Virtual Triplets: A Mixed Modal Synchronous and Asynchronous Collaboration with Human-Agent Interaction in Virtual Reality

Jingjing Zhang
University of Canterbury
Christchurch, New Zealand
jingjing.zhang@pg.canterbury.ac.nz

Binyang Han
University of Canterbury
Christchurch, New Zealand
binyang.han@pg.canterbury.ac.nz

Ze Dong
University of Canterbury
Christchurch, New Zealand
ze.dong@pg.canterbury.ac.nz

Ruoyu Wen
University of Canterbury
Christchurch, New Zealand
ruoyu.wen@pg.canterbury.ac.nz

Gun A. Lee
University of South Australia
Australia
Gun.Lee@unisa.edu.au

Simon Hoermann
University of Canterbury
Christchurch, New Zealand
simon.hoermann@canterbury.ac.nz

Wendy Zhang
University of Canterbury
Christchurch, New Zealand
wendy.zhang@canterbury.ac.nz

Thammathip Piumsomboon
University of Canterbury
Christchurch, New Zealand
tham.piumsomboon@canterbury.ac.nz

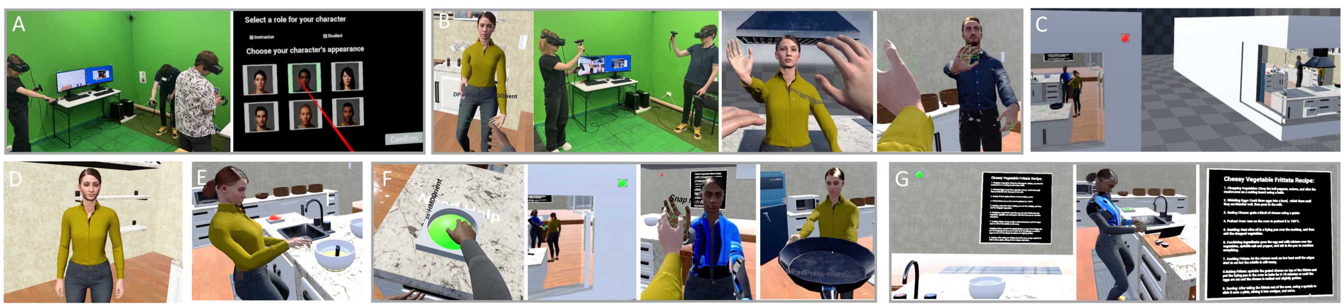

Figure 1: The Virtual Triplets system facilitates collaboration between an instructor and two students: A) Participants choose roles and avatars, B) Students control avatars; instructors switch between avatars in two classrooms, C) The instructor accesses the bird's-eye view with the left controller, D) The instructor's avatar is controlled by an agent when not in use, E) Instructors' actions are recorded in real-time; students request replays from the agent, F) The 'Ask-for-help' button in virtual classrooms alerts the instructor, G) The setting provides a structured recipe sequence for cooking guidance.

ABSTRACT

We introduce the Virtual Triplets (VTs), a Virtual Reality collaborative system designed for both synchronous and asynchronous interaction through collaboration between humans and human-agent. When a single instructor supervises two students in a showcased virtual cooking class scenario, the system enables the instructor to switch control between avatars in two separate environments, and a virtual agent takes over in the instructor's absence, ensuring continuous support. VTs facilitate the recording and playback of the instructor's demonstrations for students, coupled with a feature that allows the instructor to employ a bird's-eye view for effective classroom management. Analysis of observational data and interviews from our pilot study indicates good system usability, ease of avatar management for instructors, and a strong sense of continuous support among students. We discuss potential improvements and broader applications of VTs, aiming to enhance user experience in multitasking scenarios involving multiuser human-human and human-agent collaboration.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**.

KEYWORDS

Multi-User Collaboration, Human-Agent Interaction, Human-Agent Collaboration, Virtual Reality, Virtual Training, Embodied Virtual Agents

ACM Reference Format:
Jingjing Zhang, Binyang Han, Ze Dong, Ruoyu Wen, Gun A. Lee, Simon Hoermann, Wendy Zhang, and Thammathip Piumsomboon. 2024. Virtual Triplets: A Mixed Modal Synchronous and Asynchronous Collaboration

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3650766

with Human-Agent Interaction in Virtual Reality. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3613905.3650766

1 INTRODUCTION

Human beings, inherently bound by spatial limitations, can only exist in one location at a time. However, the advent of immersive technologies and artificial intelligence (AI) promises to transcend these constraints. Advancements in Immersive Virtual Environments (IVEs) and AI are reshaping collaborative experiences, enabling multitasking and personalised interactions that enrich the user experience. IVEs support scenarios beyond real-world confines, demonstrating their transformative impact in the future of collaboration [7, 19, 22, 26, 27, 35, 37].

This research introduces Virtual Triplets (VTs), a novel collaborative Virtual Reality (VR) system that seamlessly integrates AI agents in a mixed-modal approach to support both synchronous and asynchronous collaboration. VTs enables an instructor, aided by an AI agent, to simultaneously guide two students in an IVE, ensuring continuous engagement and consistent support. VTs leverages the advancements in four research areas, namely, virtual embodiments in collaborative settings, synchronous and asynchronous collaboration, embodied virtual agents, and immersive virtual learning.

