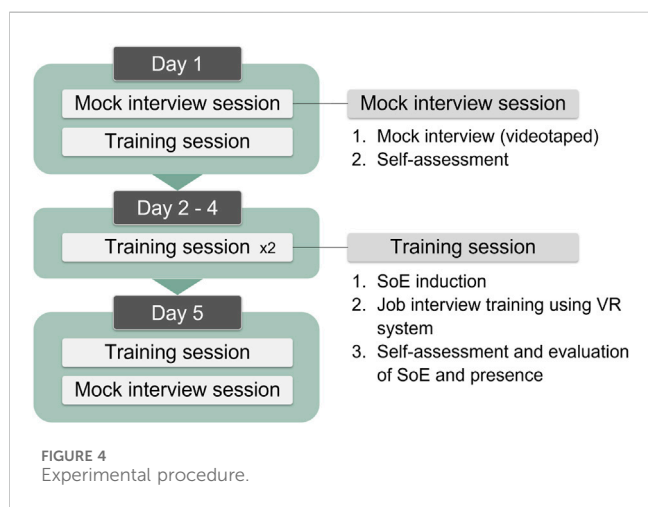
FIGURE 4
Experimental procedure.

was because the cognitive load was expected to be high when conducting VR training while considering the content of responses from the beginning. Table 3 displays the list of the interview questions, covering topics such as self-introduction, school activities, and the participant's research topic. In the experiment, the scenarios and interview questions shown in Table 2 were consistently used throughout all phases.

Additionally, participants were required to submit a photograph of their face in advance. Using the photograph, an avatar was created for each participant with Character Creator 4. The purpose of creating avatars for each participant was to standardize the impact of the avatar's appearance (Latoschik et al., 2017; Waltemate et al., 2018; Suk and Laine, 2023) on SoE.

The experiment spanned 5 days, with approximately 1 hour allocated to each day. The intervals between training dates were at least 1 day apart, with an average interval of 3.8 days (SD = 1.4). For each condition, the minimum interval was 2 (days), with 1PP having M = 3.68, SD = 0.83; Back having M = 3.34, SD = 0.71; and Front having M = 4.46, SD = 2.18. The entire experimental workflow is shown in Figure 4 and consists of two main sessions: a mock interview session and a training session.

2.6.1 Mock interview session

In this session, the experimenter played the role of an interviewer and conducted a face-to-face mock interview without

the VR system. Figure 5 illustrates the mock interview setup, which included two mannequins placed on either side of the interviewer to replicate the virtual environment as closely as possible. Participants in the mock interview were videotaped from the front for evaluation. Immediately after the mock interview, participants evaluated themselves using the questionnaire (see Table 2).
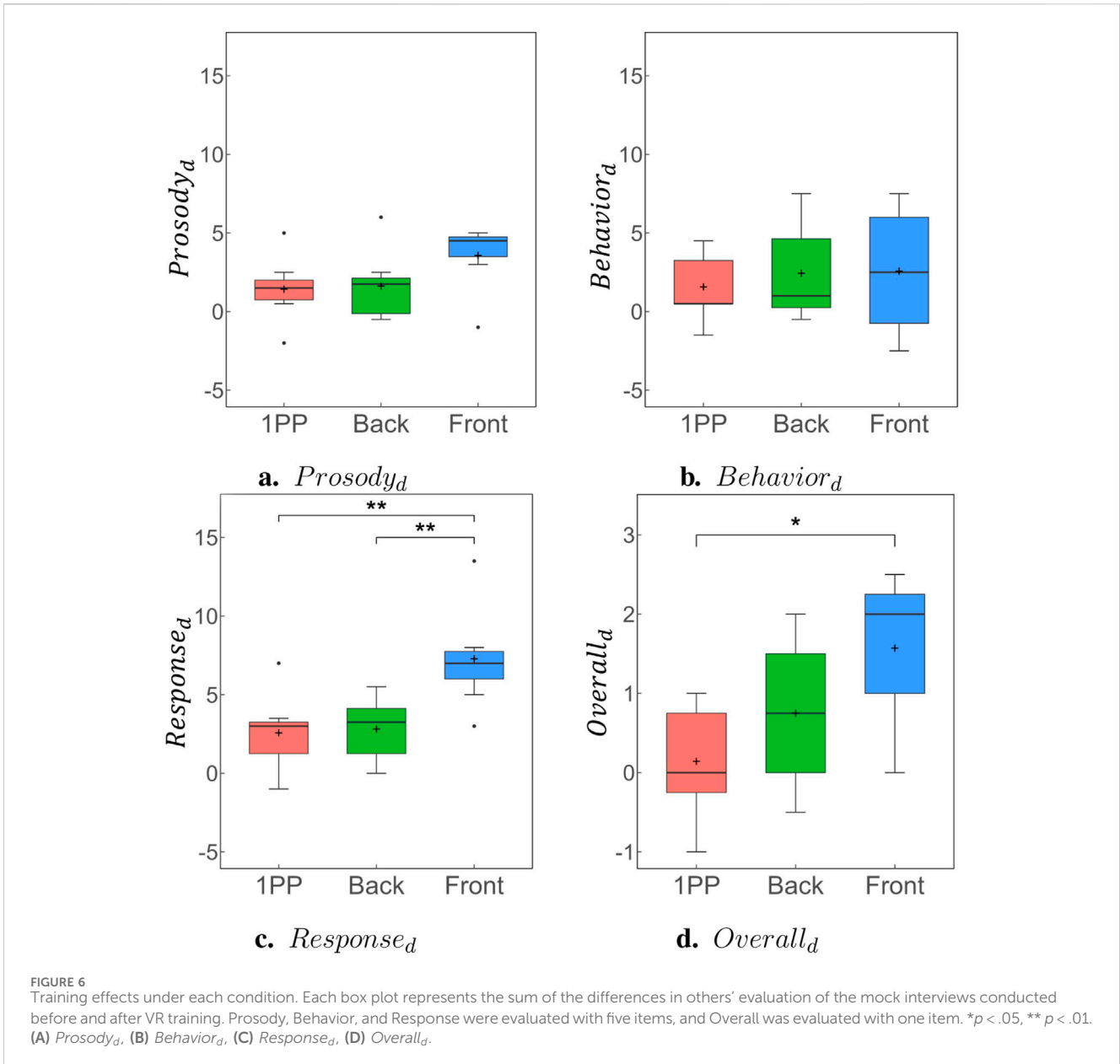
2.6.2 Training session

