Bridging physical and virtual ecologies of action: giving and following instructions in co-located VR-gaming sessions

Margarethe Olbertz-Siitonen & Arja Piirainen-Marsh

Viittausohje:

Olbertz-Siitonen, M., & Piirainen-Marsh, A. (2023). Bridging physical and virtual ecologies of action: giving and following instructions in co-located VR-gaming sessions [Fyysisen ja virtuaalisen toiminnan ekologioiden yhdistäminen: ohjeiden antaminen ja seuraaminen VR-pelisessioissa]. *Prologi – Viestinnän ja vuorovaikutuksen tieteellinen aikakauslehti*, *20*(1), 137–166. https://doi.org/10.33352/prlg.121525

To cite this article:

Olbertz-Siitonen, M., & Piirainen-Marsh, A. (2023). Bridging physical and virtual ecologies of action: giving and following instructions in co-located VR-gaming sessions. *Prologi – Journal of Communication and Social Interaction*, *20*(1), 137–166. https://doi.org/10.33352/prlg.121525

Prologi

– Viestinnän ja vuorovaikutuksen tieteellinen aikakauslehti

journal.fi/prologi/

ruotsiksi: Prologi – Tidskrift för Kommunikation och Social Interaktion
englanniksi: Prologi – Journal of Communication and Social Interaction

Julkaisija: Prologos ry.

Prologos

Avoin julkaisu / Open Access
ISSN 2342-3684 / verkko

Article

Prologi, 20(1)
137–166
https://doi.org/10.33352/prlg.121525

CC BY-NC-SA 4.0

Bridging physical and virtual ecologies of action: Giving and following instructions in co-located VR-gaming sessions

Margarethe Olbertz-Siitonen
FT, senior lecturer
Department of Language and
Communication Studies, University of Jyväskylä
margarethe.olbertz-siitonen@jyu.fi

Arja Piirainen-Marsh
Professor
Department of Language and
Communication Studies, University of Jyväskylä
arja.piirainen-marsh@jyu.fi

received 01.09.2022 / accepted 22.06.2023 / published 18.10.2023

Abstract

Using multimodal conversation analysis (CA), this study examines communicative practices in co-located encounters where participants are getting acquainted with VR-technology and games. The analysis focuses on instructional activities and investigates how an expert player guides others in learning how to handle VR equipment and getting to know the game mechanics during the initial moments of starting a new game.

The data comprises video-recordings of informal social gatherings of 3–4 young adults who take turns in trying out different VR games. The gaming situations were organized in a temporary game lab using consumer-grade VR equipment, a large screen that displayed video feed from the game console and loudspeakers for the game sound.

The analysis demonstrates how the experienced player who has no direct agency over the virtual world uses verbal and carefully placed tactile means to help novice players (who in turn have restricted access to the material world) navigate the initial stages of entering a game. Their complying actions, on the other hand, are characterized by finely tuned bodily adjustments in interplay with affordances of the technology. The instruction sequences, therefore, represent interactional moments in which the participants mutually attend to asymmetries, orienting to bridging physical and virtual ecologies of action.

KEYWORDS: fractured ecology, instructions, instructed action, manual guiding action, multimodal conversation analysis, virtual reality

Introduction

This study discusses how communicative practices are adapted to the affordances for interaction in co-located encounters where participants are getting acquainted with highly immersive virtual reality (VR) mechanics and games. As an emerging technology, VR has gained much attention recently (see Liberatore & Wagner, 2021). Research in psychology and the humanities, for example, addresses questions of presence (see Felton & Jackson, 2022) and emotional effects on user experience (Diemer et al., 2015), or how VR can be successfully implemented in mental and physical rehabilitation (see Howard, 2017; Strong, 2020) as well as in teacher education (see Billingsley et al., 2019). Indeed, the abundance of scholarly work suggests an expectation that VRs are readily applicable to and going to improve almost all areas of social life.

Interaction research, on the other hand, has explored the situatedness and complexity of human conduct in the context of technology-in-interaction. For instance, Hindmarsh et al. (2006) analyze the organization of social interaction in a desktop-based VR and note interactive difficulties in establishing shared reference, prompting considerable adjustments in the ways participants index objects in the virtual environment. Their findings reveal the spatial and local embeddedness of mutual understanding and forecast difficulties for designers "that relate to support for interaction in virtual environments as opposed to issues of graphical realism" (Hindmarsh et al., 2006, p. 812). Conversation analytic work has yielded valuable insights into virtual embodied practices for playful collaboration in virtual worlds. Findings show that participants may (have to) reconfigure properties of the technology for interaction and adapt communicative actions to the material and virtual specificities of the setting. Studies have captured, for example, the situated resources gamers employ (sometimes unsuccessfully) in the organization and coordination of joint activities (Moore et al., 2007), in projecting next actions (Bennerstedt & Ivarsson, 2010), and for peer-socialization of novices (Liang, 2021) in avatar-mediated video games.

However, there is a notable gap in research on interaction that is organized around highly immersive VR, i.e. technologized social situations that involve the use of a head-mount (fully restricting visual access to the physical environment while providing a three-dimensional view of the digital landscape) and other devices that enable interaction with the interface by translating bodily action into the virtual space (see, however, Olbertz-Siitonen et al., 2021). This means that participants' actions and orientations in settings that integrate immersive VR equipment, for example the (interactional) work that is required to enter and engage with VR, remain largely unexplored.

This study addresses this gap by investigating how an expert player guides other participants in learning how to handle VR equipment and getting to know the game mechanics during the initial moments of starting a new game. Using multimodal conversation analysis (CA), we trace the embodied practices that participants deploy in instruction sequences and show how these are skillfully fitted to different overlapping (physical and virtual) ecologies of action in the context of VR-functionality. The instructional activity is organized through sequences of 'here and now' instructions that make relevant a manual/bodily second action – instructed action (Garfinkel, 2002) – that advances the activity of starting a new game (see Deppermann, 2018). We demonstrate that the accomplishment of instructions relies on specific bodily configurations, involving close physical proximity, verbal turns, use of touch and manual guiding actions that attend not only to the different temporalities of verbal and physical activities (Lerner & Raymond, 2021; Mondada, 2016), but also to distinct spatialities and restricted affordances for actions. We further illustrate how the instruction sequences exhibit different agency of the participants and display the local competencies relevant for the task.

By focusing on immersive VR-in-interaction, then, our analysis also extends previous research on co-located video gameplay, involving interaction at the 'boundaries' of in-game activities and being (with others) in the same physical space (see Baldauf-Quilliatre & Colón de Carvajal, 2021; Mondada, 2012; Tekin, 2021; Tekin & Reeves, 2017). It contributes to understanding the organization of participation involving the player's interaction with the device, actions in the virtual world as well as engagement with other co-present participants (see Schmidt & Marx, 2021). A distinctive feature of the gaming set-up in this study is the way that the use of VR technology restricts access to the contextual resources that enable interaction with the game and others in the room. For example, the participants do not share the same visual field – only the player wearing the headset has direct agency over the full range of resources in the virtual space, but at the same time has limited access to the physical environment. The interaction thus takes place in a somewhat fractured ecology (Luff et al., 2003) that constrains the way that participants can establish and maintain intersubjectivity (see also Heath & Luff, 1993). Our analysis illustrates how the participants draw on and adjust available (embodied) resources to attend to communicative asymmetries and thus achieve mutual orientation to the task at hand.

In the next section, we introduce the data and method, after which we outline our approach to instructions and instructed action as social accomplishments. Following this, we present our analysis of instances of instructed action that are situated in consecutive moments, or steps, through which participants interactively achieve the start of a new game. We conclude by discussing our findings and their implications, for example, in terms of usability and UX design, but also regarding the social embeddedness of VR functionality.

Data and method

We draw on video-recorded instances of co-located VR-gameplay involving university students and staff (researchers). The data was collected in 2018 in a temporary game lab using PlayStation VR (Olbertz-Siitonen et al., 2021). While the equipment allowed only one person to enter the virtual world, the arrangement of the game lab included a large external screen as well as loudspeakers providing other participants with audio-visual access to in-game events. The dataset comprises four two-hour gaming sessions amounting to a total of approximately 500 minutes of recorded material. The number of participants fluctuated in each session, as they arrived and left at different times. However, altogether 15 people participated in the tryouts (3 researchers and 12 players). All participants gave their permission to utilize the collected material in research and publications.

