International Society of the Learning Sciences

"Do you see what we see?" – Perspective-taking across realities

Lucas Paulsen, Jacob Davidsen, Aalborg University,
lpauls17@student.aau.dk, jdavidsen@ikp.aau.dk
Rolf Steier, Oslo Metropolitan University, rolf.steier@oslomet.no

Abstract: Immersive virtual reality (IVR) is slowly emerging as a collaborative technology in CSCL. This form of technology mediation challenges the typical distribution of available resources known from face-to-face interactions. Now, one or more participants may be immersed in IVR, while others are 'left behind' in the physical space experiencing the virtual world on a flat screen. The collaborative activities are thus taking place across realities, and it is important to learn how the participants navigate in this complex setting. With perspective-taking as our theoretical lens, we apply Interaction Analysis methods to the participants' interactional moves in instructing and participating across realities through an excerpt in which participants are engaged in a module on designing with video. Based on the analysis we discuss some of the challenges that emerge for a teacher instructing across different realities, as well as key implications for new forms of pedagogical practices in CSCL.

Introduction

The promise of Immersive Virtual Reality (IVR) for collaborative learning is largely unexplored in everyday classroom settings. One barrier is the lack of designed material for IVR-oriented activities in a classroom setting and a second barrier is that even though high-end Head-Mounted-Displays (HMD) for IVR are becoming more affordable, schools still may not be able to afford classroom sets of high-end HMDs. In addition, IVR has historically been considered an individual tool for cognitive development (Enyedy & Yoon, 2021). An alternative is moving the place of collaboration from within virtual reality to across realities. In this paper, we are not working with collaboration inside IVR, instead we focus on interaction across realities (Steier, 2020), more precisely interaction between a student in IVR and a group of students and their teacher located in the physical surroundings (see figure 1 in Zenodo archive, https://doi.org/10.5281/zenodo.6384030).

As the participants are situated in different realities, it is necessary for the participants to engage in perspective-taking using different interactional resources which are not visible and available for all, as the participant immersed in IVR is effectively cut off from their physical surroundings, only being able to hear the others' voices. Having access to the same resources in CSCL-activities is crucial for attending to a joint space for problem solving (Roschelle & Teasly, 1995). In the triadic relationship between human-technology-human in CSCL (Ludvigsen & Steier, 2019) the participants are often working together around a shared screen or an online platform 'seeing', 'hearing' and 'pointing' at the same resources. However, new technologies like HMDs for IVR creates a new type of shared space – a problem space that remains relatively unexplored in the CSCL literature (see Enyedy & Yoon (2021) for a recent overview).

In this paper we explore how collaborative activities occur across realities and focus on how perspective-taking is accomplished through collaboratively imagining (Murphy, 2005) instruction. By attending to perspective-taking we then look at the ways teachers and students are able to navigate between different realities and overcome the challenges that IVR poses as medium for collaborative learning. We address the following question: *how do participants attend to a joint-problem space across realities through perspective-taking?*

To situate this question, we first zoom in on perspective-taking as a theoretical concept in CSCL. We then present and analyse video data from a CSCL-activity to demonstrate how, in a process of perspective-taking across realities, instruction is collaboratively imagined by the student immersed in IVR and the teacher positioned in the physical space.

Theoretical stance: Perspective-Taking and Embodiment

Perspective-taking as a general mechanism for collaborative learning involves adopting an alternative point of view to further the joint interpretation of a context (Suthers, 2006). Such alternative points of view can be understood broadly to include the perceptual field of collaborators, but also more abstractly as shifts between first- and third-person perspectives (Roberts & Lyons, 2020). Thus, the activity of perspective-taking includes both physical movement in a space as well as conceptual and imaginative work. This imaginative work entails adopting an alternative point of view without necessarily being able to attend to the perceptual conditions of that new point of view. The uneven distribution of resources between participants immersed in IVR and the ones 'left behind' means that the participants' perspectives are also not made available to each other as a shared resource. These

conditions must then be inferred or imagined. Viewing imagining as collaborative implies that imagining is not only an internal mental process, but also an external process occurring in social and material context (Murphy, 2005). Attending to imagining as a distributed socially situated activity allows for imagining to be viewed as an interactional feature. This is in line with a more general orientation in CSCL, namely the shift in attention from meaning making as created through utterances (Stahl, 2006) towards more embodied and bodily-material resources (Davidsen & Ryberg, 2017) in collaborative activities. This line of CSCL research is inspired by Goodwin's (2003) work on embodied co-operative actions focusing on human actors making sense using different modalities of interactions, yet this variety of modalities is not available when collaborating in a setting like the one outlined in this paper. As new technological-pedagogical arrangements emerge, the requirements for perspective-taking also change. We approach perspective-taking analytically by attending to language use, to embodied action, to the availability of and orientation to resources in the participants environment.