Participants initially wore an HMD, four motion trackers, and a hand tracker. They were instructed to stand in front of a chair positioned at the center of the experimental area. Subsequently, participants were asked to move their bodies, change their facial expressions, and observe the virtual environment according to voice instructions lasting approximately 2 min, aimed at inducing SoE. These instructions were based on those used in a prior study (Roth and Latoschik, 2020). However, because our VR system incorporates facial tracking and hand tracking, additional instructions were included to prompt participants to consciously engage their facial expressions and hand movements. In the 1PP and Back conditions, a virtual mirror was placed in front of the avatar to enable participants to observe the avatar's movements, similar to previous research (González-Franco et al., 2010; Banakou and Slater, 2017). Following this, three virtual interviewers replaced the mirror, and a 5–10 min VR job interview ensued, with participants responding to the questions (see Table 3). After the interviewer announced the end of the interview, participants removed the HMD, conducted self-assessments of their performance during the interview, and evaluated their experienced SoE and presence (Total 8 times: 1 time on Day 1, 2 times each on Days 2, 3, and 4, and 1 time on Day 5).

On the first day, participants received an explanation of the purpose and procedure of the experiment and then signed an experimental consent form. Participants were informed of their right to withdraw from the experiment at any time. They also completed surveys regarding their nationality and height. Their height was used to adjust the height of their avatars. Subsequently, participants underwent one mock interview session and one training session.

On days 2 through 4, participants completed two training sessions each day, with a five-minute break between the sessions.

On day 5, participants conducted one training session followed by a mock interview session. Finally, they provided

FIGURE 6
Training effects under each condition. Each box plot represents the sum of the differences in others' evaluation of the mock interviews conducted before and after VR training. Prosody, Behavior, and Response were evaluated with five items, and Overall was evaluated with one item. *$p < .05$, ** $p < .01$. **(A)** $Prosody_d$, **(B)** $Behavior_d$, **(C)** $Response_d$, **(D)** $Overall_d$.

open-ended responses to questions about their observations of the VR system.

At the start of each day, participants were given time to review the questionnaire items for self-assessment and their previously submitted answers. Upon completing the entire experimental schedule, participants received compensation of JPY 6,000.
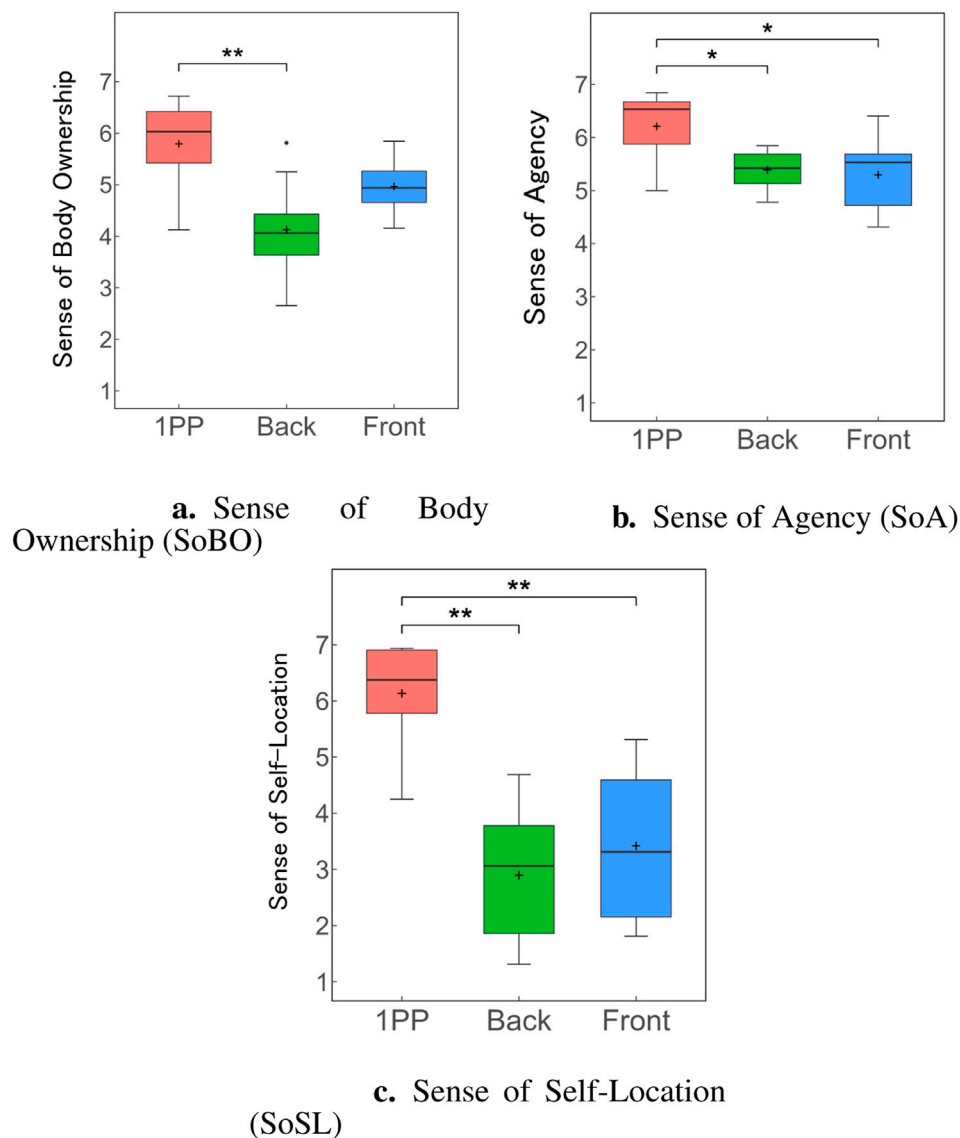
3 Results

All 22 participants completed the entire experimental schedule, and data on evaluations by others, SoE, and presence were obtained. All data were analyzed using R. Note that self-assessment data were excluded from the following report because they were supplementary measures in this experiment and did not differ significantly across conditions.