To document the activities in both the physical and the virtual environments, two cameras were set up to capture different views of the game lab in addition to automated recording of the video feed of the game that was displayed on the big screen. Thus, the material provides three angles: a view of what was happening in the game, a focused view of the player, and a wider view of the room in general (showing, for example, other participants sitting behind or beside the player, fig. 1).

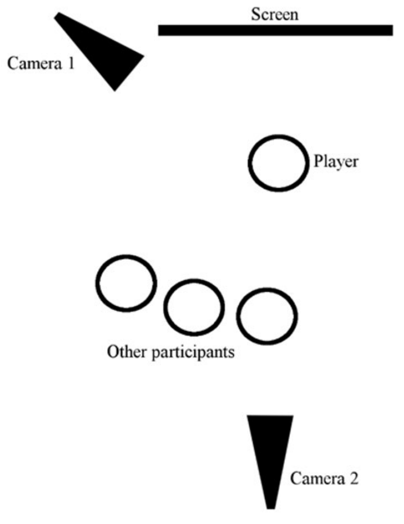


Figure 1. Setup of the game lab

While some were experienced gamers, all attendees had little to no previous experience with PlayStation VR and can be considered novices in operating parts of the equipment (such as the headset) as well as navigating the virtual game reality. This aspect is visible in the ways they handle the gear and enter different games. Indeed, one objective of these organized gaming events was to gain insights into how the participants together get acquainted with the technology and orient to playing games that are new to them in terms of content and possible immersion.

The participants took turns in trying out various games. The sessions were thus characterized by a one-at-a-time turn-taking system where each participant waited for their turn while only one had agency in the virtual world (cf. Carlin et al., 2021). Moments of switches between players emerged as particularly interesting for further analysis, as these instances made relevant different forms of assistance, revealing the situatedness of giving and following instructions within and across physical and virtual ecologies of action. Our analysis is based on a collection of 40 switches between players that included instruction sequences where the expert player (one of the researchers) provided guidance on how to handle the VR equipment and game mechanics. For this study, we selected 5 examples that are representative of practices deployed in the situated accomplishment of instruction sequences in their sequential environments. The data were transcribed applying the principles of CA (see Jefferson, 2004) and the conventions for transcribing embodied conduct developed by Mondada (2018, see appendix). Visual phenomena and written instructions provided by VR technology are presented in the transcript when relevant for the analysis. The participants spoke English and Finnish during the sessions. In the case of Finnish, an English translation is included in the transcripts.

The study applies multimodal CA (Goodwin, 2000; Mondada, 2016; 2019) which is grounded in the ethnomethodological question of "(…) how do social actors come to know, and *know in common*, what they are doing and the circumstances in which they are doing it" (Heritage, 1984, p. 76; our emphasis). Multimodal CA focuses on the mutual accomplishment of interaction, that is, the embodied communicative practices (gestures, gaze, talk, prosody, posture, etc.) through which interactants make their understandings available to each other and thereby achieve intersubjectivity – concurrently as well as sequentially (see Mondada, 2016). In the exclusive consideration of observable, embedded and contextual behavior lies an important strength of this approach for communication studies. CA draws on what becomes witnessable in the recordings of naturally occurring data (Schegloff & Sacks, 1973), and therefore studies *actual* (instead of accounts of) communication. This enables us to concentrate on situated action as the site for understanding social and communicative processes in a technological setting, paying detailed attention to the temporality, materiality and interactivity of communicative practices involving VR technology. Our analysis illustrates the practices deployed in the moments leading to starting a new game and shows how the sequential steps of getting equipped for play, managing game mechanics, and interacting 'in VR' are assembled through the participants' use of language, bodily conduct, objects, and space.

Instructions and instructed action

Starting a new game requires practical knowledge and expertise in using the device (how to handle and operate the controller, how to react to requests of the system), adjustment to a new spatial surround (virtual space, new visual field) and learning how to see and act in the new environment (see Nishizaka, 2006). The joint production of the sequential steps through which a new game is launched involves instructional sequences, where the expert player guides each new participant through the necessary activities by telling and showing them what to do next. These sequences are interactionally accomplished through the deployment of instructions coupled with manual and bodily actions (gestures, manual actions involving objects), and the embodied actions through which the recipients respond to them. Manual activity is crucial to the task at hand; getting the game started involves coordination of bodily conduct and timely accomplishment of actions in ways that are sensitive to the specific situation (Heath et al., 2018; Lerner & Raymond, 2021).

We approach instructions as part of a family of directive actions, requests and directives taking different forms, that make relevant a complying second action (Deppermann, 2018; Lindwall et al., 2015). The instructions provide resources for understanding what happens next in the current phase of activity, what to do with the controller or where to look and what to pay attention to in their visual field. Like other situations that involve physical activities and skills (see Keevallik, 2010; Lindwall & Ekström, 2012; Nishizaka, 2007, 2011), instructions in the game lab are performed multimodally and often involve multiple resources including manual and other physical guidance. The recipients show their understanding of the instruction – and the social and material environment – in the way they attend to it in their next actions, accomplished through bodily actions with or without co-occurring talk.

Previous research has shown how instructions are embedded in the ecologies of practical activities such as driving (De Stefani, 2018), handicraft (Lindwall & Ekström, 2012) and cooking (Lilja & Piirainen-Marsh, 2022) and emerge from specific local circumstances in these activities. In our data, the organization of instructions and the actions projected by them are sensitive to the specific spatial and material arrangements of the game lab, which provide asymmetrical access to the VR technology and restrict the affordances for action available for different participants. As the analysis will show, restrictions of visual access that are experienced by the player once they are wearing the headset, necessitate the use of verbal descriptions and tactile engagements with the devices to negotiate how an instruction is to be understood and how the projected second action can be accomplished. The way that different resources are made relevant and mobilized depends on the position of the instruction in the trajectory of the activity and the situational contingencies that the participants experience.

Recent research on the linguistic and embodied design of instructions has shown that the choice of action formats depends on the temporal relevance for the response (immediate, *here and now* vs. later response), the participants' displayed competences related to prior knowledge (epistemic status) and interactional history (Deppermann, 2018; de Stefani, 2018). Instructions are recipient designed so that they are sensitive to the relative expertise and skills of the participants in performing the task or activity. In our data, the experienced player draws on his expertise in the domain of VR and games in guiding the players, some of whom are experienced players but have little or no prior experience of VR.

Like other task-centered activities, setting up and interacting with a game involves practical, manual and bodily know-how which is essential for successful play. In the context of VR, even experienced players need to adjust to a new environment and develop a new kind of *mastery of the body*, knowing in action (Ekström, 2012; Schön, 1987) to manage the game. In the game lab the participants take turns in trying out the technology, which enables those who are not using the device to observe the interactions that occur between the expert and the current player and develop at least an initial understanding of what is involved in starting the game. In the following sections we elucidate how the multimodal design of instructions is sensitive to the varied expertise and degree of common ground that participants display in their talk and actions.

Another crucial aspect of the technological setting that impacts the participants' project of starting a new game is the way they attend to the visual world that the player experiences through the headset and the other participants see on a large screen. The VR technology design provides instructions in written form and in the form of visual images and prompts that are meant to guide the user in how to proceed and make choices from the available options/games. The participants orient to this in their conduct in different phases of preparing for and negotiating entry into the game. Although the specific design of the virtual environment offers resources for understanding how to use it, the guidance provided by the expert player is often crucial for the practical accomplishment of the instructed actions. This is visible, for example, in the way that the instructions attend to specific ways of using the devices to locate relevant features (e.g., visual instructions and cues in VR) and carry out actions that are necessary for starting the game.