Context of this study and method

The collected data is from a study on how students learn to perform Interaction Analysis (Jordan & Henderson, 1995) using a dedicated IVR software called AVA360VR (Annotate, Visualise and Analyse 360 video in Virtual Reality) (McIlvenny et al., 2021). The students are part of a module on the use of "video for design" and have been introduced to Interaction Analysis – including how to collect and transcribe video data. In total 8 groups tried out the software and each group were allocated 1.5 hours to perform their interaction analysis. Each of the groups were given a brief introduction to the hardware and software and they all gave consent to being recorded. For each session, we recorded their interaction with a 360 camera, a 2D camcorder and a screen-recording of the software, allowing for simultaneous viewing of the participant immersed in IVR, participants situated outside of VR and the screen showing the IVR output.

A total of 12 hours of video data was collected, which has been viewed systematically to identify sequences of perspective-taking across realities. In this corpus, we identified sequences in which uneven perceptual experiences across the mixed realities produced confusion or breakdowns between participants. In this paper we present in depth one such 50-second-long sequence to illustrate how collaboration is interactionally organised across realities in a process of perspective-taking where instruction is collaboratively imagined between the student immersed in IVR and the teacher outside IVR. An important feature of AVA360VR in relation to the presented example is that AVA360VR allows for participants outside of IVR to follow-along what the person immersed in IVR is doing in one of two ways on a flat screen: a first-person view where outside observers see what the participant sees and a third-person view in which observers see the larger scene from "behind the head" of the immersed participants avatar. Prior to the sequence the student immersed in IVR, Andy, has accidentally switched from third to first person view for the participants outside of IVR and they are now figuring out how to switch back. The sequence starts when the teacher enters the room to check up on the students. What follows is Andy, Bob and Dylan establishing a joint-problem space that is situated across realities, rather than exclusively in the virtual reality by attending to the physical and virtual perspectives of the technological-pedagogical arrangement. We jump into the sequence just as the joint-problem space has been established and Dylan orients himself towards Andy and begins instructing.

Analysis

Throughout the sequence we will show how instruction is collaboratively imagined between the teacher in the physical room and the student tasked with solving the problem, who is immersed in IVR and therefore not able to perceive the teacher's bodily activity. A transcript of the sequence can be located in Zenodo archive https://doi.org/10.5281/zenodo.6384030, see figure 2.

Dylan initiates the instruction in line 1 by orienting himself towards Andy who is immersed in IVR saying "oh (.) if you (.) if you turn your" in line 2. Dylan then positions himself next to Andy in line 3 and raises his right arm, positioning it so that it embodies the positioning of Andy's right arm. After a 1.2 second pause Dylan elaborates his gesture in line 5 by saying "right controller" marking it as relevant for Andy, who acknowledges the relevancy by saying "yes" and looking down at the right controller in VR in lines 6 and 7. As a joint focus has now been established, Dylan is able to begin the instruction in line 8 and 9 by saying "twisting it ninety degrees" while rotating his right wrist ninety degrees towards himself. As previously mentioned, these gestures performed by Dylan are not visible to Andy as Andy is immersed in IVR. Dylan's gesture shows how being immersed in IVR is a bodily experience, and that the knowledge of manipulating objects in VR is a form of bodily knowledge. Dylan's gestures are then the only way for him to make the instructions available to himself before being able to instruct Andy using speech. This also means that Andy has to imagine the gestures that Dylan is performing, establishing a collaborative process of imagining as Dylan verbally supports his gestures while performing them in the physical space. This allows Andy to take the perspective of Dylan and thereby imagine

the gesture of twisting his right controller. Andy follows the instruction and initially twists the controller ninety degrees to the right. Dylan gazes at the screen in line 10, presumably looking for feedback that the action has been performed. This is further shown in line 12 by the utterance “do you see a picture” and gazing at Andy in line 13. As no visual feedback on whether the action has been performed is shown on the output outside of VR, Andy and Dylan end up overlapping in lines 14 and 15 after a 0.3 second pause as Andy says "yes” and Dylan says, “try turning it the other way”.

