Exploring the Contributing Factors of First-Person Perspective Video on Instruction: A Preliminary Study

NEIL JUSTIN ROMBLON and YASUYUKI SUMI, Future University Hakodate, Japan

The advancements in mixed reality (MR) research have opened new doors for computer-supported cooperative work (CSCW). To address the lack of co-presence asynchronous CSCW systems, multiple strategies varying from the utilization of embodied avatars of the co-collaborator; to displaying first-person perspective video recording of peers have been employed, and have found success in task performance improvements and recall. This paper presents a preliminary study to explore whether this effect is achieved by the immersiveness of a mixed reality platform, or simply from the first-person perspective video recording.

CCS Concepts: • **Human-centered computing** → *Empirical studies in HCI*.

Additional Key Words and Phrases: First-person Perspective Video, Eye Tracking, Assembly Task

ACM Reference Format:
Neil Justin Romblon and Yasuyuki Sumi. 2025. Exploring the Contributing Factors of First-Person Perspective Video on Instruction: A Preliminary Study. 1, 1 (March 2025), 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction and Background

Mixed Reality (MR) research has steadily been more developed in recent years. As such, it has also brought new life to the field of Computer-Supported Cooperative Work (CSCW). Novel ways to enable collaboration have emerged, such as embodied avatars [8, 16], immersive first-person perspectives [17, 18], and spatial annotations [9, 10]. With the variety of features that MR can introduce to CSCW, efforts have been made to categorize MR collaboration in various dimensions [6, 19], such as remote vs co-located, technologically symmetric vs asymmetric, etc. Among these dimensions exists the dichotomy between synchronous vs asynchronous.

Synchronous and asynchronous CSCW systems are differentiated by whether or not the collaborators are engaging at same time; regardless of whether they may be co-located or not [6, 11]. Traditionally, synchronous CSCW systems are represented through videoconferencing, telephone, whiteboards, among others. While traditional asynchronous CSCW systems can be exemplified by technologies such as emails and instant messaging [11].

Some advantages of asynchronous CSCW systems over synchronous CSCW systems is that it allows for flexible time-coordination, work parallelism, and reviewability [12, 15]. However, one limitation in asynchronous CSCW platforms is the typical lack of co-presence among collaborators, which may lead to difficulties in coordination and awareness [4]. However, with mixed reality brought into the picture, a variety of possibilities are suddenly available–from showcasing spatially placed annotations in one's physical workspace; to presenting a multimodal virtual embodiment of one's collaborator as a recorded message.

Asynchronous collaboration in mixed reality typically envisions the avatar of the co-collaborator as a separate entity from the subject to assist in complications brought by the lack of co-presence [3, 13]. A more direct approach of displaying the first-person video recording of the co-collaborator has also been studied [14]. This study has found improvements in

Authors' Contact Information: Neil Justin Romblon, n-romblon@sumilab.org; Yasuyuki Sumi, sumi@fun.ac.jp, Future University Hakodate, Hakodate, Hokkaido, Japan.

Manuscript submitted to ACM

comprehension and more accurate task recall, despite being less complex than the embodied avatar approach. Another study [17] explored the utilization of first-person instructional video within a virtual reality environment, which found improvements in task performance but did not contribute much towards conceptual knowledge.

However, these studies utilize first-person perspective videos in an immersive lens–within virtual reality. The study by Bréchet et al. [2] points out that bodily presence is necessary for aiding the recall of episodic memories. While short-term memory may not necessarily equate to episodic memory, it might be possible that bodily presence is also a prerequisite for short-term memory retention. As such, we would like to explore the question, "How much does the utilization of first-person perspective alone contribute towards improvements in task performance?" For the following sections, we will briefly discuss the methodology, and preliminary results we have obtained from a small-scale experiment.

2 Methods

For the experiment, participants are tasked to assemble a toy block model, wherein depending on their randomly-assigned condition, the instructions on building the toy block model would be taken from either the toy block set's provided physical instruction manual (i.e., *Manual* condition); or a first-person video recording of someone performing the assembly with overlaid eye gaze information (i.e., *first-person perspective video* or *FPV* condition).

The study uses the *Neon* eye tracking glasses by Pupil Labs [1] as the primary tool for data collection of the experiment. The Neon module is capable of keeping track of fixations [5], and movement data. Additionally, Pupil Labs' own analysis platform, Pupil Cloud, is also used for manual and automated annotation.