3.1 Training effects by perspective condition

Figure 6 displays training effects for each category under each condition. The Spearman's rank correlation coefficient between the two raters was $0.721 \, (p < .001)$.

The Shapiro-Wilk test and Bartlett's test were performed to assess the normality and equality of variances ($\alpha = .05$). Normality was verified in all categories except for $Prosody_d$. Equality of variances was confirmed in every category. Therefore, the Kruskal-Wallis test was conducted for $Prosody_d$, while one-way ANOVA was conducted for $Behavior_d$, $Response_d$, and $Overall_d$. Both of these tests were carried out between-participants. The one-way ANOVA indicated significant effects of perspective on $Response_d$ ($F(2, 19) = 7.40, p = .004$) and $Overall_d$ ($F(2, 19) = 4.86, p = .020$). None of the other comparisons were significant. Post hoc comparisons using the Tukey HSD test revealed

FIGURE 7
The results of SoBO, SoA, and SoSL for each condition. Each box plot shows the average of the eight VR training sessions. $*p < .05$, $** p < .01$. **(A)** Sense of Body Ownership (SoBO) **(B)** Sense of Agency (SoA) **(C)** Sense of Self-Location (SoSL).

that the mean score of $Response_d$ was significantly higher in Front ($M = 7.29, SD = 3.24$) than in 1PP ($M = 2.57, SD = 2.52$) and Back ($M = 2.81, SD = 2.00$) (Front vs. 1PP: $p = .008$, Front vs. Back: $p = .009$). The mean score of $Overall_d$ was also significantly higher in Front ($M = 1.57, SD = 0.93$) than in 1PP ($M = 0.14, SD = 0.75$) (Front vs. 1PP, $p = .015$).

3.2 SoE by perspective condition

Figure 7 displays the results of SoBO, SoA, and SoSL for each condition.

The Shapiro-Wilk test and Bartlett's test were performed to assess the normality and equality of variances ($\alpha = .05$). Normality was verified in every category except for SoSL. Equality of variances was confirmed in every category. Therefore, one-way ANOVAs were conducted for SoBO and SoA, while the Kruskal-Wallis test was conducted for SoSL. The one-way ANOVAs revealed significant effects of perspective on SoBO ($F(2, 19) = 7.40, p = .004$) and SoA ($F(2, 19) = 4.86, p = .020$). The Kruskal-Wallis test confirmed significant differences for SoSL ($H(2) = 11.92, p = .0005$). Based on these results, *post hoc* comparisons were performed using the Tukey HSD test for SoBO and SoA, and the Steel-Dwass test for SoSL. The results showed that the mean SoBO was significantly higher in 1PP ($M = 5.79, SD = 0.95$) than in Back ($M = 4.13, SD = 1.01$) (1PP vs. Back: $p = .004$). The mean SoA was significantly higher in 1PP ($M = 6.21, SD = 0.71$) than in both Back ($M = 5.39, SD = 0.37$) and Front ($M = 5.29, SD = 0.74$) (1PP vs. Back: $p = .048$, 1PP vs. Front: $p = .031$). Regarding SoSL, significant differences were identified between 1PP ($M = 6.13, SD = 1.01$) and Back ($M = 2.90, SD = 1.22$) and between 1PP and Front ($M = 3.42, SD = 1.48$) (1PP vs. Back: , $p = .002$, 1PP vs. Front: $p = .012$).
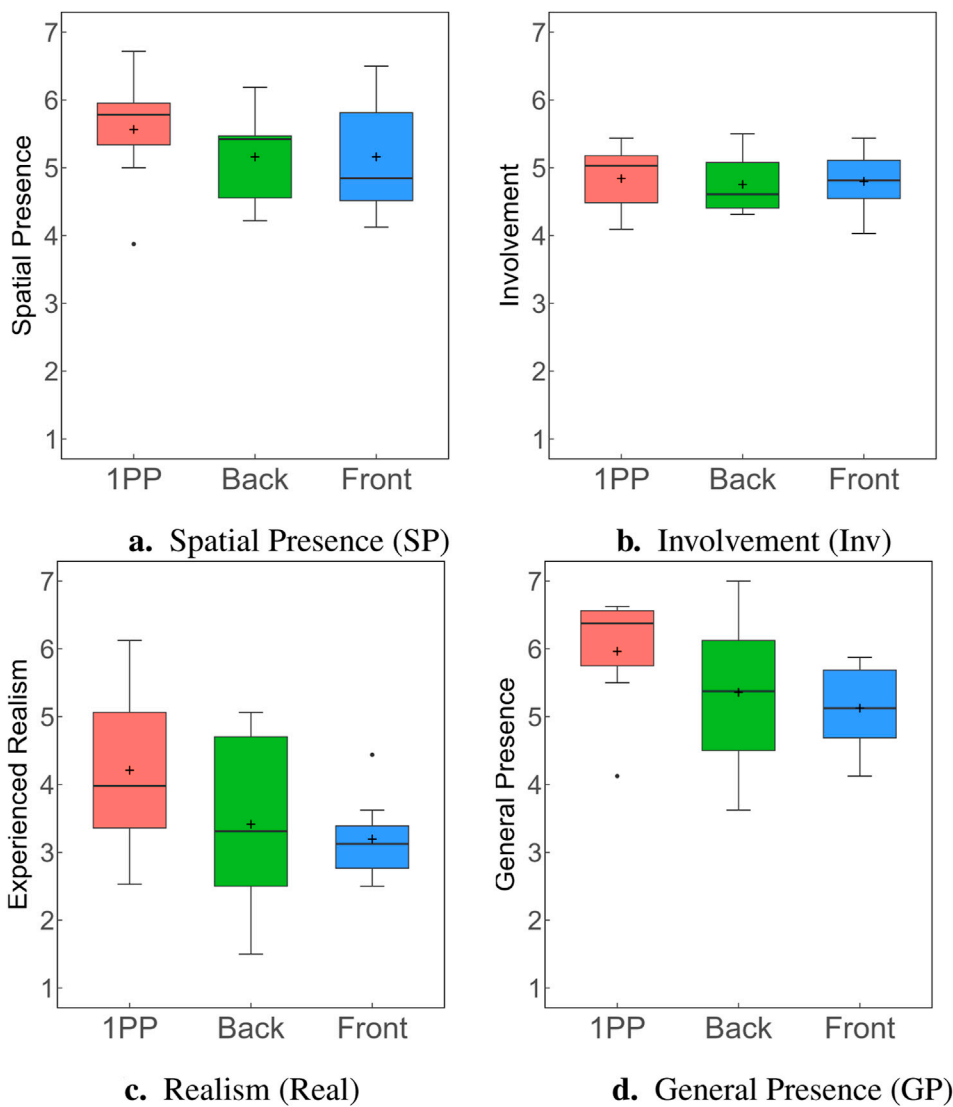
FIGURE 8
Presence results for each condition. From left to right, the figure shows the three IPQ subscales Spatial Presence (SP), Involvement (Inv), and Experienced Realism (Real), as well as an additional general item, General Presence (GP). Each box plot shows the average of the eight VR training sessions. **(A)** Spatial Presence (SP) **(B)** Involvement (Inv) **(C)** Realism (Real) **(D)** General Presence (GP).