Here, the screen is assumed by the participants to be a shared visual resource, yet, in the moment that Andy rotates his right controller ninety degrees to the right what is seen by Andy in VR and what is shown on the flat output becomes incongruent. For Andy a video feed appears next to the rotated controller showing the output produced on the screen in the physical space, yet on the screen nothing but a rotated controller is shown. As the video feed is not available for Dylan, he assumes that the action has not been performed correctly and asks Andy to rotate it the other way. Yet, as is indicated by Andy uttering “yes” the correct action has already been performed.

Meanwhile Andy may also assume that the picture he sees should be visible for the participants outside of VR, therefore does not respond immediately. Andy follows the repair initiated by Dylan, rotating the right controller one hundred and eighty degrees to the left. In line 19, after a 0.5 second pause, Dylan again asks for confirmation that the correct action has been performed, as yet again no visual feedback has been provided on the screen outside of VR, by uttering “do you see (.) do you see what we see” and after a 0.2 second pause in line 21 “so do you see a kind of video (.) picture”. Here Dylan asks for Andy to clarify his perspective in VR as Dylan is unable to assert the feedback provided by the output on the screen. With “what we see” Dylan is not asking Andy to take the vantage point that Dylan possesses in the physical space. He is rather asking Andy to account for his own perspective, asking if he sees the preview of the video output that is shown when rotating the controller ninety degrees, as he elaborates in the second utterance. Meanwhile Andy rotates his controller one hundred and eighty degrees to the right, effectively repeating the first instruction issued by Dylan, and in line 22 says, “i see like my headset”, shaking his head, possibly orienting himself to the output shown in VR, and then “>and then i see myself<” in line 24. This last part overlaps with Dylan saying “yes” in line 25, confirming that Andy has performed the necessary action. Here the act of perspective-taking is made visible to the participants by Dylan asking Andy what he sees, allowing Dylan to determine whether the correct action has been performed even though the output that he is viewing in the physical space does not provide him with the necessary visual feedback. As the twisting action has been performed, Dylan moves on to the next part of the instruction using touch as a mechanism for marking Andy’s left pointer finger as relevant, as he briefly touches it using his right pointer finger and in line 26 says, “and then you press the one (.) the one over here”, and in line 27 “with with erhm:”. Here, the touch allows Dylan to specify what “over here” is referring to, making the action that needs to be performed available to Andy. Andy presses different buttons on the left controller, indicating that he is unsure what action needs to be performed, raising the left controller. In line 30, after a 0.6 second pause Dylan guides Andy towards the correct button saying “beneath (.) yes yes with your pointer finger yes”, which leads to Andy pressing the button with his right pointer finger. Dylan then gazes towards the screen showing the VR output, and after establishing that the output has been changed from being first- to third person says “yes” in line 34.

Discussion

While talk seems to be the most pervasive resource for interaction in collaborative activities across realities our work starts to uncover how important perspective-taking and collaborative imagination are in this pedagogical arrangement across realities. This process of perspective-taking makes the interaction across realities collaborative by nature, as the person immersed in VR needs to make certain objects relevant, hearable and see-able, and the participants situated outside VR need to position the view by framing the orientation of the participant immersed in IVR. This means that even though the classroom scenario proposed in this paper only involves one participant being immersed in IVR, the shown sequence demonstrates how the mediating technology, IVR, still facilitates collaboration through the constraints the technology poses for the interaction.

The sequence also outlines how collaborative imagining can be utilised as a mechanism for organising perspective-taking across realities. This means that even though instruction in classrooms typically is embodied by nature, collaborative imagining allows the participants situated outside of VR to still make their bodily actions available to the student immersed in IVR. While the sequence illustrates how perspective-taking is relevant when solving problems situated across different realities, the pedagogical arrangement of collaborating across realities also shows how the act of perspective-taking can be seen as a relevant learning mechanism. The students are tasked with performing Interaction Analysis, a method for qualitative video-based analysis relying not one specific focus, but entailing multiple foci (Jordan & Henderson, 1995). To avoid the analysis being guided by just one student the group must be able to attend to each other’s perspectives in order for them to build a collaborative analysis. Viewing perspective-taking as a learning mechanism is in alignment with recent literature (e.g., Roberts

& Lyons, 2020), suggesting that inductive methods for analysis, such as Interaction Analysis, may benefit from students attending to each other's perspectives while learning and performing such methods.