Virtual Embodiments in Collaborative Settings—From telepointer [14] to representing user's physique, including limbs [10, 33] and full-body reconstructions [38]; these developments have enhanced social interaction, spatial awareness, and the sense of presence in remote collaboration. High-fidelity avatars, in particular, have been shown to increase social presence and realism [21, 39], and motion-controlled avatars enhancing user co-presence [17].

The term "Virtual Triplets" arose from our exploration of overcoming human limitations and allowing a user to be present in multiple locations simultaneously. This concept resembles out-of-body experiences or "mental bilocation," where a person seems to exist in two places at once [13]. Importantly, our concept allows the user to maintain a disembodied perspective, observing their multiple avatars as distinct entities. While previous research has focused on one-to-one control of virtual avatars, we are investigating the challenges of managing this unique disembodied viewpoint alongside multiple avatars.

Synchronous and Asynchronous Collaboration—The realm of synchronous collaboration offers several tools, each designed to enhance the virtual collaboration experience. Systems like Mini-me [26] utilise adaptive avatars for improved social presence, while others explore scalable avatars with handheld cameras [27] or enhanced expressions for lifelike communication [16]. Additionally, systems like ObserVAR [36] streamline gaze visualisation to support instructors, and Tutor In-sight [34] introduces miniature avatars for focused student guidance. By mirroring the immediacy of face-to-face interactions, these synchronous systems facilitate prompt feedback and a strong sense of co-presence in IVEs.

On the other hand, asynchronous collaboration offers unique advantages by providing flexibility in when and how users interact. Tools like XRStudio [25], TutoriVR [20], XR-LIVE [35], and the system by Cho et al. [9] all support asynchronous learning or training through features like recorded guidance, self-paced instructional materials, and immersive demonstrations. This flexibility accommodates individual schedules, promotes self-paced learning, and eases the instructional burden. To our knowledge, few systems effectively integrate synchronous and asynchronous collaboration despite the potential benefits of combining both modes. We refer to this unification as mixed-modal collaboration. For the comparison with previous research, refer to the CSCW Time-Space Matrix in Figure 2. Our research focuses on this integration, seeking to combine the immediacy of synchronous interactions with the flexibility of asynchronous engagement for a richer collaborative experience. However, a significant challenge lies in achieving a seamless blend of these modes without sacrificing the quality of user interactions.

Embodied Virtual Agents (EVAs)—EVAs are finding broader applications across education [32], healthcare [4], and specialised training [1]. Their roles range from mentors and counsellors [4, 29, 32] to training partners for job interviews and support for those with autism [6, 31]. This demonstrates their potential to improve real-world skills. However, using these agents within complex, multi-user collaborations in IVEs is still nascent. Existing virtual agent systems often lack the lifelike presence and interaction capabilities necessary for IVEs. We aim to develop agents that more authentically represent human instructors within IVEs. These EVAs will go beyond answering queries and demonstrating tasks, as humans do, using both verbal and non-verbal communication. This capability will enhance collaborative learning environments.

Interactivity in Immersive Virtual Learning—Interactivity in IVEs is transforming education by enhancing experiential learning [18] in labs, hands-on tasks, and training. XR-LIVE [35] introduced a virtual lab for unconfined compression tests for civil engineering education, enabling practical tasks such as soil weight calculations within an interactive IVE. Similarly, TutoriVR [20] improves VR painting tutorials by incorporating interactive 3D elements, promoting self-paced learning. Systems like that of Cho et al.[9] facilitate mixed reality task recording and playback, offering guided instructions in assembly tasks, thereby improving spatial understanding. Even cultural training benefits, with VR applications teaching Japanese cultural gestures [8].

Our system builds on this foundation by offering a virtual cooking scenario akin to a practical chemistry lab, mirroring XR-LIVE's approach [35]. It includes interactive tasks such as chopping and baking to simulate real-life kitchen activities, thereby ensuring an engaging learning experience. This choice underlines our system's flexibility in delivering diverse interactive learning experiences. By incorporating authentic tool manipulation, we aim to provide a realistic demonstration simulating real-world tasks.

Design Requirements — Based on the background research in the four areas, we have identified the gap and drawn the design requirements for VTs as follows:

R1 **Embodiments**—VTs must enable seamless control and switch between multiple avatars and a disembodied view, maintaining presence across virtual locations and fostering co-presence for collaboration.

R2 **Mixed-modality**—VTs must support mixed-modal collaboration for real-time interaction and immediate responsiveness, enabling flexible engagement on individual schedules, even asynchronously. VTs should also offer intuitive transitions between modes.

R3 **EVAs**—VTs must offer EVAs mimicking human instructors within IVEs utilising both verbal and non-verbal communication to deliver instructions, perform demonstrations, and engage in realistic interaction.

R4 **Interactivity**— VTs must deliver an authentic virtual cooking scenario, equipping users with highly interactive and realistic experiences using virtual kitchen tools to ensure that virtual tasks mirror the engagement and authenticity of real-world activities.