3.3 Presence by perspective condition

Figure 8 shows the results for each of the IPQ subscales (SP, Inv, Real, and GP) under each condition.

The Shapiro-Wilk test and Bartlett's test were performed to assess the normality and equality of variances ($\alpha = .05$). Normality was verified in every category except for GP. Equality of variances was confirmed in every category. Therefore, one-way ANOVAs were conducted for SP, Inv, and Real, while the Kruskal-Wallis test was used for GP.

One-way ANOVAs revealed no significant effect of perspective on SP ($F(2, 19) = 0.58$, $p = .572$), Inv ($F(2, 19) = 0.06$, $p = .939$), and Real ($F(2, 19) = 1.52$, $p = .244$). Similarly, the Kruskal-Wallis test showed no significant differences among conditions for GP ($H(2) = 2.98$, $p = .232$).

3.4 Relationship between training effects and SoE, and between training effects and presence

We examined the correlations between the training effects as SoE, as well as between training effects and presence. Table 4 illustrates the correlation coefficients ($\alpha = 0.05$) for each condition between training effects ($Prosody_d$, $Behavior_d$, $Response_d$, and $Overall_d$) and the SoE factors (SoBO, SoA, and SoSL), as well as between training effects and the presence factors (SP, Inv, Real, and GP). Spearman's rank correlation coefficient was used to calculate the correlation coefficients.

In 1PP, a statistically significant negative correlation was found between $Overall_d$ and Real ($r = -.764$, $p = .046$). In Back, significant negative correlations were observed between $Overall_d$

TABLE 4 Correlation coefficients for each condition.

Condition	Training effect	Sense of embodiment			Presense			
		SoBO	SoA	SoSL	SP	Inv	Real	GP
1PP	$Prosody_d$	0.108	0.252	−0.101	−0.382	−0.126	−0.523	−0.300
	$Behavior_d$	−0.056	0.019	−0.152	−0.340	−0.356	−0.468	−0.321
	$Response_d$	−0.144	−0.054	−0.128	−0.336	0.126	−0.234	−0.491
	$Overall_d$	−0.255	−0.146	−0.454	−0.716	−0.291	**−0.764***	−0.615
Back	$Prosody_d$	−0.539	−0.687	−0.554	−0.224	0.285	−0.518	−0.337
	$Behavior_d$	−0.461	−0.410	−0.530	0.400	0.321	0.036	0.205
	$Response_d$	**−0.831***	−0.132	−0.527	−0.229	−0.199	−0.503	−0.359
	$Overall_d$	−0.636	−0.590	−0.470	−0.449	0.097	**−0.843****	−0.651
Front	$Prosody_d$	0.055	0.582	−0.257	−0.400	0.091	−0.182	−0.147
	$Behavior_d$	−0.036	0.071	0.505	−0.107	**0.786***	0.036	−0.414
	$Response_d$	0.446	0.685	0.118	0.234	**0.757***	0.090	0.118
	$Overall_d$	0.204	0.532	0.222	0.018	**0.771***	−0.018	−0.130

$*p < .05,**$ $p < .01.$

and Real ($r = −.843$, $p = .009$) and between $Response_d$ and SoBO ($r = −.831$, $p = .011$). On the other hand, in Front, significant positive correlations were found between $Behavior_d$ and Inv ($r = .786$, $p = .036$), $Response_d$ and Inv ($r = .757$, $p = .049$), and $Overall_d$ and Inv ($r = .771$, $p = .042$).

4 Discussion

Please note that this research is exploratory in nature. Also, due to the limitations of the sample size and current data analysis, we will avoid making definitive claims.

4.1 Training effects by perspective condition

The quality of the responses to questions improved more significantly when the training was done from the front third person perspective than when it was done from the first person or behind third person perspective. In addition, the overall quality of the interviews improved more significantly when the participants were trained from the third-person perspective than when they were trained from the first-person perspective. Therefore, Hypothesis 1 was partially supported in that training effects were higher in Front.