Analysis and findings

In this section, we elucidate the interactional character of instructions and their situated multimodal accomplishment during the initial moments of starting a new game. The analysis is organized into three sections that illustrate how instruction sequences are embedded in and constitute three different activity environments: (1) passing of the controller(s) to a new player and thereby equipping them for action, (2) establishing controller functionality and calibrating the game, and (3) managing in-game mechanics. These environments are representative of the sequential steps involved in managing entry into a new game. The order in which they occur and the way they are realized depends on the local circumstances, such as the game of choice and the hardware required, as well as the expertise shown by the participants in handling the equipment. As the analysis will show, the way instruction sequences are assembled is sensitive to these circumstances.

In negotiating the sequential steps, the participants manage the intersecting organizations of giving and following instructions and manipulating objects. This makes relevant specific kinds of configurations and practices that attend to the different spatialities involved and the manual activities (or manually organized courses of action, see Lerner & Raymond, 2021). In what follows we show how the experienced player who has no direct agency over the virtual world uses verbal and carefully placed tactile means to help novice players (who in turn have no visual access to the material world) navigate the initial stages of entering a game. The examples illustrate the configurations that enable actions to be recognizable as instructions and the practices through which they are accomplished and understood in the three different environments. First, we show how instructions enable the accomplishment of object transfer in preparation for starting the game. This involves various adjustments to the material ecology involving restricted access to visual resources. After that we focus on configurations that are deployed when establishing controller functionality and calibrating the game. This section highlights the importance of tactile resources for making instructions recognizable while managing the different spaces. In the third section we turn to instruction sequences that are related to managing in-game mechanics, which rely on close coordination of virtual and physical resources for interaction.

1) Equipping the player for action – passing of objects

At the beginning of each game session, the expert player (Max) addresses the group of participants and gives general information about VR and the games available. After this he invites one of the participants to take the first turn trying out a game. Each time a new player takes over, they take position facing the large screen, put on the VR headset and wait for Max to hand over the controller(s). This involves a series of instruction-giving sequences through which the participants negotiate control over the devices and how to use them to navigate the virtual space. The ways in which these sequences are assembled through the deployment of verbal and bodily resources show how the focal participants attend to changes in the perceptual environment, specifically, lack of shared visual field, and their relative opportunities for action. The expert player's instructions that project handling the devices, structure and support the player's transition from being a member of the group of co-present participants to an active player with ability to act in the virtual ecology.

This section examines instructions that occur in situations where the participants renegotiate their agency by handling the devices and moving them from one participant to another. The analysis focuses on the moments leading up to and accomplishing object transfer (Tuncer & Haddington, 2019), i.e., handing over the controller(s) to the player-to-be to enable them to use the controller(s) in starting the game. The situations in focus involve intersecting organizations: instruction sequences that intertwine with bodily initiation and achievement of object transfer or manipulation of objects.

The first two examples illustrate (a) how the participants orient to the lack of shared field of vision and attendant change of the perceptual environment once the player has put on the VR headset and (b) how they accomplish object transfer in the circumstances where the player-to-be has restricted visual access to the objects around them. The first extract begins just after the expert player, who is standing holding the controller, has switched the PlayStation VR on. Prior to this he has already attended to the player's (Ari) displayed need to adjust the headset. Ari continues to adjust the headset through lines 1–6, where Max now orients to the visual field that Ari can access through the headset but that is only partially available to the others. The screen shows a text that instructs the player to position themselves "1.5–2.0 meters from the camera". Max's verbal turn (l. 1–2, 4) comments on what is visible to the player wearing the headset.

Extract 1

```
1  Max     *+eli niinku nyt sulle n:äkyy siellä* (.)
           so like now you can see there
           * gaze to large screen-->           *glances twd A
   ari     +adjusts headset--->

2          *pitäs näkyy (.) eiks vaa
           should see (.) can't you
           *gaze to large screen-->

3  Ari     >kyl*lä<
           yes
            -->*

4  Max     *(sä oot) su- suunnillee [siin keskellä] ja
           (you're) rou- roughly there in the middle and
           *walks to the other side of A -->

5  Ari                              [näen itseni  ]
                                    I can see myself

6  Ari     kyllä
           yes

7  Max     *ja tossa mä (.)
           and there I
           *turns to face A holding controller w/both hands

8          *jos sä aje- an-# ojennat #kä+tes eteenpäin (.)
           if you gi-  giv- reach your hands forwards
           * releases LH, holds controller w/RH --->
   ari                               -->+ +extends BH, palms up
                          #fig2     #fig.3
```

(Figure 2)

(Figure 3)

```
9          *nii mä annan sulle +#ohjaimen (.)
           I will give you the controller
           *places controller in A's hands
  ari                          +takes controller
                                #fig.4
```

(Figure 4)

The turn-initial particle ("eli", l. 1) marks a shift from the previous activity to a focus on the visual field. The following turn-constructional unit (TCU) is addressed to the player and designed as a verbal report on what the player can be expected to see through the headset. While gazing at the screen, Max refers to the visual phenomena mediated by the headset, thereby orienting to the divergent perceptual fields. Linguistically the turn is formulated with the verb in passive form so that it claims knowledge of something that is visible to the player but does not specify what (*to you is visible*). With this design, Max explicates a specific kind of professional vision by doing "seeing by proxy" (Carlin et al., 2021). During the micropause (l. 1), Max briefly shifts his gaze from the screen towards the player (possibly to check whether he is positioned correctly). This is followed by a self-repair whereby Max downgrades his epistemic stance by reformulating his utterance as a candidate description of something that *should be* visible (l. 2). This understanding is confirmed when Ari provides an affirmative answer to Max's tag question (l. 3).

After this, Max begins to orient to the next stage of the transition: calibration and object transfer. He refers to Ari's position as *roughly in the middle* (l. 4), which Ari takes as a request for confirmation and confirms that he is able to see himself (l. 5–6). Concurrently with this exchange (l. 4–5), Max walks from one side of the player to the other, at the same time moving the controller from his left to the right hand. He takes position facing Ari and extends the hand holding the controller slightly so that it is placed in front of Ari, while verbally indexing the action with a deictic adverb (*there*, l. 7). With this multimodal action Max attempts to move the interaction forward by proffering the object to Ari.

Previous research has shown that bodily actions such as walking and manipulation of objects in the physical environment can project next actions (Mondada, 2014). In this example, the action of passing the controller becomes recognizable as soon as Max stops close to the player, and the extension of the hand can already be seen as the *preparation phase* of the object transfer (Lerner & Raymond, 2021). However, these actions are not visible to the player, who at this point still has his hands on the headset and does not react immediately. Max attends to the momentary delay by holding the object in place and formulating a verbal instruction in the form of a two-part declarative statement which specifies what the player needs to do (reaching hands forwards) to enable him to pass the controller (l. 8). Ari demonstrates understanding of the instruction in progress during the self-repair of the verb (l. 8) by extending both his hands in front of him with the palms up in readiness to receive the object. The object transfer is accomplished smoothly concurrently with the latter part of Max's utterance (l. 9) when Max places the controller in Ari's hands, releases it and withdraws from the face-to-face arrangement.

Extract 2 provides another case where the participants realign themselves and move the interaction forward by accomplishing object transfer concurrently with explicit verbal instruction. Max has switched the PlayStation on and suggested a game featuring an enormous beast. He describes the game to the whole group whilst the next player gets ready and adjusts the headset. Max closes his multiunit turn with an assessment "it's very interesting" during which he switches the controller from his right to the left hand. He then takes a step toward Jan, holding the controller in the slightly extended hand and verbally marks the transition with "so uhm" (l. 1, fig. 5). Next, he produces a verbal request instructing Jan to hold out his hands while holding the controller in front of him at chest level (fig. 6).

Extract 2

```
1  Max     *#^so uhm (.) #please hold out your hands
           *steps twd J, controller in LH, extended --->
   jan       ^sits still, hands on legs
   fig      #fig.5       #fig.6
```

(Figure 5)

(Figure 6)

```
2          (0.5)^#(0.2)
   jan          ^...extends both hands, palms up
   fig           #fig.7
```

(Figure 7)

```
3          >nosta #*käsiä mä  *#annan tän ohjaimen täältä<
           raise your hands I'll give you the controller from here
                   *lowers LH *places controller in J's hands
   fig            #fig.8       #fig.9
```

(Figure 8)

(Figure 9)

Jan does not react immediately, but begins to raise his hands, both palms up, at the end of the 0.7 s. silence (l. 2, fig. 7). Concurrently with this, Max reformulates the instruction in Finnish (l. 3). At the point where Max announces that he will *give* the controller, Jan already has his hands positioned in readiness to receive the controller (fig. 8), and the object transfer is completed by the end of the verbal turn.