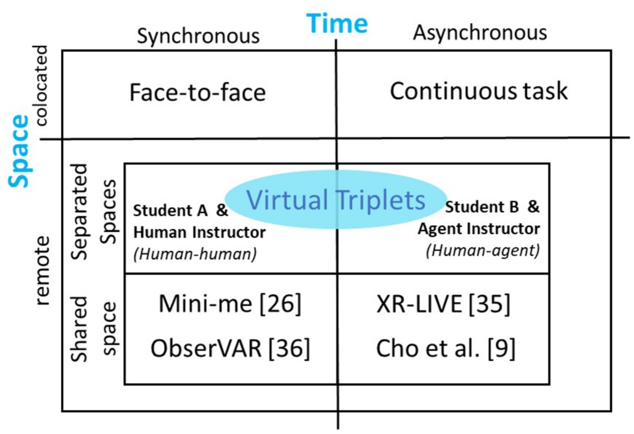

Figure 2: The Time-Space Matrix depicted in the figure highlights Virtual Triplets' capability to support collaboration in multiple, separated workspaces across both synchronous and asynchronous time domains in virtual environments.

2 VIRTUAL TRIPLETS SYSTEM DESIGN

2.1 Design Solutions

To ensure effective mixed-modal VR collaboration, our design solutions address the four outlined design requirements. For the conceptual overview of VTs, see Figure 3.

2.1.1 *Embodiments.*

S1 VTs provides various pre-customised humanoid avatars for users to choose from, promoting a stronger sense of ownership within the IVE.

S2 VTs supports instructors' control of two or more avatars, including avatar disembodiment.

2.1.2 *Mixed-modality.*

S3 VTs supports synchronous collaboration, allowing students to interact with instructors in real time.

S4 VTs supports asynchronous collaboration, allowing students to interact with instructors through EVAs for Q&A and to replay recorded demonstrations. This frees up instructors to manage multiple workspaces simultaneously.

S5 VTs' user interface offers seamless switching between synchronous and asynchronous collaboration modes, allowing instructors to multitask effectively across student engagements.

S6 VTs provides a notification system that enables students to promptly reach out to instructors for assistance, aligning with Maestro's approach [11] to enhance situational awareness and facilitate quick instructional intervention.

2.1.3 *EVAs.*

S7 VTs integrates EVAs capable of providing conversational responses and procedural demonstrations through recordings of the instructor's actions, akin to vAcademia's system [24]. This ensures the accurate replication of instructor actions, enhancing instructional continuity and presence.

2.1.4 *Interactivity.*

S8 VTs provides an interactive toolset specifically designed for a virtual cooking task scenario, supporting hands-on learning and practical skill development in an IVE.

2.2 Example Scenario of Interaction in VTs

The features of the system are illustrated in Figure 1 and listed as follows with corresponding design solutions:

(A) Initially, participants select their roles (instructor or student) and choose customised virtual avatars (S1).

(B) Students directly control their avatars. Instructors can switch between two identical avatars, each in separate virtual classrooms with a student, using a designated button on the right motion controller (S2, S3).

(C) Instructors can access a bird's-eye view by pressing a button on the left motion controller. Head position and orientation are tracked by the VR headset sensors, updating the user's IVE view while motion controller data adjusts the avatar's posture (S2, S4).

(D) When the instructor relinquishes control of the avatar, the agent assumes control (S4, S5, S7).

(E) Real-time recording of instructor actions is enabled using Unreal Engine 5's 'Take Recorder' function. Students can verbally request replays of these demonstrations from the agent. Communication with the agent is initiated using a push-to-talk button, with preset keywords triggering specific demonstrations (S4, S7).

(F) Each virtual kitchen classroom features an 'ask-for-help' button. Pressing this button in one room triggers visual indicators outside that room and inside the other room to alert the instructor (S6).

(G) A structured sequence of cooking steps for a specific recipe (see Appendix A) is provided within the system, systematically guiding participants through the cooking process (S8).

2.3 System Implementation

Virtual Triplets (VTs) is a multiuser networked system that operates on three high-performance PCs, each equipped with an Intel i7 8700 3.2 GHz processor, 32 GB RAM, an NVIDIA GeForce RTX2080 graphics card, and HP Reverb G2 VR headsets. We developed the system using Unreal Engine version 5.1.1 (UE5) on Microsoft Windows 11, integrating it with HP Reverb G2 via Windows Mixed

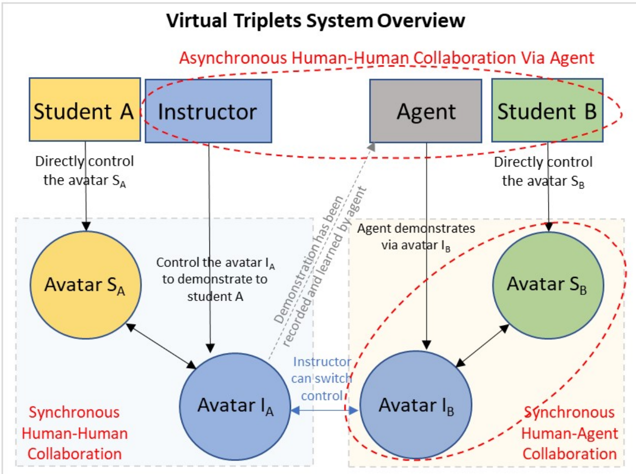

Figure 3: The overview of Virtual Triplets which supports three distinct collaborations: synchronous human-human, synchronous human-agent, and asynchronous human-human via an agent.