One possible reason for this difference is the influence of cognitive load (Sweller, 1988). Job interviews require attentiveness to speech content, language usage, and behavior. Additionally, in the VR training, the virtual agents moved independently of the participants' intentions, potentially increasing cognitive load. In Front, cognitive load was expected to be lower since the agents were not visible from the front. This could have allowed participants to focus on the training.

Another factor may be explained by the theory of objective self-awareness (Duval and Wicklund, 1972). This theory suggests that

people compare themselves to their own evaluation criteria when their attention is directed towards themselves, such as when standing in front of a mirror added. This often leads to self-evaluation, where one feels they are not meeting their own standards and experiences negative emotions. Consequently, people either attempt to distract themselves from this discrepancy or take action to mitigate it. In post-experimental interviews, all participants in Front reported that they mainly focused on the avatar's face during the VR training. Based on this theory, the participants in Front paid more attention to themselves, which may have led to higher training effects.

On the other hand, there was no significant difference in the improvement of the quality of behavior during the interview, regardless of which perspective the participants were trained in. Therefore, Hypothesis 2 was not supported. Similarly, there was no significant difference in the improvement of speech prosody during an interview, regardless of which perspective the participants were trained from. In the VR training, the agents only provided simple responses to the participants' answers and did not show facial expressions or gestures. In fact, some participants mentioned that they were not sure if their intentions were conveyed due to the limited reactions from the agents. The lack of flexibility in communication with the current implementation of the agent may have affected the training of nonverbal skills (prosody and behavior).

In summary, the use of the avatar's frontal perspective, which is not common in VR, can be beneficial for VR public speaking training.

4.2 SoE by perspective condition

SoBO was significantly lower in Back than in 1PP, consistent with previous studies (Gorisse et al., 2017; Bellido Rivas et al., 2021; Maselli and Slater, 2013). Meanwhile, SoBO in Front showed no

significant difference from that in 1PP. Participants in Front could see the avatar's face, which may have enhanced SoBO.

Some prior studies have reported differences in SoA depending on perspective (Hoppe et al., 2022), while others have not (Gorisse et al., 2017). According to Gorisse et al. (Gorisse et al., 2017), SoBO and SoSL were significantly affected by perspective, whereas SoA was less influenced by perspective.

SoSL was significantly lower in Back and Front than in 1PP. SoSL is known to be significantly affected by perspective (Galvan Debarba et al., 2017), and this is also observed in this study.

4.3 Presence by perspective condition

Regarding presence, there were no significant differences among the conditions in any component of the IPQ, thus it cannot be concluded that perspective has an effect on presence. Compared to the impact of perspective on SoE, the impact of perspective on presence is considered minor. This is consistent with previous research (Gorisse et al., 2017).

4.4 Relationship between training effects and SoE, and between training effects and presence

Note that the discussion in this subsection is currently based solely on the correlation between training effects and SoE/Presence, and it cannot be concluded a causal relationship.

4.4.1 Sense of embodiment

When trained from a third-person perspective from behind, the degree of improvement in the quality of responses to questions showed a negative correlation with SoBO. Low SoBO is believed to indicate a state in which one can detach oneself from the avatar (Scattolin et al., 2022). Therefore, one possible explanation for this negative correlation is that participants who experienced lower SoBO may have trained from a more objective standpoint, tends to result in better training effects.

4.4.2 Presence

There was a significant negative correlation between $Overall_d$ and Real in both 1PP and Back. This result contradicts Hypothesis 4. A possible explanation for this negative correlation could be that the more people perceived the virtual environment as realistic, the more stressed they felt, which negatively impacted the training. In fact, previous studies indicate that higher realism of virtual environments leads to increased levels of stress and cognitive load (Slater et al., 2009; Skulmowski, 2022).

As mentioned earlier, a job interview itself is a task that involves a high cognitive load. In addition, participants in 1PP and Back may have experienced a higher cognitive load because they faced the virtual agents. In Front, the virtual agent's front was not visible and the cognitive load was relatively low, which may have resulted in the lack of a significant correlation between Real and training effects.

On the other hand, in Front, Inv was significantly positively correlated with $Behavior_d$, $Response_d$, and $Overall_d$. This result partially supports Hypothesis 4. Participants who experienced a high

level of Inv were less attentive to the real environment and were able to concentrate more on the training, as Inv represents the sense of attention and involvement in a virtual environment.

4.5 Limitations

In our experiment, Front exhibited the most pronounced training effects among the three conditions. Combining this finding with other systems, such as a real-time feedback system that displays icons on an HMD to encourage trainees to improve their speech Palmas et al. (2021), has the potential to yield greater training effects. Designing a combination with other systems and user interfaces is a subject for future discussion. Moreover, in our VR job interview training, the virtual interviewer provided only simple responses to participants' answers and did not react with body language. This inflexibility in communication with the virtual interviewer may have impacted the training effects. The use of agents with more interactive communication functions and more flexible dialogue functions is a topic for the future. In addition, it will be interesting to analyze the relationship between the social presence/co-presence that users perceive in agents and training effectiveness.

Additionally, the current experiment was conducted in the context of a job interview task, with participants were limited to graduate students. In the future, it will be necessary to investigate whether the results are applicable to other public speaking scenarios, such as presentations, and to different participant groups.

Although the experiment yielded statistically significant results, the sample size for each condition consisted of 7 or 8 participants. Hence, the results of the correlations between training effects and SoE and presence should be interpreted carefully. Nevertheless, our findings, which suggest a similar trend in the correlation between 1PP and Back and a different trend in Front, may provide valuable insights.

5 Conclusion

This study pursued two objectives in VR job interview training: the first was to examine the variance in training effects depending on perspective, and the second was to explore the relationship between training effects and SoE, as well as between training effects and presence. Three experimental conditions were employed: first-person perspective (1PP), third-person perspective from behind the avatar (Back), and third-person perspective from in front of the avatar (Front).