Within the context of this problem solving, the practical aspect of learning to use a new tool coincides with a more reflective and process-oriented aspect of learning in order to attend to the role of one's perspective in collaboration. The student immersed in IVR is learning how change from first-person perspective to third-person perspective for the audience, along with the implications that this move has for collaboration. Yet it is also clear that the instructor must learn to provide guidance in this challenging context with implications for future teaching activities in such mixed reality settings.

With this paper we begin looking at how IVR may be integrated as a CSCL technology that is socially situated rather than individually oriented. By attending to IVR as a CSCL technology embedded in a classroom setting where one student is immersed and the rest of the class is following on a flat screen, we can examine the interactional work of instructing and participating across realities. This technology arrangement constrains the amount of mutually available resources that are usually used in establishing and attending to a joint problem space in CSCL. Through our analysis we show how such instruction and participation across realities unfolds through a process of perspective-taking. The use of collaboratively imagining embodied gestures is here shown as a way to accommodate to the constraints imposed by the mediating technology. The integration of IVR into classrooms does hold great potential to create interesting pedagogical activities, but as we show in the analysis the setup is also challenging the means through which participants may attend to a joint problem space. Teachers would need to engage in perspective-taking with the student in and outside IVR to facilitate collaborative learning.

Moving forward, this study also suggests that IVR may prompt new and interesting collaborative learning arrangements in the classroom. In a recent review, Enyedy and Yoon (2021) found that historically IVR has been viewed as an individually oriented cognitive technology, but our study shows that IVR is providing a collaborative environment – with interesting affordances for CSCL scholars and designers. We suspect that these kinds of pedagogical activities will become more common, and it is important that the CSCL community takes a role in shaping such work.

References

Davidsen, J., & Ryberg, T. (2017). "This is the size of one meter": Children's bodily-material collaboration. *International Journal of Computer-Supported Collaborative Learning*, *12*(1), 65–90.

Enyedy, N., & Yoon, S. (2021). Immersive Environments: Learning in Augmented + Virtual Reality. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (pp. 389–405). Springer International Publishing.

Goodwin, C. (2003). The Body in Action. In J. Coupland & R. Gwyn (Eds.), *Discourse, the Body, and Identity* (pp. 19–42). Palgrave Macmillan UK.

Jordan, B., & Henderson, A. (1995). Interaction Analysis: Foundations and Practice. *Journal of the Learning Sciences*, *4*(1), 39–103.

Ludvigsen, S., & Steier, R. (2019). Reflections and looking ahead for CSCL: Digital infrastructures, digital tools, and collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *14*(4), 415–423.

McIlvenny, P., Davidsen, J., Christensen, N. H., Tanderup, S. H., & Kovács, A. B. (2021). *AVA360VR: Annotate, Visualise, Analysis 360 Video in Virtual Reality*. https://vbn.aau.dk/da/publications/ava360vr-annotate-visualise-analysis-360-video-in-virtual-reality

Murphy, K. M. (2005). Collaborative imagining: The interactive use of gestures, talk, and graphic representation in architectural practice. *Semiotica*, *2005*(156).

Roberts, J., & Lyons, L. (2020). Examining Spontaneous Perspective Taking and Fluid Self-to-Data Relationships in Informal Open-Ended Data Exploration. *Journal of the Learning Sciences*, *29*(1), 32–56.

Roschelle, J., & Teasley, S. D. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. In C. O'Malley (Ed.), *Computer Supported Collaborative Learning* (pp. 69–97). Springer Berlin Heidelberg.

Stahl, G. (2006). Group cognition computer support for building collaborative knowledge. MIT Press.

Steier, R. (2020). Designing for Joint Attention and Co-presence Across Parallel Realities. In M. Gresalfi & I. Horn (Eds.), *The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020* (Vol. 3, pp. 1309–1316). International Society of the Learning Sciences.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, *1*(3), 315–337.