Reality. The MetaHuman Creator[1] was utilized for creating realistic, customisable avatars, set in a fully-equipped kitchen scenario with 3D objects from the UE5 online store and Quixel Bridge[2].

The system employs HP Reverb G2 headsets and controllers, with Unreal Engine scripting that links physical movements to in-game actions. This includes spatial audio and voice communication for enhanced user immersion. We use the Convai plugin[3] in UE5, which provides conversational AI agents capable of natural language communication in English and basic humanoid animations like natural idle poses and mouth movements during speech.

For networking, the system utilizes the VR Expansion Plugin[4] for its suite of multiuser and networking tools. VTs comprises a multi-user server and two client computers, each running the VTs instance of UE5. The system allows clients to access the central server remotely or locally via a Local Area Network. Up to three users can simultaneously share a virtual scene, with the server ensuring synchronized updates across each participant's virtual environment. Users interact through VR equipment, controlling avatars designed for realistic movement. Full-body skeletons, rigged to follow the motion controllers' movements using the Dragon IK Plugin[5], enable this realistic avatar control.

3 PILOT STUDY

The pilot study was designed to gain insights into the system's usability, identify areas for improvement, and explore potential applications. For this study, a virtual cooking task was chosen for its rich interactive potential and its demand for a combination of both procedural and narrative instruction, which effectively showcases the capabilities of the VTs, including AI assistance and mixed-modal transitioning in collaboration. The study has been approved by the University of Canterbury's Human Research Ethics Committee.

Our pilot study engaged three triads of nine university-affiliated individuals with six postgraduate students and three staff members (two females, seven males), with an average age of 32.7 years (SD = 4.53). Expertise was determined by participants' experience in digital technology research, resulting in two triads with all experts and one with all non-experts.

In the initial phase of our pilot study, the features of the VTs system were methodically introduced to the participants. The participant assuming the role of an instructor received hands-on training on how to switch control between avatars and to use the bird's-eye view of the IVE. The bird's-eye view is akin to the monitoring tool in VisTA-LIVE [12], which has proven useful for overseeing the progress of a larger class simultaneously. Conversely, participants in a student role were instructed on how to effectively communicate with the AI agent. This instruction covered querying the agent and activating task demonstrations via keywords that triggered the playback of instructor actions. They also learned how to signal for the instructor's assistance when needed.

Following these initial orientations, participants explored the virtual kitchen space and interacted with its array of tools to ensure familiarity with the VTs system's interactive capabilities. After mastering basic interactions and tool usage in the virtual kitchen, all participants were tasked with following a virtual recipe (see Appendix A). This task provided structure to the activity and emulated a genuine learning environment, demonstrating the VTs system's support for experiential learning processes [18].

3.1 Data Analysis from the Pilot Study

We monitored how participants interacted with the system, focusing on their understanding of its technology, functions, and interaction techniques. Key observations and any issues encountered were carefully noted.

Feedback Collection:

- *System Usability:* Using the System Usability Scale (SUS) questionnaire [5], participants evaluated the usability of our system.
- *Social Presence:* The social presence questionnaires [15] helped us understand the impact of substituting the instructor with an agent on the perceived social presence.

Interviews and Discussions:

- *Instructor Feedback:* Focused on their experiences in managing multiple avatars and interactions with students. Central question: *"How did you find managing multiple avatars simultaneously with the Virtual Triplets? What were the challenges and solutions?"*
- *Student feedback:* Centered on their experiences receiving support during virtual lessons. Key question: *"Describe your experience with the Virtual Triplets System's support during lessons. Were there moments of varying levels of support?"*

Insight and Future Direction:

- *System Improvements:* Suggestions for enhancing the current system.

[1]https://metahuman.unrealengine.com/
[2]https://quixel.com/bridge
[3]https://www.convai.com/
[4]https://vreue4.com/
[5]https://www.unrealengine.com/marketplace/en-US/product/dragon-ik-animal-inverse-kinematics

- *Future Applications:* Perspectives on potential future uses for the technology.
- *Technological Development:* Ideas for incorporating new features or future technological explorations with the system.

3.2 Results and Feedback

Following the pilot study, participants were asked to complete the SUS questionnaire, comprised of 10 statements rated on a 5-point Likert scale, where VTs received an average score of 79 (SD=4.2), surpassing the average benchmark score of 68. This score suggests a good level of usability but also highlights areas where further improvements can be made [3]. The social presence questionnaire with 36 statements across six subscales, rated on a 7-point Likert scale, was used to assess social presence using VTs. The results varied across subscales as shown in Figure 4: Co-presence scored highly (M=5.57, SD=1.51), indicating a strong sense of community; Attentional Allocation had a moderate score (M=3.98, SD=1.66), reflecting average attention management; both Perceived Message Understanding (M=4.54, SD=1.84) and Perceived Affective Understanding (M=4.37, SD=1.48) scored well, suggesting effective communication and emotional perception; Perceived Emotional Interdependence (M=3.56, SD=1.72) and Perceived Behavioral Interdependence (M=4.59, SD=1.23) showed varied levels of emotional and behavioural connection.