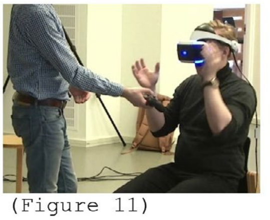
These extracts show how the participants' orientations to the technological objects and their asymmetrical access to the visual field and material environment matter for the organization of interaction. The participants' verbal and bodily-visual actions display their relative agency and competence in manipulating the devices. While Max unproblematically adopts the role of expert and instructor, it takes collaboration and careful coordination of verbal and bodily resources for the participants to negotiate the passing of the controller to the player to equip them for the next step. Although passing of the controller is to be expected for the play session to begin, the actual exchange and its timing needs to be negotiated in a way that is sensitive to the player's restricted visual field. Max's bodily visual actions (walking, adopting face-to-face position towards the player, the way he holds the controller) that project passing of the device are not available to the player (although may be partially perceived by them). The preparation phase leading to object transfer therefore requires additional work, in this case explicit verbal requests, to ensure collaboration. On the other hand, the players' bodily actions whereby they show readiness to receive the controller are performed (partly) simultaneously with the verbal turns. This illustrates the social organization of "manually-implemented courses of action" (Lerner & Raymond, 2021, p. 278) more generally. This type of action often involves reciprocal bodily actions by two participants and need to synchronize these actions in a timely way so that as soon as an initiating action (such as proffering an object) is recognizable, the recipient rearranges the body in readiness for the next action and thereby shows understanding of the first action (as in positioning of hands in readiness to receive the object).

Extracts 1 and 2 illustrate exchanges between the expert and the first player in the session. However, object transfer and the realignment of participation can also take place in a routinized way without explicit verbal instruction. Extract 3 depicts how passing of the controller to a next player (Ari, see Ex. 1) is achieved simultaneously with Max's engagement in talk with other participants.

Extract 3

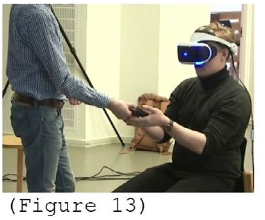
```
1  Max      *itse #tota se  *olento on kyllä aivan*(0.7#)+
            the actual like the actual creature is really
            *...steps twd A *turns                 *ext. RH -->
   ari      adjusts headset -- >                         +...
   fig           #fig.10                           #fig.11
```


(Figure 10)


(Figure 11)

```
2           +#todella hieno siinä (0.6)#+sii*nä tota
            truly amazing there        there like
            -->                         -->* *releases contr.
   ari      +lowers hands palms up       +takes controller
   fig       #fig. 12                   #fig.13
```


(Figure 12)


(Figure 13)

```
3           (*0.7) VR-maailmassa ku se tulee nii lähelle (0.5)
                   in the VR-world when it comes so close
             *moves away from player --->
```

During the moments before the start of this exchange, Max has shown the choice of games available to the group and Ari has put on the headset, which he continues to adjust through line 1. Max is responding to another participant (A), who has commented on the games. While he comments on the beast in one of the games, Max concurrently takes a couple of steps towards Ari (l. 1, fig. 10). Standing next to him, he then extends his hand holding the controller in front of Ari (fig. 11), who reacts to this without delay by releasing the headset and beginning to lower his hands. As soon as he adopts the ready-to-receive configuration with both hands extended (l. 2), Max places the controller in his hands (figs. 12 and 13), releases it and moves further away, while continuing his verbal turn addressed to A.

In sum, extract 3 illustrates how two distinct organizations of conduct – the passing of an object to equip the player and verbal description of a virtual experience – intersect and can be managed without overt trouble (see Raymond & Lerner, 2014). Unlike extracts 1 and 2, here Max's bodily conduct – preparation for the handoff by moving physically close to the player and taking position facing him – is enough to make the proffering of the object recognizable to the recipient and the passing of the object is achieved smoothly without talk. In our data this way of managing object transfer is found in situations where the player-to-be is not the first person to take a turn with the VR. This suggests that the participants have developed a sense of an emerging routine in the way that events are ordered in the session (cf. Deppermann & Pekarek Doehler, 2021). Having observed how previous players are walked through the steps of starting the game, they are able to recognize actions that progress the activity without explicit verbal instruction.

2) Establishing controller functionality and calibrating – getting ready to play

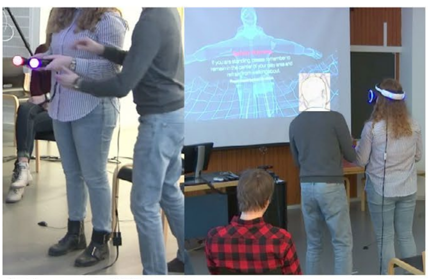
After passing the controller to a new player, its functionality may become relevant for the participants – especially during the first rounds of trying out the VR. The instructional activities that are embedded in and constitute this sequential moment of getting ready to play are carefully designed, attending to different competencies of the players, the technology, as well as to the fractured ecology. While in some cases in our data the functionality of the controller(s) is dealt with very briefly, assuming prior knowledge about or a certain familiarity with handling different types of console controllers, extract 4 is representative of instances that involve embodied interaction in close proximity, including guiding and corrective touch coupled with detailed descriptions. The example thus illustrates the kind of bodily instruction that is necessary for a novice player to negotiate the step of getting ready to play in VR.

Prior to the following fragment (Ex. 4a), the new player, Tea, has equipped and adjusted the headset and received the controllers from Max who placed them in her extended hands (ready-to-receive, see Ex. 3) after he selected a game ("Everest"). After that, Max carefully repositioned the player in the room by giving verbal instructions while attending to the cable of the headset. Right before the transcript starts, a calibration image of a person becomes visible on the screen together with a safety warning and an additional instruction stating, "Press move button to close." (fig. 14), which Tea reads out loud. Max, who is positioned slightly behind Tea and looks over her left shoulder, initiates a shift to a new activity (with the turn-initial particle *so*), directing attention to "these controllers" (l. 1).

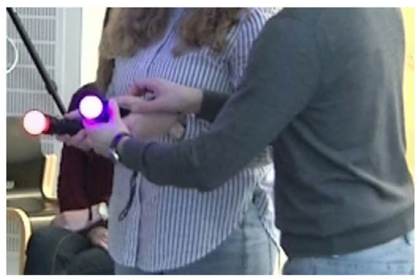
Extract 4a

(Figure 14. Safety Warning / "Press move button to close.")

```
1   Max      *so. (.) #these *these control*#lers¿
             *...............*touches T's LH w/RH, and L controller w/LH
                                         *slightly lifts T's L thumb
    fig               #fig.15               #fig.16
```



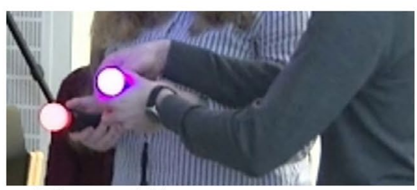
(Figure 15)



(Figure 16)

```
2            *they have #here¿     * (.) *the kind of [#trigger.]
             *turns L controller R*      *......points at trigger--->
    fig                 #fig.17                      #fig.18
```

(Figure 17)



(Figure 18)

```
3   Tea                                               [yea.     ]

4            I can see what you (are doing.)*
    max                                  --->*

5   Max      *yeah.
             *removes hands

6   Tea      uhuh,
```

In addition to establishing the new focus verbally, Max moves both hands towards Tea's left arm, until they reach her hand at "these". He lightly takes hold of the motion controller with his left hand and at the same time touches the top of Tea's left hand with his right hand (figs. 15

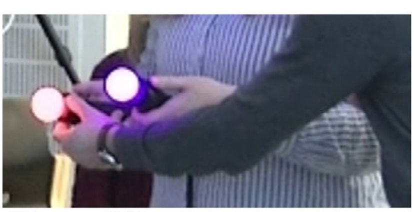
and 16). At "controllers" he then carefully lifts Tea's thumb with a pincer grip, loosening the grip of her hand around the motion controller, and continues to explain a feature of the controller, pointing out the location of a "trigger" (l. 2). This is done in combination with turning the motion controller in her hand to the right (while holding Tea's thumb in place) and then pointing at the trigger with his right index finger (figs. 17 and 18). Tea verbally displays her understanding (l. 3–4), which is acknowledged by Max (l. 5).