The toy block set to be assembled by the participants is a 205-piece model of a two-story cafe building. The Manual condition participants must rely on the toy block set's provided 2-page 15-step paper instruction manual in completing the assembly task (Fig 1). They may flip between the front and the back side of the manual freely.

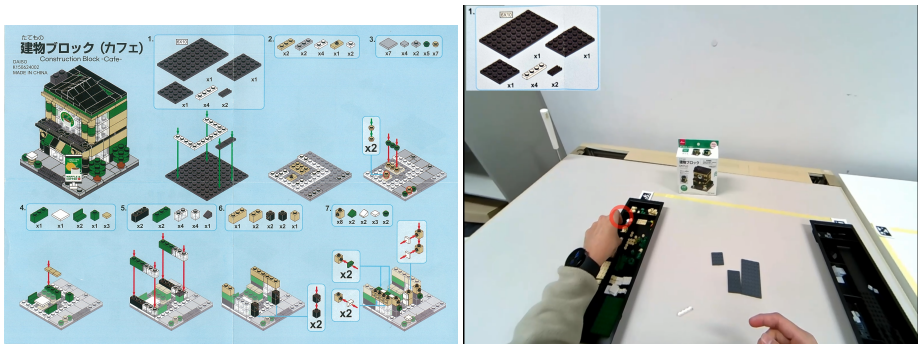

Fig. 1. (left) Manual condition instruction medium. (right) First-person Perspective Video instruction medium.

On the other hand, participants under the FPV condition are provided a laptop with a 34-minute first-person perspective video of the task being accomplished (Fig. 1). The video has been slightly edited to reduce idle times and to eliminate mistakes made in the initial recording. Additionally, the list of pieces to be used in the step is also overlaid on at the top-left side of the video footage–a detail that is taken from the instruction manual itself. Lastly, the gaze of the person in the recording is also illustrated in the video with a red ring. Participants under the FPV condition are free to pause, play, and seek through the video using the laptop's media player.

After the briefing, participants are asked to wear the eye tracking glasses and are given a maximum of two hours to accomplish the task. The assembly period ends once the participant has completed building the toy block set, or if the time limit has been reached.

3 Results and Discussion

A total of 10 university students (from undergraduate to doctoral level) with prior experience in similar assembly tasks, participated in the preliminary study, ranging from 19 to 33 years old (*M*: 22.9, *SD*: 4.20). All participants were able to complete the task within the allotted time.

3.1 Differences in Overall and Step-by-Step Time to Completion

The FPV condition participants have been found to have a longer time-to-completion, possibly due to the fact that they are constrained by the length of the instructional video itself (i.e., 33 minutes and 54 seconds long). This is reflected with the data collected, as shown in Figure 2. The fastest FPV condition participant is participant 2, who completed the task in 62 minutes; and the slowest FPV participant is participant 4, who finished the task in 91 minutes. On the other hand, the fastest and slowest participants under the Manual condition are participants 9 and 8, finishing in 37 and 53 minutes, respectively.

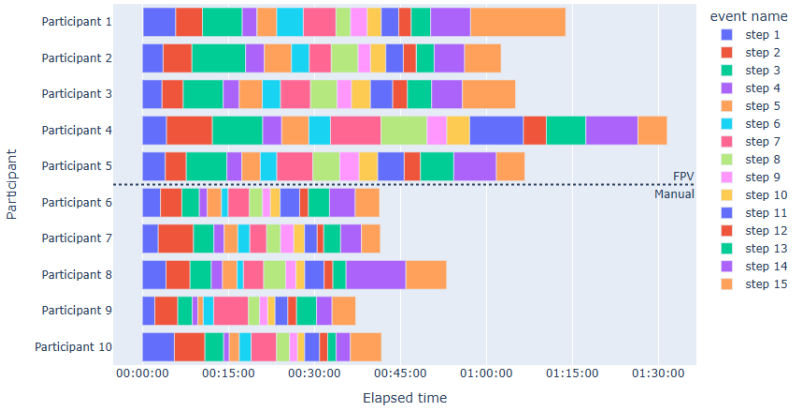

Fig. 2. Step-by-step time to completion for every participant.