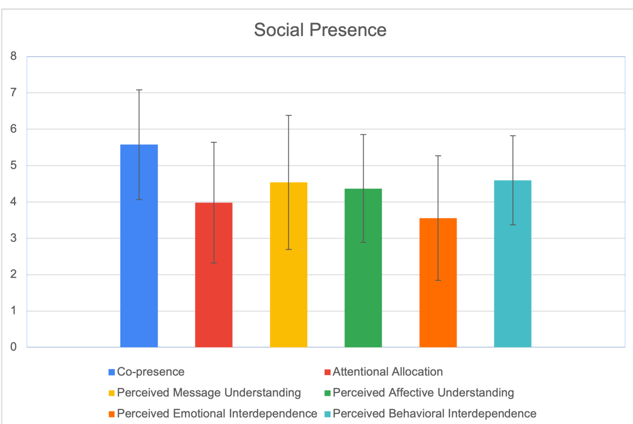

Figure 4: Social presence questionnaire results.

In our study, most participants in the student role reported feeling consistently supported by both the instructor and the AI agent, though one student noted feeling more supported by the AI agent compared to the instructor. From the instructors' perspective, all found managing multiple avatars and interacting with students straightforward. In refining our system, instructors and students provided key feedback as follows:

(1) Ask-for-Help Indicator: Two instructors occasionally overlooked the ask-for-help indicator. One suggested adding a sound notification, while another proposed relocating the indicator to the wrist, akin to a wearable device, for better visibility.
(2) Classroom Differentiation: All instructors struggled to distinguish between the two classrooms despite different student avatars. They recommended customising the virtual environments, such as changing wallpapers, for clearer differentiation.
(3) Student Progress Monitoring: Instructors expressed a need for a feature that provides real-time updates on each student's progress to enhance their ability to monitor and recognise each student's learning journey. This feedback highlights a challenge noted in prior research that varied learning paces and engagement in group workshops hinder instructors' ability to effectively track and support individual learning journeys [11]. Systems like Tutor In-sight [34], Maestro [11], and VisTA-LIVE [12] address this with dashboards that visualise student attention and provide activity, progress, and intervention suggestions. These examples emphasise the value of such features, suggesting potential integration into VTs to enhance the instructional experience.
(4) Agent Interaction and Guidance: Two students wanted the agent to follow them and provide clear starting points for demonstrations.
(5) Instructor's Virtual Presence: One student preferred a consistent location for the instructor's avatar, either beside them or at a fixed position, rather than where the real instructor last left it.
(6) Agent Presence and Assistance: Another student suggested that the agent should remain out of sight and only appear upon request to reduce the feeling of constant surveillance. This consideration resonates with previous findings indicating that constant tracking of student behaviours, such as eye movement or facial expressions, can evoke discomfort due to privacy concerns [30].

Participants provided a variety of suggestions for potential applications of the system across different fields, such as *science classes*, *on-board training*, *tour guiding*, *rehabilitation*, *assembly tasks*, and *operation of dangerous machinery*. Upon reviewing these suggestions, a common theme emerged: the system seems ideally suited for fields that involve simple and repetitive hands-on tasks but also require customised information provision. This insight aids in identifying areas where the system could be most effectively employed, balancing the need for repetitive skill practice with personalised instructional content.

4 DISCUSSION AND FUTURE WORK

Based on our observations and user feedback, we recognise the importance of training sessions, particularly for those in the instructor role. Ensuring instructors are well-versed in VR technology is vital, as it directly impacts the effectiveness of the learning experience. Similarly, acquainting non-expert users with VR technology, including navigation and basic controller interaction techniques, is equally crucial to ensure effective implementation of the system. This aligns with the findings of Merchant et al., who highlighted the importance of orienting users to VR environments to optimise learning outcomes [23]. Furthermore, user feedback has been instrumental in guiding our approach towards more suitable task scenarios that effectively leverage this technology. Designing task scenarios based on user feedback ensures a more user-centric approach, which can significantly enhance the overall VR experience.

Currently, our system presents a limitation in the interaction dynamics between the instructor, student, and the AI agent. It is structured so that students can interact solely with either the instructor or the AI agent. This setup allows the instructor to adopt a bird's-eye view when students are engaging with the agents. However, our observations indicate a preference among instructors for more active engagement with students rather than just operating a bird's-eye view. This preference for the immersive involvement of instructors in VR environments, enhancing the learning experience, is echoed in the work of Radianti et al. [28].