The experimental results demonstrated that the Response score, which assesses verbal communication skills, exhibited a higher training effect in Front compared to 1PP and Back. Moreover, the Overall score, which measures the overall impression of the interview, displayed a higher training effect in Front than in 1PP. Thus, it can be concluded that job interview training from the avatar's front perspective is effective, even though it is not a commonly used perspective in VR.

Furthermore, we examined the correlations between SoE and presence in relation to training effects for each perspective condition. As a result, a significant negative correlation was identified between SoBO and the Response score in Back. In terms of presence, negative correlations were found between Real and the Overall score in 1PP and Back. In contrast, positive correlations were found between Behavior, Response, and Overall

scores with respect to Inv in Front. These results suggest that the correlation trends vary depending on perspective.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethical review board of Nara Institute of Science and Technology. 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

FU: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. YF: Conceptualization, Methodology, Project administration, Supervision, Writing–review and editing. TS: Methodology, Supervision, Writing–review and editing. MK: Methodology, Supervision, Writing–review and editing. HK: Methodology, Supervision, Writing–review and editing.

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.

Publisher's note

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

References

Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. A., Schmertz, S. K., Zimand, E., et al. (2013). Virtual reality exposure therapy for social anxiety disorder: a randomized controlled trial. *J. Consult. Clin. Psychol.* 81, 751–760. doi:10.1037/a0033559

Aymerich-Franch, L., Kizilcec, R. F., and Bailenson, J. N. (2014). The relationship between virtual self similarity and social anxiety. *Front. Hum. Neurosci.* 8, 944. doi:10.3389/fnhum.2014.00944

Banakou, D., and Slater, M. (2017). Embodiment in a virtual body that speaks produces agency over the speaking but does not necessarily influence subsequent real speaking. *Sci. Rep.* 7, 14227. doi:10.1038/s41598-017-14620-5

Bellido Rivas, A. I., Navarro, X., Banakou, D., Oliva, R., Orvalho, V., and Slater, M. (2021). The influence of embodiment as a cartoon character on public speaking anxiety. *Front. Virtual Real.* 2, 695673. doi:10.3389/frvir.2021.695673

Bhandari, N., and O'Neill, E. (2020). Influence of perspective on dynamic tasks in virtual reality. *Proc. 3DUI (IEEE)*, 939–948. doi:10.1109/VR46266.2020.00114

Burin, D., Kilteni, K., Rabuffetti, M., Slater, M., and Pia, L. (2019). Body ownership increases the interference between observed and executed movements. *PLOS ONE* 14, e0209899. doi:10.1371/journal.pone.0209899

Chollet, M., Wörtwein, T., Morency, L.-P., Shapiro, A., and Scherer, S. (2015). "Exploring feedback strategies to improve public speaking: an interactive virtual audience framework," in *Proc. UbiComp* (New York: Association for Computing Machinery), 1143–1154. doi:10.1145/2750858.2806060

Cmentowski, S., Krekhov, A., and Krüger, J. (2019). "Outstanding: a multi-perspective travel approach for virtual reality games," in *Proc. CHI PLAY* (New York: Association for Computing Machinery), 287–299. doi:10.1145/3311350.3347183

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., and Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Front. Psychol.* 6, 26. doi:10.3389/fpsyg.2015.00026

Duval, S., and Wicklund, R. A. (1972). A theory of objective self awareness.

Ebrahimi, O. V., Pallesen, S., Kenter, R. M., and Nordgreen, T. (2019). Psychological interventions for the fear of public speaking: a meta-analysis. *Front. Psychol.* 10, 488. doi:10.3389/fpsyg.2019.00488

El-Yamri, M., Romero-Hernandez, A., Gonzalez-Riojo, M., and Manero, B. (2019). "Emotions-responsive audiences for vr public speaking simulators based on the speakers' voice," in *Pcoc. ICALT (IEEE)*, 349–353. doi:10.1109/ICALT.2019.00108

Fydrich, T., Chambless, D. L., Perry, K. J., Buergener, F., and Beazley, M. B. (1998). Behavioral assessment of social performance: a rating system for social phobia. *Behav. Res. Ther.* 36, 995–1010. doi:10.1016/S0005-7967(98)00069-2

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends cognitive Sci.* 4, 14–21. doi:10.1016/S1364-6613(99)01417-5

Galvan Debarba, H., Bovet, S., Salomon, R., Blanke, O., Herbelin, B., and Boulic, R. (2017). Characterizing first and third person viewpoints and their alternation for embodied interaction in virtual reality. *PLOS ONE* 12, e0190109. doi:10.1371/journal.pone.0190109

Girondini, M., Stefanova, M., Pillan, M., and Gallace, A. (2023). Speaking in front of cartoon avatars: a behavioral and psychophysiological study on how audience design impacts on public speaking anxiety in virtual environments. *Int. J. Human-Computer Stud.* 179, 103106. doi:10.1016/j.ijhcs.2023.103106

Gonzalez-Franco, M., Ofek, E., Pan, Y., Antley, A., Steed, A., Spanlang, B., et al. (2020). The rocketbox library and the utility of freely available rigged avatars. *Front. Virtual Real.* 1. doi:10.3389/frvir.2020.561558

González-Franco, M., Pérez-Marcos, D., Spanlang, B., and Slater, M. (2010). "The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment," in *Proc. Virtual reality conference (VR)* (IEEE), 111–114. doi:10.1109/VR.2010.5444805

Gorisse, G., Christmann, O., Amato, E. A., and Richir, S. (2017). First-and third-person perspectives in immersive virtual environments: presence and performance analysis of embodied users. *Front. Robotics AI* 4. doi:10.3389/frobt.2017.00033

Hoorens, V. (1993). Self-enhancement and superiority biases in social comparison. *Eur. Rev. Soc. Psychol.* 4, 113–139. doi:10.1080/14792779343000040