By taking hold of her hand and lifting her thumb (l. 1), Max projects complex manual guiding actions that involve manipulation of the device, thereby orienting to Tea as a novice in handling motion controllers. Max's embodied actions constitute a finely tuned ensemble of multimodal activities (a "complex multimodal Gestalt", Mondada, 2016, p. 344), and they also respond to Tea's reaction (l. 4–5) without delay. For example, precisely at the end of the indexical "here" in line 2, the motion controller has been turned far enough to make the trigger recognizable, and at "this kind of trigger", Max's index finger reaches the location of the button in question. This requires carefully timed preparation (e.g., in terms of the anticipatory turning of the controller). His actions thus observably orient to the progressivity of the current activity. When Tea ultimately claims understanding, he immediately releases her hand from his touch (l. 4–5).

Having indicated the location of the trigger on the left motion controller, Max next underlines the relevance of the button, describing it as the main button for interacting with the PlayStation. At the same time, he manually directs the attention to the other motion controller (which also has a trigger) resting in Tea's right hand, as he takes hold of it and then slightly pulls it away from the player's body (without removing it from her hand), forecasting some corrective action (l. 7, figs. 19 and 20).

Extract 4b

```
7  Max     *and the trigger is the *#m↑ain button *you will be #using.
           *......................*touches R controller w/ L hand--->
                                                 *adjusts controller
   fig                             #fig.19                   #fig.20
```


(Figure 19)


(Figure 20)

```
8          (0.6) [I think.   ]

9  Tea           [•#this one.]
                  •points with R thumb
   fig             #fig.21
```

```
10 Max     no no. •#the- *with the index finger.
                         *touches T’s R index finger w/ L index finger
   tea             •lifts R thumb
   fig              #fig.22

11         *so here. (0.5) [on the  * *underside.]
           *directs T’s index finger* *withdraws LH from R controller

12 Tea                     [uh (yeah I          ]*see.)
   max                                           *nods, walks
                                                 backward--->

13 Max      (0.7) yea.*
                  --->*
```

(Figure 21)

(Figure 22)

Simultaneously with Max downgrading his epistemic authority in line 8 (“I think.”), Tea indicates a problem with locating the correct button on the right controller (l. 9): she offers a candidate understanding (“this one.”) and points at a nearby button with the thumb, thereby indexing the referent while assuming a shared focus (fig. 21). This prompts corrective actions involving talk and (guiding) touch by Max (l. 10–11) who thus treats Tea’s understanding as inaccurate and in need of repair. Max verbally rejects the proposed option, which Tea immediately ratifies by lifting the thumb (l. 10). He then moves on to locate the trigger, first by clarifying which finger (the index finger) should operate the button in question and next by drawing attention to the bottom of the controller. This explanation is coupled with cautiously placed and finely coordinated manual actions directing Tea’s index finger to the right place: when he announces, “with the index finger” (l. 10), he touches Tea’s left index finger (with his index finger reaching around the controller from below, fig. 22) and then guides it to the trigger as he refers to the underside of the device with the indexical “here“ (l. 11). This is acknowledged by another verbal display of understanding from Tea, produced in overlap with Max’s further clarification (“on the underside”), which leads to an embodied closure of the corrective instruction (Max removing his hands, nodding, and moving away from Tea).

Max now gazes at the screen, which still features the safety warning and the instruction “Press move button to close” (fig. 14). He then initiates a new instructional sequence by moving back to Tea and beginning to locate the button indicated in the VR instruction (l. 14).

Extract 4c

```
14 Max      *(0.6) u:m (0.6) *and um *and then    *#this button¿
            *gaze at screen  *moves back to T, gaze at T's LH
                                    *...........*lifts T's L thumb
                                                with pincer grip (LH)
   fig                                           #fig.23

15          (0.6) *here-                  *#if you look down?
                  *touches controller *...both hands to T's LH, gaze at
                                           screen--->
   fig                                     #fig.24
```


(Figure 23)


(Figure 24)

```
16          (0.6) look at your hand¿
            --->

17 Tea      •ø(1.8)
            •turns head downward
   VR        øimage of the controllers becomes visible

18 Max      *#u:m (0.9)
            *gaze at T's LH, lifts T's L thumb with pincer grip (w/ LH)
   fig       #fig.25

19          *here is actually a button [#here  ] in the    *middle    *
            *places T's L thumb on button and points w/ it *pushes
                                                            button w/
                                                            T's L thumb*
   fig                                #fig.26

20 Tea                                [mm hmm¿]
```


(Figure 25)


(Figure 26)

```
21 Max      =*øthat's like move button.
             *moves backward, looks at screen
   VR         øimage changes

22          •ø(1.5) .hh
   tea      •turns head back up
   VR        øimage of person with stretched-out arms comes back into
             view; text ("hold both arms out…")
```

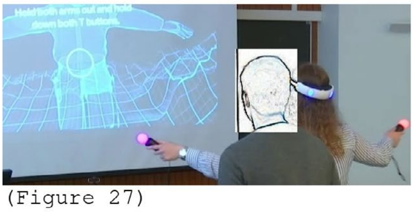
Like before, Max draws on tactile resources to refer to relevant features of the controller and to locate the next button (l. 14, fig. 23; l. 18–19, figs. 25 and 26). However, he now also involves the VR perspective by asking Tea to look down (l. 15). When Tea does not respond, Max updates his instruction and specifies what she should look at (l. 16) thereby treating her non-responsiveness as a problem related to the design of the prior turn and initiating self-repair. Now Tea displays understanding, turning her head downward, which brings a virtual representation of the motion controllers into view. Next, Max lifts Tea's left thumb (fig. 25) and uses it to not only point at the button while describing its position, but also to press it (l. 19, Fig. 26). Pushing the button in the material ecology lights up the corresponding button on the virtual controller at the same time, making it possible for Tea to see the feature in the virtual world. Max thus provides her with the haptic and visual experience of pressing the button and manually ensures that Tea follows the VR instruction allowing them to move forward. Only after that he reveals that this is the move button (l. 21). The action of pressing the button with Tea's thumb responds to the VR command (fig. 14) and prompts a change to a new instruction ("Hold both arms out and hold both t buttons"), which becomes apparent when Tea lifts her head to face forward in the VR (l. 22).

Throughout lines 1–22, Max thus orients to Tea's restricted visual access to the material ecology by employing finely tuned manual guiding actions that enable Tea to locate relevant buttons and carry out the actions required to proceed. However, Max also orients to Tea's visual field by accommodating the visual representation of the motion controllers, adapting the affordance of the technology for establishing mutual reference (and thereby instructing Tea where to find the move button for future use).

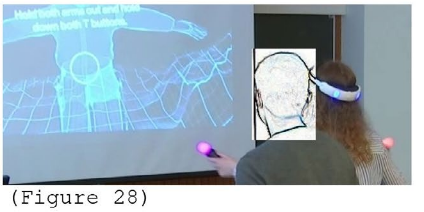
Next, Tea demonstrates learning as she is now able to operate the controllers in response to the new request by the VR – without requiring further manual instructions by Max (Ex. 4d). Nevertheless, Tea displays some uncertainty prompting Max to step in again and provide verbal guidance in collaboration with the VR, mediating between the virtual and material world. Both participants begin to read out the text that has appeared after pressing the move button together, and Max completes the turn (l. 23–25).