To address this, we are considering the introduction of a 'ghost view' feature. This feature would allow the instructor to be present alongside the student in a non-obtrusive manner, enabling autonomous learning from the agent while still offering instructor oversight. Additionally, we are exploring the possibility of overlapping two or more virtual workspaces. This concept is intended to maintain a bird's-eye view for the instructor, enabling them to simultaneously monitor the activities students are engaged in. By doing so, we aim to provide a more comprehensive and holistic view of the student's actions and interactions within the VR environment. Importantly, this feature is designed to enhance the instructor's observational capabilities without overwhelming or distracting them.

A major challenge observed in the study is the confusion experienced by several students regarding whether their interactions are with a real human or an AI agent. This aligns with the previous finding [2] that social perceptions of AI can vary based on communication dynamics and the AI's portrayal as human-like or mechanical. Addressing this, an expert participant suggested exploring technologies that enable seamless transitions between the real instructor and the agent, aiming for a smooth and indistinguishable shift. In contrast, feedback from two students highlighted a different approach proposing a hint system that would clearly indicate when the real instructor takes over, favouring distinct and recognisable transitions. This diverse feedback has inspired us to investigate the transition behaviour between human instructors and AI agents further.

This exploration examines the transition behaviours between human instructors and AI agents, echoing previous research on the impact of virtual characters' realism in IVEs. Earlier studies [21, 39] explored how varying rendering styles, from highly realistic to cartoonish avatars, affect user engagement. Applying this to our context, we plan to test a spectrum of transition behaviours between human instructors and AI agents, ranging from seamless to distinct switches, including intermediate states. Our investigation focuses on how the nature of these transitions influences student perceptions and interactions within the VR system.

At the same time, we aim to explore how these transitions affect instructors' perspectives, especially in terms of body ownership when switching the embodiment with AI agents of varying realism and voice. Specifically, we question whether instructors would perceive a stronger connection to the AI agent closely resembles their own appearance and voice. Our future study will address this by comparing instructors' experiences across different AI representations, aiming to correlate realism with body ownership and the perceived quality of the delivery of their teaching.

In follow-up research, we will focus on two primary aspects:

(1) To understand the impact of AI integration, we will conduct comparative studies. We'll compare the effectiveness of our VR system in three scenarios: 1) human instructor with AI agent support, 2) without AI support, i.e. avatar paused during absence, and 3) AI only without a human instructor. This will allow us to critically evaluate the system's performance and determine the most effective way to use AI for enhanced learning.
(2) We aim to delve deeper into the subtleties of how instructors and students experience transitions between human and AI entities in IVEs. We plan to investigate a range of transition styles and their impact on the instructor's sense of body ownership and student perception. Testing will compare varying degrees of avatar and voice realism, from highly lifelike to distinctly artificial, to determine the optimal balance that preserves instructor presence and enhances the learning experience.

5 LIMITATIONS

One of the limitations identified in our pilot study was the absence of a comprehensive training session prior to the experiment, an aspect we plan to incorporate in the formal study. Secondly, the experimental setup confined all three users to a single room due to the limited experimental space. In the ideal case, the space should be separated to provide a more realistic and less constrained environment. Finally, we observed that some users, particularly those without prior immersive virtual reality experience, experienced discomfort or dizziness when engaging with the technology.

6 CONCLUSION

In this paper, we introduce 'Virtual Triplets,' a system for mixed-modal collaboration within an immersive virtual environment, as demonstrated through a virtual cooking scenario. This innovative system enables simultaneous control of multiple avatars by either humans or AI agents, effectively reducing human workload while ensuring dynamic interaction. Our pilot study of the Virtual Triplets system has provided substantial insights into its potential and versatility. The system's application in fields that require repetitive, hands-on tasks demonstrates its broad scope for future deployment. Crucially, the feedback received from both instructors and students has been pivotal in refining our current system and directing our future technological explorations.

REFERENCES

[1] Keith Anderson, Elisabeth André, T. Baur, Sara Bernardini, M. Chollet, E. Chryssafidou, I. Damian, C. Ennis, A. Egges, P. Gebhard, H. Jones, M. Ochs, C. Pelachaud, Kaśka Porayska-Pomsta, P. Rizzo, and Nicolas Sabouret. 2013. The TARDIS Framework: Intelligent Virtual Agents for Social Coaching in Job Interviews. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* Vol. 8253 LNCS. 476–491. https://doi.org/10.1007/978-3-319-03161-3{_}35
[2] Zahra Ashktorab, Casey Dugan, James Johnson, Qian Pan, Wei Zhang, Sadhana Kumaravel, and Murray Campbell. 2021. Effects of Communication Directionality and AI Agent Differences in Human-AI Interaction. 1–15. https://doi.org/10.1145/3411764.3445256

[3] A Bangor, P Kortum, J Miller Journal of usability studies, and undefined 2009. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *uxpajournal.org* 4 (2009), 114–123. https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Bangor_May2009.pdf

[4] Timothy W. Bickmore, Daniel Schulman, and Candace Sidner. 2013. Automated interventions for multiple health behaviors using conversational agents. *Patient Education and Counseling* 92, 2 (2013). https://doi.org/10.1016/j.pec.2013.05.011

[5] John Brooke. 1996. SUS -A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996).