Hoppe, M., Baumann, A., Tamunjoh, P. C., Machulla, T.-K., Woźniak, P. W., Schmidt, A., et al. (2022). "There is no first-or third-person view in virtual reality: understanding the perspective continuum," in *Proc. CHI* (New York: Association for Computing Machinery), 1–13. doi:10.1145/3491102.3517447

Hoque, M. E., Courgeon, M., Martin, J.-C., Mutlu, B., and Picard, R. W. (2013). "Mach: my automated conversation coach," in *Proc. UbiComp* (New York: Association for Computing Machinery), 697–706. doi:10.1145/2493432.2493502

Howard, M. C., and Gutworth, M. B. (2020). A meta-analysis of virtual reality training programs for social skill development. *Comput. and Educ.* 144, 103707. doi:10.1016/j.compedu.2019.103707

Jeunet, C., Albert, L., Argelaguet, F., and Lécuyer, A. (2018). "do you feel in control?": towards novel approaches to characterise, manipulate and measure the sense of agency in virtual environments. *IEEE Trans. Vis. Comput. Graph.* 24, 1486–1495. doi:10.1109/TVCG.2018.2794598

Jun, J., Jung, M., Kim, S.-Y., and Kim, K. (2018). "Full-body ownership illusion can change our emotion," in *Proc. CHI* (New York: Association for Computing Machinery), 1–11. doi:10.1145/3173574.3174175

Kilteni, K., and Ehrsson, H. H. (2017). Body ownership determines the attenuation of self-generated tactile sensations. *Proc. Natl. Acad. Sci.* 114, 8426–8431. doi:10.1073/pnas.1703347114

Kilteni, K., Groten, R., and Slater, M. (2012). The sense of embodiment in virtual reality. *Presence Teleoperators Virtual Environ.* 21, 373–387. doi:10.1162/PRES_a_00124

King, P. E., and Finn, A. N. (2017). A test of attention control theory in public speaking: cognitive load influences the relationship between state anxiety and verbal production. *Commun. Educ.* 66, 168–182. doi:10.1080/03634523.2016.1272128

Koek, W. J. D., and Chen, V. H. H. (2023). "The effects of avatar personalization and human-virtual agent interactions on self-esteem," in *Proc. VRW* (IEEE), 927–928. doi:10.1109/VRW58643.2023.00306

Kothgassner, O. D., Felnhofer, A., Beutl, L., Hlavacs, H., Lehenbauer, M., and Stetina, B. (2012). "A virtual training tool for giving talks," in *Proc. ICEC* (Berlin: Springer-Verlag), 53–66. doi:10.1007/978-3-642-33542-6_5

Kruger, J. (1999). Lake wobegon be gone the below-average effect and the egocentric nature of comparative ability judgments. *J. Personality Soc. Psychol.* 77, 221–232. doi:10.1037//0022-3514.77.2.221

Kwon, J. H., Powell, J., and Chalmers, A. (2013). How level of realism influences anxiety in virtual reality environments for a job interview. *Int. J. Human-Computer Stud.* 71, 978–987. doi:10.1016/j.ijhcs.2013.07.003

Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., and Botsch, M. (2017). "The effect of avatar realism in immersive social virtual realities," in *Proc. VRST* (New York: Association for Computing Machinery), 1–10. doi:10.1145/3139131.3139156

Maister, L., Slater, M., Sanchez-Vives, M. V., and Tsakiris, M. (2015). Changing bodies changes minds: owning another body affects social cognition. *Trends Cognitive Sci.* 19, 6–12. doi:10.1016/j.tics.2014.11.001

Maselli, A., and Slater, M. (2013). The building blocks of the full body ownership illusion. *Front. Hum. Neurosci.* 7, 83. doi:10.3389/fnhum.2013.00083

Matsumiya, K. (2021). Awareness of voluntary action, rather than body ownership, improves motor control. *Sci. Rep.* 11, 418. doi:10.1038/s41598-020-79910-x

Medeiros, D., dos Anjos, R. K., Mendes, D., Pereira, J. M., Raposo, A., and Jorge, J. (2018). "Keep my head on my shoulders! why third-person is bad for navigation in vr," in *Proc. VRST* (New York: Association for Computing Machinery), 1–10. doi:10.1145/3281505.3281511

Mostajeran, F., Balci, M. B., Steinicke, F., Kuhn, S., and Gallinat, J. (2020). "The effects of virtual audience size on social anxiety during public speaking," in IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE.

Naim, I., Tanveer, M. I., Gildea, D., and Hoque, M. E. (2015). "Automated prediction and analysis of job interview performance: the role of what you say and how you say it," in *Proc. FGIEEE*, 1–6. doi:10.1109/FG.2015.7163127

Newport, R., Pearce, R., and Preston, C. (2010). Fake hands in action: embodiment and control of supernumerary limbs. *Exp. Brain Res.* 204, 385–395. doi:10.1007/s00221-009-2104-y

Nikitina, A. (2011). *Successful public speaking* (Bookboon)

North, M. M., North, S. M., and Coble, J. R. (1998). Virtual reality therapy: an effective treatment for the fear of public speaking. *Int. J. Virtual Real.* 3, 1–6. doi:10.20870/IJVR.1998.3.3.2625

Ogawa, N., Narumi, T., and Hirose, M. (2019). Virtual hand realism affects object size perception in body-based scaling. *Proc. 3DUI (IEEE)*, 519–528. doi:10.1109/VR.2019.8798040

Osimo, S. A., Pizarro, R., Spanlang, B., and Slater, M. (2015). Conversations between self and self as sigmund freud—a virtual body ownership paradigm for self counselling. *Sci. Rep.* 5, 13899. doi:10.1038/srep13899

Palmas, F., Reinelt, R., Cichor, J. E., Plecher, D. A., and Klinker, G. (2021). "Virtual reality public speaking training: experimental evaluation of direct feedback technology acceptance," in *Proc. 3DUI (IEEE)*, 463–472. doi:10.1109/VR50410.2021.00070