3.2 Analysis on Fixation-on-Instruction

The fixation instances of the participants have also been recorded. The resulting data is illustrated in a timeline graph in Figure 3, wherein each solid block represents a fixation instance. Long solid color segments indicate that the fixation instance occurs continuously for a long duration. Whereas, thin color segments indicate that the fixation may be more akin to a brief look towards the instruction manual. It can be seen from the graph that participant 4 does not have as many fixation instances as the other participants. This is due to an error in the recording process, wherein the Neon glasses were not worn on properly, thus preventing the system to detect the participant's fixations accurately. As such, participant 4 will be excluded from further fixation analysis.

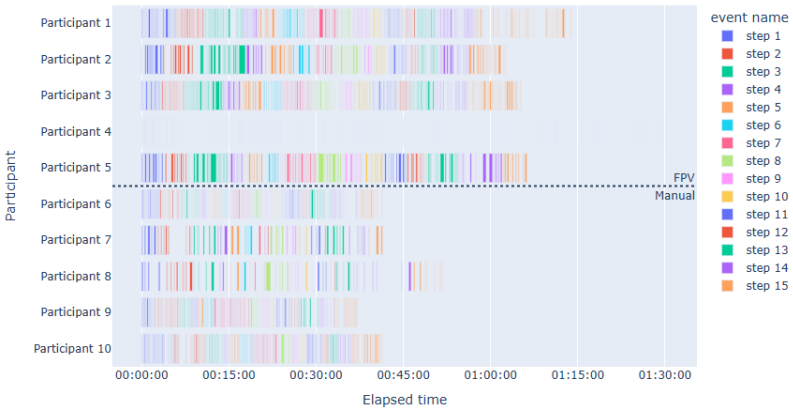

Fig. 3. Fixation-on-instruction instances per participant.

To evaluate whether there is a significant difference between an *independent samples t-test* is performed. Testing at $\alpha = 0.05$, $N_{Manual} = 965$, $N_{FPV} = 1950$, there is no significant difference found in the fixation durations between the FPV condition and the Manual condition participants ($p = 0.97125$), despite the paper manual participants seemingly having shorter fixation durations as seen from figure 3. This indicates that both Manual and FPV condition participants' fixation durations generally last for the same amount of time.

4 Conclusion and Future Work

This paper provided a preliminary look on how the use of first-person perspective video as an instruction material could affect the behavior of participants on an assembly task, as opposed to using a more traditional physical instruction manual. After conducting the experiment on 10 participants, it is found that the time to completion for FPV participants is much longer than the Manual (i.e., traditional) participants. This result is expected, as the FPV participants are limited by the length of the FPV instruction itself. Additionally, the instances of when the participants look into the instructions (i.e., fixation-on-instruction) is also noted. Despite the initial assumption, it has been found that FPV participants tend to look into the instructions more than the Manual participants. There was also no difference in the average fixation-on-instruction duration across both conditions. Possible causes could be the increased cognitive load of having to gather information from a dynamic and transient source as opposed to a static one [20, 21]. However, definite conclusions can not be made yet, due to the insufficient amount of participants.

For future work, it is crucial to have more participants in the experiment, to have sufficient statistical power in conducting hypothesis testing. Further refinement of the FPV instruction is also necessary, to reduce the impact of the video length on the participants' performance. Intentional pauses to serve as *segments* may also be included in the FPV instruction, as it may reduce cognitive workload [7]. Additionally, it might be worth keeping track of the participants interactions with the FPV instruction (e.g., pause, play, seek back and forward), which in turn could lead to more avenues for analysis.

References

[1] Chris Baumann and Kai Dierkes. 2023. Neon Accuracy Test Report. https://doi.org/10.5281/zenodo.10420388

[2] Lucie Bréchet, Robin Mange, Bruno Herbelin, Quentin Theillaud, Baptiste Gauthier, Andrea Serino, and Olaf Blanke. 2019. First-person view of one's body in immersive virtual reality: Influence on episodic memory. 14, 3 (2019), e0197763. https://doi.org/10.1371/journal.pone.0197763

[3] Kevin Chow, Caitlin Coyiuto, Cuong Nguyen, and Dongwook Yoon. 2019. Challenges and Design Considerations for Multimodal Asynchronous Collaboration in VR. 3 (2019), 40:1–40:24. Issue CSCW. https://doi.org/10.1145/3359142

[4] A. Dix. 1994. Computer Supported Cooperative Work: A Framework. In *Design Issues in CSCW*, Duska Rosenberg and Christopher Hutchison (Eds.). Springer London, 9–26. https://doi.org/10.1007/978-1-4471-2029-2_2 Series Title: Computer Supported Cooperative Work.