Excerpt 4d

```
23 Tea      [hold both]

24 Max      [hold both] arms •#out
   tea                       •holds both arms out--->
   fig                        #fig.27

25 Max      and hold• [both ] t buttons¿
   tea          --->•

26 Tea                [•#uh-]
                       •brings arms quickly back in
   fig                  #fig.28
```

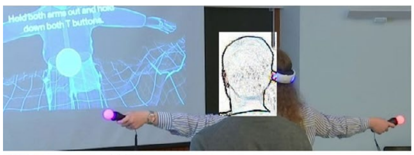


(Figure 27)



(Figure 28)

```
27 Max      yeah- do •do what (.) what he (.) sh:ows.
   tea               •holds both arms out again--->

28          yes. and then: the triggers.
   tea      (arms)--->

29 Tea      •(1.5)
            (arms)--->
            •holds down the trigger buttons--->

30 Max      ø(for the-) (1.3)#     ø
   tea      (arms/triggers)--->
   VR       øcircle fills up, blingø
   fig                      #fig.29
```



(Figure 29)

```
31 Max      øyeah¿
   tea      (arms)--->
   VR       ø"calibration complete"

32 Tea      •°calibration complete.° •
            •.....brings arms back in•
```

While Max reads the end of the first part of the instruction, Tea begins to comply, stretching out her arms in accordance with the VR command (l. 24). However, when Max continues to read the second part of the instruction, (l. 25), Tea quickly retracts her arms indicating difficulties in understanding what exactly is expected of her. Max responds to this display by encouraging her to indeed "do what he shows" (l. 27), now drawing on the image of a person holding out both arms that is visible on the screen. Concurrently with this, Tea repeats the action of stretching out the arms. She thus visibly orients to the progressivity of the current activity and displays understanding. Max accepts her alignment verbally and by attending to the next relevant action ("and then the triggers") (l. 28), which clearly builds on the previous instructional work of locating the triggers on the motion controllers, assuming knowledge of where they are and which fingers to use to operate them. Tea in turn demonstrates understanding and displays acquired competence in handling the controllers: she holds down both triggers until the circle in the middle of the image fills up, which calibrates the system and renders the game ready for play.

In sum, Extract 4 illustrates how touch, verbal instruction and manual action help locate features of the controllers and get a sense of their functioning in VR, as well as their temporal and complementary relationship with requests of the system and changes taking place in VR. The experienced player is adjusting his embodied actions to displayed competencies and understandings and skillfully draws on and adapts interactive resources in both, the material and virtual ecology of action, thereby attending to (or even bridging) asymmetries inherent to the specific set-up of the VR game lab. In addition, the segment contains a concrete example of learning-in-interaction, as Tea eventually shows expertise in operating the motion controllers correctly (see Deppermann & Pekarek Doehler, 2021; Zemel & Koschmann, 2014).

3) Managing in-game mechanics – starting to play

In the final passage (Ex. 5) instructional activities revolve around essential game mechanics and shift from establishing how to control the adventure (i.e., game-specific functions of the controller) to determining what to do in the game (moving and looking around in the virtual environment and interacting with in-game objects). Extracts 5a–d demonstrate how the organization of instructional practices reflect the complexity of starting to play within a fractured ecology of interaction.

The VR-game in question is a demo version of "The Last Guardian" developed by Team Ico. There are two main characters, the "boy" who is directed by the player from a first-person perspective, and "Trico", a supersized but friendly fantasy beast that needs to be fed and taken care of, but also offers help. The design of the game requires the player to interact with teleports, or "hotspots", to move around in the virtual environment of what appear to be ruins of old, large buildings. Teleports are activated by focusing on them by aligning the head (direction of gaze), after which the player can transport to them by pressing "X" on the controller.

Before the episode, a new player, Elo, has taken the seat, followed by an instructional sequence that involves equipping and manipulating the headset, passing the controller, and selecting the game. Now the game has loaded, and Elo is *in* VR, which is acknowledged by Max with a short turn-initial *so*, projecting the beginning of a new activity that the now accessible game makes relevant (l. 1). Max first looks at the screen, but as he continues to talk, he approaches Elo from the left and turns his gaze towards him. However, when his gaze reaches Elo's hands holding the controller, he abandons his ongoing turn (l. 2) and initiates a new one: he extends his arm into the direction of the controller (fig. 30), and lightly touches Elo's hand while indicating a problem with what he sees (l. 4–5).

Extract 5a

```
1   Max     *eli.
            so.
            *looks at screen

2           *(0.5) täs pelissä-,
            in this game,
            *walks toward E, turns gaze to E’s hands

3   Elo     =°m°

4   Max     *tota:# (.) vaikka sullakin on *#nyt
            well (.) even though you also have now
            *..............................*lightly touches E’s right
                                            thumb on the right analog
                                            stick
    fig           #fig.30                    #fig.31
```

```
5          *(.) heti menee tähän [tää?]
           (.) immediately this goes here¿
           *lightly taps E's thumb repeatedly with R index finger--->

6  Elo                            ¤[mm¿ ]
                                  ¤nods
   max     --->
```


(Figure 30)


(Figure 31)

```
7  Elo     [joo.]
   max     --->

8  Max     [niin] tällä ei liikuta.
           with this one doesn't move.
           --->

9  Elo     joo.*
   max     --->*
```

The embodied design of Max's turn serves to correct a manual problem which is consequential for the game (Lerner & Raymond, 2021). Drawing on carefully adjusted interactive resources, including touch, Max elaborates that the left analog stick on which Elo's thumb currently rests is not used for moving around (or changing the perspective) in the game (l. 4–5 and 8). A first indication of a problem is noticeable in the restart in line 4, which Max slightly delays with the hesitation marker "tota", coupled with the extension of the hand towards Elo's hands. With "vaikka" (*even though*) and the clitic "-kin" on the pronoun ("sullakin"/ *you also have*), Max continues to project a negative assessment of the way that Elo is visibly preparing for play, but at the same time treats the problem as a common and understandable one for someone who has prior experience of gaming (l. 4–5). This reading of a routinized starting position is also emphasized by Max's *immediately* and the use of the intransitive verb "menee" (*goes*) when referring to the thumb in line 5 (*immediately this goes here¿*). Elo in turn displays understanding verbally (l. 6, 7, and 9) and by nodding (l. 6).

By pointing (fig. 31) and slightly tapping Elo's left thumb with his index finger (l. 4–9), Max attends to Elo's restricted visual access to the material ecology. The use of the mutually available resource of touch enables Elo to feel what Max is referring to with the indexical expressions "tähän" (*here*) and "tää" (*this*) in line 5, and "tällä" (*with this*) in line 8, which are contextualized by the continued tapping. This allows him to build an interactional space (Mondada, 2013) for performing the corrective instruction. Max thus assumes a shared focus, achieved by tactile deictic reference.

Having indicated how the device is not used, Max next orients to the right side of the controller in Elo's hands, giving instructions on how to look around in the game instead (*rather you look with this other stick*, l. 10–11). While he continues to gesture, now his actions do not involve touch. Rather, he first points at the right analog stick (l. 10, fig. 32) after which he produces a small pincer grip gesture that imitates manipulation of the stick (l. 11, fig. 33). However, these bodily actions remain invisible to Elo.

Extract 5b

```
10 Max      *#.h vaan
            rather
            *points at right analog stick
   fig       #fig.32

11          (.) *#sä tällä toisella tatilla* *katot,
            you look with this other stick,
                *pincer grip gesture  ,,,,,* *retracts arm,
                                             straightens,
                                             and gaze L at screen
   fig           #fig.33
```

(Figure 32)

(Figure 33)

```
12          >eiku nii *¤øsä katot          ¤päälläs tietysti.<=
            ah no you look with your head of course.
                      *slaps R hand in the air
   elo                 ¤focuses on hotspot ¤moves gaze around--->
   VR                   øhotspot reacts

13 Elo      =joo.
            --->
```

Although Max continues to mobilize multiple resources, Elo can only respond to Max's talk (and possibly audible proximity). Indeed, Max now verbalizes the referent, "tällä toisella tatilla" (*with this other stick*, l. 11), by which he ensures understandability. While the interactive meaning of his pointing gestures remains unclear, the fact that Max does not touch Elo anymore indicates that they have now established the target of the ongoing activity (determining how to operate the game).