Pelachaud, C. (2017). *Greta: realtime platform to control socio-emotional virtual characters.*

Piryankova, I. V., Wong, H. Y., Linkenauger, S. A., Stinson, C., Longo, M. R., Bülthoff, H. H., et al. (2014). Owning an overweight or underweight body: distinguishing the physical, experienced and virtual body. *PLOS ONE* 9, e103428. doi:10.1371/journal.pone.0103428

Poeschl, S. (2017). Virtual reality training for public speaking—a quest-vr framework validation. *Front. ICT* 4, 13. doi:10.3389/fict.2017.00013

Pyasik, M., Ciorli, T., and Pia, L. (2022). Full body illusion and cognition: a systematic review of the literature. *Neurosci. and Biobehav. Rev.* 143, 104926. doi:10.1016/j.neubiorev.2022.104926

Roth, D., and Latoschik, M. E. (2020). Construction of the virtual embodiment questionnaire (veq). *IEEE Trans. Vis. Comput. Graph.* 26, 3546–3556. doi:10.1109/TVCG.2020.3023603

Salamin, P., Thalmann, D., and Vexo, F. (2006). "The benefits of third-person perspective in virtual and augmented reality?," in *Proc. VRST* (New York: Association for Computing Machinery), 27–30. doi:10.1145/1180495.1180502

Scattolin, M., Panasiti, M. S., Villa, R., and Aglioti, S. M. (2022). Reduced ownership over a virtual body modulates dishonesty. *iScience* 25, 104320. doi:10.1016/j.isci.2022.104320

Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: factor analytic insights. *Presence Teleoperators Virtual Environ.* 10, 266–281. doi:10.1162/105474601300343603

Skulmowski, A. (2022). Is there an optimum of realism in computer-generated instructional visualizations? *Educ. Inf. Technol.* 27, 10309–10326. doi:10.1007/s10639-022-11043-2

Slater, M., Khanna, P., Mortensen, J., and Yu, I. (2009). Visual realism enhances realistic response in an immersive virtual environment. *IEEE Comput. Graph. Appl.* 29, 76–84. doi:10.1109/MCG.2009.55

Slater, M., and Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Front. Robotics AI* 3. doi:10.3389/frobt.2016.00074

Slater, M., and Wilbur, S. (1997). A framework for immersive virtual environments (five): speculations on the role of presence in virtual environments. *Presence Teleoperators and Virtual Environ.* 6, 603–616. doi:10.1162/pres.1997.6.6.603

Smith, T. E., and Frymier, A. B. (2006). Get 'real': does practicing speeches before an audience improve performance? *Commun. Q.* 54, 111–125. doi:10.1080/01463370500270538

Suk, H., and Laine, T. H. (2023). Influence of avatar facial appearance on users' perceived embodiment and presence in immersive virtual reality. *Electronics* 12, 583. doi:10.3390/electronics12030583

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Sci.* 12, 257–285. doi:10.1016/0364-0213(88)90023-7

Takac, M., Collett, J., Blom, K. J., Conduit, R., Rehm, I., and De Foe, A. (2019). Public speaking anxiety decreases within repeated virtual reality training sessions. *PLOS ONE* 14, e0216288. doi:10.1371/journal.pone.0216288

Tanaka, H., Negoro, H., Iwasaka, H., and Nakamura, S. (2017). Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PLOS ONE* 12, e0182151. doi:10.1371/journal.pone.0182151

Tay, C., Ang, S., and Van Dyne, L. (2006). Personality, biographical characteristics, and job interview success: a longitudinal study of the mediating effects of interviewing self-efficacy and the moderating effects of internal locus of causality. *J. Appl. Psychol.* 91, 446–454. doi:10.1037/0021-9010.91.2.446

Thakkar, J. H., Rao, P. S. B., Shubham, K., Jain, V., and Jayagopi, D. B. (2022). "Understanding interviewees' perceptions and behaviour towards verbally and non-verbally expressive virtual interviewing agents," in *Proc. ICMI* (New York: Association for Computing Machinery), 61–69. doi:10.1145/3536220.3558802

Tsakiris, M., Prabhu, G., and Haggard, P. (2006). Having a body versus moving your body: how agency structures body-ownership. *Conscious. cognition* 15, 423–432. doi:10.1016/j.concog.2005.09.004

Tsakiris, M., Schütz-Bosbach, S., and Gallagher, S. (2007). On agency and body-ownership: phenomenological and neurocognitive reflections. *Conscious. cognition* 16, 645–660. doi:10.1016/j.concog.2007.05.012

Valls-Ratés, Ï., Niebuhr, O., and Prieto, P. (2022). Unguided virtual-reality training can enhance the oral presentation skills of high-school students. *Front. Commun.* 7. doi:10.3389/fcomm.2022.910952

Waltemate, T., Gall, D., Roth, D., Botsch, M., and Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Trans. Vis. Comput. Graph.* 24, 1643–1652. doi:10.1109/TVCG.2018.2794629

Wang, Y., Hu, Z., Li, P., Yao, S., and Liu, H. (2022). Multiple perspectives integration for virtual reality-aided assemblability assessment in narrow assembly spaces. *Int. J. Adv. Manuf. Technol.* 119, 2495–2508. doi:10.1007/s00170-021-08292-9

Witmer, B. G., and Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence Teleoperators Virtual Environ.* 7, 225–240. doi:10.1162/105474698565686

Zhou, H., Fujimoto, Y., Kanbara, M., and Kato, H. (2021). Virtual reality as a reflection technique for public speaking training. *Appl. Sci.* 11, 3988. doi:10.3390/app11093988

Zopf, R., Truong, S., Finkbeiner, M., Friedman, J., and Williams, M. A. (2011). Viewing and feeling touch modulates hand position for reaching. *Neuropsychologia* 49, 1287–1293. doi:10.1016/j.neuropsychologia.2011.02.012