[5] Michael Drews and Kai Dierkes. 2024. Strategies for enhancing automatic fixation detection in head-mounted eye tracking. 56, 6 (2024), 6276–6298. https://doi.org/10.3758/s13428-024-02360-0

[6] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billinghurst. 2019. Revisiting collaboration through mixed reality: The evolution of groupware. 131 (2019), 81–98. https://doi.org/10.1016/j.ijhcs.2019.05.011

[7] Logan Fiorella and Richard E. Mayer. 2018. What works and doesn't work with instructional video. 89 (2018), 465–470. https://doi.org/10.1016/j.chb.2018.07.015

[8] Guo Freeman, Dane Acena, Nathan J. McNeese, and Kelsea Schulenberg. 2022. Working Together Apart through Embodiment: Engaging in Everyday Collaborative Activities in Social Virtual Reality. 6 (2022), 1–25. Issue GROUP. https://doi.org/10.1145/3492836

[9] Inma García-Pereira, Cristina Portalés, Jesús Gimeno, and Sergio Casas. 2020. A collaborative augmented reality annotation tool for the inspection of prefabricated buildings. 79, 9 (2020), 6483–6501. https://doi.org/10.1007/s11042-019-08419-x

[10] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. In touch with the remote world: remote collaboration with augmented reality drawings and virtual navigation. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (Edinburgh Scotland, 2014-11-11). ACM, 197–205. https://doi.org/10.1145/2671015.2671016

[11] Jonathan Grudin and Steven Poltrock. 2014. 27. Computer Supported Cooperative Work. In *The Encyclopedia of Human-Computer Interaction* (2 ed.). https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/computer-supported-cooperative-work

[12] Jim Hollan and Scott Stornetta. 1992. Beyond being there. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92* (Monterey, California, United States, 1992). ACM Press, 119–125. https://doi.org/10.1145/142750.142769

[13] Lara Sofie Lenz, Andreas Rene Fender, Julia Chatain, and Christian Holz. 2024. Comparing Synchronous and Asynchronous Task Delivery in Mixed Reality Environments. 30, 5 (2024), 2776–2784. https://doi.org/10.1109/TVCG.2024.3372034 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[14] Akos Nagy, Yannis Spyridis, Gregory Mills, and Vasileios Argyriou. 2025. MemoryPods: Enhancing Asynchronous Communication in Extended Reality. (2025). https://doi.org/10.48550/ARXIV.2502.15622 Publisher: arXiv Version Number: 1.

[15] Gary M. Olson and Judith S. Olson. 2000. Distance Matters. 15, 2 (2000), 139–178. https://doi.org/10.1207/S15327051HCI1523_4

[16] 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 *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC Canada, 2018-04-19). ACM, 1–13. https://doi.org/10.1145/3173574.3173620

[17] Maxime Ros, Lorenz S. Neuwirth, Sam Ng, Blaise Debien, Nicolas Molinari, Franck Gatto, and Nicolas Lonjon. 2021. The Effects of an Immersive Virtual Reality Application in First Person Point-of-View (IVRA-FPV) on The Learning and Generalized Performance of a Lumbar Puncture Medical Procedure. 69, 3 (2021), 1529–1556. https://doi.org/10.1007/s11423-021-10003-w

[18] Mona W. Schmidt, Karl-Friedrich Kowalewski, Sarah M. Trent, Laura Benner, Beat P. Müller-Stich, and Felix Nickel. 2020. Self-directed training with e-learning using the first-person perspective for laparoscopic suturing and knot tying: a randomised controlled trial. 34, 2 (2020), 869–879. https://doi.org/10.1007/s00464-019-06842-7

[19] Mickael Sereno, Xiyao Wang, Lonni Besançon, Michael J. McGuffin, and Tobias Isenberg. 2022. Collaborative Work in Augmented Reality: A Survey. 28, 6 (2022), 2530–2549. https://doi.org/10.1109/TVCG.2020.3032761 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[20] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. 12, 2 (1988), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

[21] John Sweller, Paul Ayres, and Slava Kalyuga. 2011. Measuring Cognitive Load. In *Cognitive Load Theory*, John Sweller, Paul Ayres, and Slava Kalyuga (Eds.). Springer, 71–85. https://doi.org/10.1007/978-1-4419-8126-4_6

Received 14 March 2025; accepted 24 March 2025