In line 12, however, Max initiates self-repair with the particles "eiku nii" (*ah no*), which index a problem with his previous instruction and project a revision. Right before that (at the end of l. 11), he retracts his arm, straightens himself and gazes at the large screen to his left, thereby withdrawing from close proximity with Elo and projecting a new activity that possibly does not require immediate physical intervention (such as instructive touch). Max now corrects his instruction by stating that the player indeed uses the head(set) for looking around in the game. Concurrently with "päälläs" (*with your head*), Elo starts to move the head, changing the main character's/player's visual field in the VR, which is followed by a verbal receipt in line 13. In a finely coordinated way, he thus not only ratifies Max's repair, but he also treats it as an instruction that makes a complying action relevant. His embodied response further *mediates* Max's revision, displaying (an understanding of) what *you look with your head* means in VR. As in the previous examples, the temporality of the player's complying actions (starting to move the head at the same time as *head* is verbalized) reflects an orientation to the progressivity of the current task at hand and reveals careful, anticipatory monitoring of what Max is saying. This becomes even more apparent as the participants move on to the next task, *playing the game*.

The second part of the sequence, then, marks a shift to in-game activities, which – in retrospect – has been prepared by Max by changing his posture and position in the physical space (l. 11). Extract 5c shows how Elo silently demonstrates understanding of the game mechanics as he begins to interact with features of the virtual world and how this is finally coordinated both with Max's turn and with a visual instruction provided by the game. First, however, Max directs the attention to *hotspots* in the virtual environment, pointing at the screen with his outstretched left arm, palm down (l. 14).

Extract 5c

```
14 Max      ja-¤* sitten kun siellä on n:oita hotspot¤#teja¿
            and then since there are those hotspots¿
            -->¤                                  ¤......
                *points with whole R hand at screen, palm down--->
   fig                                             #fig.34

15 Elo      ø#[mm. ]            ¤
            ..focuses on hotspot¤
   max      --->
   VR       øhotspot reacts
   fig       #fig.35
```

(Figure 34)

(Figure 35)

```
16 Max         [niin] kun sä fokusoidut sillä katseel¤las¿
               so when you focus with your gaze¿
            --->
   elo                                           ¤interacts with the
                                                 hotspot (presses X)
```

```
17          ø(.) ja painat äksää¿ ^ niin sit se siirtyy siihen.
            and press X¿ then it moves there.
            --->
   VR       øloads new perspective^

18 Elo      ¤(0.5) °joo°.
            ¤turns gaze to next hotspot
   max      --->
```

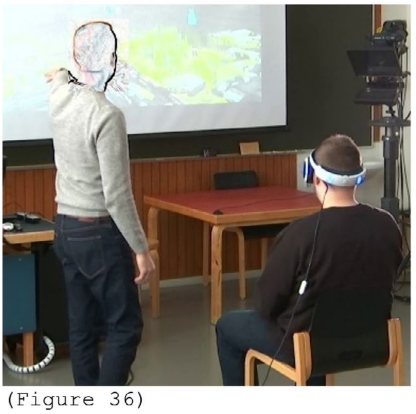
Simultaneously to Max's "hotspotteja", Elo begins to change the direction of his head/gaze until the focal point (or the "camera" angle, see Laurier & Reeves, 2014) aligns with the hotspot situated in the virtual visual field in front of the main character (l. 14–15, figs. 34 and 35). He thus continues to display close listening, locating the referent of Max's talk while also anticipating the next step: *focus with your gaze*. As the gaze engages with the hotspot in line 15, it reacts by changing the color and producing a small sound. At the same time, a sign occurs on the screen displaying an "X" in addition to the word "Move", which indicates that pressing the X-button of the controller will now teleport the character to the location of the hotspot, indeed making this very same action conditionally relevant for the player. However, Elo does not immediately follow the VR-directive. Rather, he slightly delays interaction with the hotspot until Max's explanation (that now positions focusing with the gaze as a prerequisite for some subsequent action) reaches a junction (at the end of l. 16). Elo then presses the X-button, which – after a short moment of loading during which Max continues his instructive talk by adding *and press X¿* (l. 17) – changes the location and perspective of the main character in the virtual environment. In line 17, Max initiates completion of the sequence, describing the consequence of *focusing* and *pressing X* in the virtual world. This is acknowledged by Elo with a silently produced "joo*"* and visible orientation to the next hotspot, preparing to move forward in the game.

The design of Max's turn (l. 14–17) forecasts several steps of action that build on each other and make relevant a specific order of interacting with the interface. Each step is marked with rising intonation, projecting continuation until the interactive work of learning how to see things and move around in the game is completed (l. 17–18). While both participants orient to this to-do-list, their working through the list appears somewhat out of sync as Elo's complying actions precede Max's instructions. However, in this instance, Elo also responds to instructions from the game. His instructed action reflects a double orientation of sorts, where he maneuvers the different temporalities of the game on the one hand and Max's talk on the other. In this way, Elo makes available his understanding of the game mechanics and shows agency in terms of being able to move forward in the game, which at the same time changes the essence of Max's instructions. Both participants display mutual orientation to the progressivity of the current gameplay-activity, projecting the nature of the unfolding instructions in relation to in-game requirements.

Next, Max suggests looking around, to which Elo, who was already targeting the next hotspot, complies without delay (l. 19), thus realigning with Max's focus. The design of the instruction suggests that Max orients to Elo as settled in and fully immersed. At the same time, he begins to move backward, further away from Elo, limiting the possibility of quickly returning to attending to potential problems with operating the hardware.

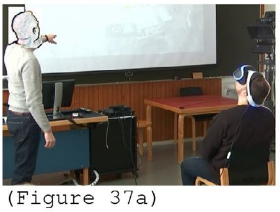
Extract 5d

```
19 Max      .h ja (.) nyt ¤(.) sä voit katella sit ympäril*les,
            and (.) now (.) you can then look around,
            --->                                         *....
   elo                    ¤looks to the left, away from hotspot

20          #ja sit jos katot (0.8) <ylöspäin¿>*
            and then if you look up¿
            --->       ...walks backward       *
   fig      #fig.36
```



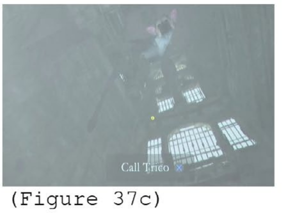
(Figure 36)

```
21 Elo      ¤(3.6)
            ¤looks up and around
   max      --->

22          ø#ee
            øTrico becomes visible
   max      --->
   fig       #figs.37 a,b,c
```



(Figure 37a)

(Figure 37b)



(Figure 37c)

```
23 Irja?    heh hh*h
   max        --->*
```

As Max disengages from the local material ecology of action, he continues to draw attention to the virtual environment by telling Elo to look up (l. 20). Elo responds to the instruction by looking up and around (l. 21), until Trico becomes visible (l. 22, figs. 37a, b, and c), sitting far above the main character on a ledge, looking down at him.

Extract 5, then, is another example of the participants' observable orientations to divergent visual fields and to different agency over physical and virtual communicative resources – here embedded in the moments of starting/entering an unfamiliar VR game. By carefully employing touch that contextualizes deictic expressions and clearly attends to Elo's restricted visibility, Max guides Elo through the initial steps of determining game-specific commands, while acknowledging the player's expertise as an experienced (yet novice VR-)gamer. The instructional activities are thus not only situated in the boundaries of the virtual and physical environments, but also reveal close monitoring of the player's starting position and incipient game play, thereby projecting (and attending to the) sequential progressivity of the task at hand. At the same time the passage involves moments of embodied disengagement from Elo's reality: For example, Max's prolonged pointing to the large screen (l. 14–23) is not visible to the player, and therefore can be seen to fulfil a different instructional function (e.g., in terms of engaging others that are in the room, while still talking to Elo). Similarly, Elo skillfully navigates Max's and the VR's instructions. Both participants thus display multiple orientations to the different ecologies of action.