[6] Zoraida Callejas, Brian Ravenet, Magalie Ochs, and Catherine Pelachaud. 2014. A computational model of social attitudes for a virtual recruiter. In *13th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2014*, Vol. 1.

[7] Yuanzhi Cao, Xun Qian, Tianyi Wang, Rachel Lee, Ke Huo, and Karthik Ramani. 2020. An Exploratory Study of Augmented Reality Presence for Tutoring Machine Tasks. In *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3313831.3376688

[8] Alan Cheng, Lei Yang, and Erik Andersen. 2017. Teaching language and culture with a virtual reality game. In *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 2017-May. https://doi.org/10.1145/3025453.3025857

[9] Hyunwoo Cho, Bowen Yuan, Jonathon Derek Hart, Eunhee Chang, Zhuang Chang, Jiashuo Cao, Gun A. Lee, Thammathip Piumsomboon, and Mark Billinghurst. 2023. An Asynchronous Hybrid Cross Reality Collaborative System. *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (10 2023), 70–73. https://doi.org/10.1109/ISMAR-ADJUNCT60411.2023.00022

[10] Andre Doucette, Carl Gutwin, Regan Mandryk, Miguel Nacenta, and Sunny Sharma. 2013. Sometimes when we touch: How arm embodiments change reaching and coordination on digital tables. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. https://doi.org/10.1145/2441776.2441799

[11] Volodymyr Dziubak, Ben Lafreniere, Tovi Grossman, Andrea Bunt, and George Fitzmaurice. 2018. Maestro: Designing a System for Real-Time Orchestration of 3D Modeling Workshops. (2018). https://doi.org/10.1145/3242587.3242606

[12] Pak Ming Fan, Santawat Thanyadit, and Ting Chuen Pong. 2023. VisTA-LIVE: A Visualization Tool for Assessment of Laboratories in Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* (2023). https://doi.org/10.1109/TVCG.2023.3341079

[13] Tiziano Furlanetto, Cesare Bertone, and Cristina Becchio. 2013. The bilocated mind: new perspectives on self-localization and self-identification. *Frontiers in Human Neuroscience* 7 (2013). https://doi.org/10.3389/fnhum.2013.00071

[14] S. Greenberg, C. Gutwin, and M. Roseman. 1996. Semantic telepointers for groupware. In *Proceedings - 6th Australian Conference on Computer-Human Interaction, OzCHI 1996*. https://doi.org/10.1109/OZCHI.1996.559988

[15] Chad Harms and Frank Biocca. 2004. Internal Consistency and Reliability of the Networked Minds Measure of Social Presence. *Seventh Annual International Workshop: Presence 2004* (2004).

[16] Jonathon Derek Hart, Thammathip Piumsomboon, Gun A. Lee, Ross T. Smith, and Mark Billinghurst. 2021. Manipulating Avatars for Enhanced Communication in Extended Reality. In *Proceedings - 2021 IEEE International Conference on Intelligent Reality, ICIR 2021*. https://doi.org/10.1109/ICIR51845.2021.00011

[17] P. Heidicker, E. Langbehn, and F. Steinicke. 2017. Influence of avatar appearance on presence in social VR. In *2017 IEEE Symposium on 3D User Interfaces, 3DUI 2017 - Proceedings*. 233–234. https://doi.org/10.1109/3DUI.2017.7893357

[18] D. A. Kolb. 2015. Experiential Learning: Experience as the source of Learning and Development Second Edition. *Pearson Education, Inc. Upper Saddle River* (2015), 41. https://doi.org/10.1002/job.4030080408

[19] B.T. Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman. 2019. Loki: Facilitating remote instruction of physical tasks using Bi-directional mixed-reality telepresence. In *UIST 2019 - Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 161–174. https://doi.org/10.1145/3332165.3347872

[20] B.T. Kumaravel, C. Nguyen, S. DiVerdi, and B. Hartmann. 2019. TutoriVR: A video-based tutorial system for design applications in virtual reality. In *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300514

[21] Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. 2017. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*, Vol. Part F131944. https://doi.org/10.1145/3139131.3139156

[22] Gun Lee, Seungjun Ahn, William Hoff, and Mark Billinghurst. 2019. AR Tips: Augmented first-person view task instruction videos. *Adjunct Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2019* (10 2019), 34–36. https://doi.org/10.1109/ISMAR-ADJUNCT.2019.00024

[23] Zahira Merchant, Ernest T. Goetz, Lauren Cifuentes, Wendy Keeney-Kennicutt, and Trina J. Davis. 2014. Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education* 70 (1 2014), 29–40. https://doi.org/10.1016/J.COMPEDU.2013.07.033

[24] Mikhail Morozov, Alexey Gerasimov, Mikhail Fominykh, and Andrey Smorkalov. 2013. Asynchronous immersive classes in a 3D virtual world: Extended description of vAcademia. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 7848. https://doi.org/10.1007/978-3-642-38803-3{_}5

[25] Michael Nebeling, Shwetha Rajaram, Liwei Wu, Yifei Cheng, and Jaylin Herskovitz. 2021. XRStudio: A Virtual Production and Live Streaming System for Immersive Instructional Experiences. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3411764.3445323

[26] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. Mini-me: An adaptive avatar for Mixed Reality remote collaboration. In *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 2018-April. https://doi.org/10.1145/3173574.3173620

[27] Thammathip Piumsomboon, Gun A. Lee, Andrew Irlitti, Barrett Ens, Bruce H. Thomas, and Mark Billinghurst. 2019. On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction. In *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300458

[28] Jaziar Radianti, Tim A. Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt. 2020. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers and Education* 147 (2020). https://doi.org/10.1016/j.compedu.2019.103778

[29] Albert Rizzo, Stefan Scherer, David Devault, Jonathan Gratch, Ronald Artstein, Arno Hartholt, Gale Lucas, Stacey Marsella, Fabrizio Morbini, Angela Nazarian, Giota Stratou, David Traum, Rachel Wood, Jill Boberg, and Louis Philippe Morency. 2016. Detection and computational analysis of psychological signals using a virtual human interviewing agent. *Journal of Pain Management* 9, 3 (2016).

[30] Kyoungwon Seo, Joice Tang, Ido Roll, Sidney Fels, and Dongwook Yoon. 2021. The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education* 18, 1 (2021). https://doi.org/10.1186/s41239-021-00292-9

[31] Matthew J. Smith, Emily J. Ginger, Katherine Wright, Michael A. Wright, Julie Lounds Taylor, Laura Boteler Humm, Dale E. Olsen, Morris D. Bell, and Michael F. Fleming. 2014. Virtual reality job interview training in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 44, 10 (2014). https://doi.org/10.1007/s10803-014-2113-y

[32] William Swartout, Benjamin D. Nye, Arno Hartholt, Adam Reilly, Arthur C. Graesser, Kurt VanLehn, Jon Wetzel, Matt Liewer, Fabrizio Morbini, Brent Morgan, Lijia Wang, Grace Benn, and Milton Rosenberg. 2016. Designing a personal assistant for life-long learning (PAL3). In *Proceedings of the 29th International Florida Artificial Intelligence Research Society Conference, FLAIRS 2016*.

[33] Anthony Tang, Carman Neustaedter, and Saul Greenberg. 2007. VideoArms: Embodiments for Mixed Presence Groupware. In *People and Computers XX — Engage*. Springer London, London, 85–102. https://doi.org/10.1007/978-1-84628-664-3{_}8

[34] Santawat Thanyadit, Matthias Heintz, and Effie L.C. Law. 2023. Tutor In-sight: Guiding and Visualizing Students' Attention with Mixed Reality Avatar Presentation Tools. In *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3544548.3581069

[35] S. Thanyadit, P. Punpongsanon, T. Piumsomboon, and T.-C. Pong. 2022. XR-LIVE: Enhancing Asynchronous Shared-Space Demonstrations with Spatial-temporal Assistive Toolsets for Effective Learning in Immersive Virtual Laboratories. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022). https://doi.org/10.1145/3512983

[36] Santawat Thanyadit, Parinya Punpongsanon, and Ting Chuen Pong. 2019. ObserVAR: Visualization system for observing virtual reality users using augmented reality. In *Proceedings - 2019 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2019*. https://doi.org/10.1109/ISMAR.2019.00023

[37] T.-Y. Wang, Y. Sato, M. Otsuki, H. Kuzuoka, and Y. Suzuki. 2019. Effect of full body avatar in augmented reality remote collaboration. In *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*. 1221–1222. https://doi.org/10.1109/VR.2019.8798044

[38] Boram Yoon, Hyung Il Kim, Gun A. Lee, Mark Billinqhurst, and Woontack Woo. 2019. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*. https://doi.org/10.1109/VR.2019.8797719

[39] K. Zibrek, E. Kokkinara, and R. McDonnell. 2018. The effect of realistic appearance of virtual characters in immersive environments - Does the character's personality play a role? *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1681–1690. https://doi.org/10.1109/TVCG.2018.2794638

A COOKING RECIPE: CHEESY VEGETABLE FRITTATA

Step 1: Chopping Vegetables: chop the bell peppers, onions, and slice the mushrooms on a cutting board using a knife.

Step 2: Whisking Eggs: crack three eggs into a bowl, whisk them until they are blended well, then pour in the milk.

Step 3: Grating Cheese: grate a block of cheese using a grater.

Step 4: Preheat Oven: turn on the oven to preheat it to 190°C.

Step 5: Sautéing: heat olive oil in a frying pan over the cooktop, and then add the chopped vegetables.

Step 6: Combining Ingredients: pour the egg and milk mixture over the vegetables, sprinkle salt and pepper, and stir in the pan to combine everything.

Step 7: Cooking Frittata: let the mixture cook on low heat until the edges start to set but the middle is still runny.

Step 8: Baking Frittata: sprinkle the grated cheese on top of the frittata and put the frying pan in the oven to bake for 8-10 minutes or until the eggs are set and the cheese is melted and slightly golden.

Step 9: Serving: after taking the frittata out of the oven, students can use a spatula to slide it onto a plate, slice it into wedges, and serve.