Conclusion

The instruction sequences that were examined in this study represent interactionally complex moments in which the participants mutually attend to asymmetries, bridging physical and virtual ecologies. Instructions and instructed actions at the boundaries of the material environment and VR are accomplished through carefully adjusted and integrated embodied and technologized resources that reveal overlapping orientations to distinct visual fields and possibilities for action. Instructions are situated in co-located VR-gameplay in that they involve tactile engagement with the devices, guiding the player in how to receive the controller(s), how to *see* in VR, how to handle the equipment and how to interact with (objects in) the PlayStation VR.

Our findings show how the sequential environment, different spatialities and material organization are consequential for the organization of instruction sequences. Getting equipped for play involves explicit verbal instruction, bodily adjustments and movements that enable the participants to accomplish object transfer in circumstances of divergent access to visual resources. When establishing controller functionality and calibrating the game, manual guiding and corrective touch, and affordances of the VR, emerge as important resources for locating features of the controllers and getting a sense of their functioning to gain agency in the virtual space. Managing in-game mechanics relies on close coordination and skillful navigation of the different ecologies of action. This becomes particularly apparent in extract 4, where the participants establish shared attention to the virtual representation of the controllers, which then works as a reference point for locating the correct controller button in the physical space.

The analysis highlights the social nature of co-located gaming (e.g., Baldauf-Quilliatre & Colón de Carvajal, 2021; Tekin, 2021). Despite asymmetrical access and distributed agency that characterize the VR events in our data, they emerge as inherently social activities. The analysis shows that expert assistance may be beneficial, if not required, for novice players to start and successfully operate a VR game. Entering VR involves embodied (interactional) effort and adjustments to intersecting ecologies (i.e., a sense of operating in a physical space while being in VR) in addition to technical knowledge, for example. In view of this, implications of previous research on a wide range of applications of immersive VR according to which the technology may be easily and successfully implemented should be taken with caution (see also Fuchs, 2017, on gamification).

The pivotal role of embodied action and sociality in the context of VR-in-interaction has implications for UX-design and sociological research on HCI, calling for more attention to the complexities of human conduct situated in the domains of technology-use. Even though this has been pointed out already in the 2000s by Hindmarsh et al. (2006, p. 814), when stating that "[...] sociological accounts might be richer through serious consideration of the practical, embodied production of 'virtual' action", this remains an important task for design and research. Bearing in mind that our study is limited to the specifics of interaction in a lab, future work on instructed action at the intersection of physical and virtual ecologies should expand to natural settings that increasingly involve utilization of VR, such as training, entertainment, or rehabilitation.

Acknowledgements

We would like to express our gratitude to the two anonymous reviewers. Their valuable insights and suggestions helped improve our manuscript.

References

Baldauf-Quilliatre, H., & Colón de Carvajal, I. (2021). Spectating: How non-players participate in videogaming. *Journal für Medienlinguistik, 4*(2), 123–161. https://doi.org/10.21248/jfml.2021.33

Bennerstedt, U., & Ivarsson, J. (2010). Knowing the way. Managing epistemic topologies in virtual game worlds. *Computer Supported Cooperative Work, 19*, 201–230. https://doi.org/10.1007/s10606-010-9109-8

Billingsley, G., Smith, S., Smith, S., & Meritt, J. (2019). A systematic literature review of using immersive virtual reality technology in teacher education. *Journal of Interactive Learning Research, 30*(1), 65–90.

Carlin, A., Marques, J., & Moutinho, R. (2021). Seeing by proxy: Specifying professional vision. *Learning, Culture and Social Interaction, 30*, Part A. https://doi.org/10.1016/j.lcsi.2021.100532

Deppermann, A. (2018). Instruction practices in German driving lessons: Differential uses of declaratives and imperatives. *International Journal of Applied Linguistics, 28*(2), 265–282. https://doi.org/10.1111/ijal.12198

Deppermann, A., & Pekarek Doehler, S. (2021). Longitudinal conversation analysis – Introduction to the special issue. *Research on Language and Social Interaction, 54*(2), 127–141. https://doi.org/10.1080/08351813.2021.1899707

De Stefani, E. (2018). Formulating direction: Navigational instructions in driving lessons. *International Journal of Applied Linguistics, 28*(2), 283–303. https://doi.org/10.1111/ijal.12197

Diemer, J., Alpers G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology, 6*, 1–9. https://doi.org/10.3389/fpsyg.2015.00026

Ekström, A. (2012). *Instructional work in textile craft. Studies of interaction, embodiment and the making of objects*. Stockholm University.

Felton, W. M., & Jackson, R. E. (2022). Presence: A review. *International Journal of Human–Computer Interaction, 38*(1), 1–18. https://doi.org/10.1080/10447318.2021.1921368

Fuchs, M. (2017). Between drudgery and "Promesse du Bonheur." Games and gamification. In R. Kowert, & T. Quandt (Eds.), *New Perspectives on the Social Aspects of Digital Gaming: Multiplayer 2* (1st ed.) (pp. 185–199). Routledge.

Garfinkel, H. (2002). Instructions and instructed actions. In H. Garfinkel (Ed.), *Ethomethodology's program* (pp. 197–218). Rowman & Littlefield.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32*(10), 1489–1522. https://doi.org/10.1016/S0378-2166(99)00096-X

Heath, C., & Luff, P. (1993). Disembodied conduct: interactional asymmetries in video-mediated communication. In G. Button (Ed.), *Technology in Working Order: Studies of Work, Interaction, and Technology* (pp. 35–54). Routledge.

Heath, C., Luff, P., Sanchez-Svensson, M., & Nicholls, M. (2018). Exchanging implements: The micro-materialities of multidisciplinary work in the operating theatre. *Sociology of Health & Illness, 40*(2), 297–313. https://doi.org/10.1111/1467-9566.12594

Heritage, J. (1984). *Garfinkel and Ethnomethodology*. Cambridge: Polity Press.

Hindmarsh, J., Heath, C., & Fraser, M. (2006) (Im) materiality, virtual reality and interaction: Grounding the 'virtual' in studies of technology in action. *The Sociological Review, 54*(4), 795–817. https://doi.org/10.1111/j.1467-954X.2006.00672.x

Howard, M. C. (2017). A meta-analysis and systematic literature review of virtual reality rehabilitation programs. *Computers in Human Behavior, 70*, 317–327. https://doi.org/10.1016/j.chb.2017.01.013

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. Lerner (Ed.), *Conversation Analysis. Studies from the First Generation* (pp. 13–31). Benjamins.

Keevallik, L. (2010). Bodily quoting in dance correction. *Research on Language and Social Interaction, 43*(4), 401–426. https://doi.org/10.1080/08351813.2010.518065

Laurier, E., & Reeves, S. (2014). Cameras in video games: Comparing play in Counter-Strike and the Doctor Who Adventures. In M. Broth, E. Laurier, & L. Mondada (Eds.), *Studies of Video Practices* (pp. 181–207). Taylor & Francis.

Lerner, G., & Raymond, G. (2021). Body trouble: Some sources of difficulty in the progressive realization of manual action. *Research on Language and Social Interaction, 54*(3), 277–298. https://doi.org/10.1080/08351813.2021.1936994

Liang, M.-Y. (2021). Pragmatic socialization through gameplay directives: Multimodal conversation analysis of avatar-embodied interactions. *Journal of Pragmatics, 171*, 36–48. https://doi.org/10.1016/j.pragma.2020.09.028

Liberatore, M. J., & Wagner, W. P. (2021). Virtual, mixed, and augmented reality: A systematic review for immersive systems research. *Virtual Reality, 25*, 773–799. https://doi.org/10.1007/s10055-020-00492-0

Lilja, N., & Piirainen-Marsh, A. 2022. Recipient design by gestures. *Social Interaction. Video-Based Studies on Human Sociality, 5*(1). https://doi.org/10.7146/si.v5i2.130874

Lindwall, O., & Ekström, A. (2012). Instruction-in-interaction: The teaching and learning of a manual skill. *Human Studies, 35*(1), 27-49. https://doi.org/10.1007/s10746-012-9213-5

Lindwall, O., Lymer, G., & Greiffenhagen, C. (2015). The sequential analysis of instruction. In N. Markee (Ed.), *The Handbook of Classroom Discourse and Interaction* (pp. 142–157). Wiley.

Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., & Oyama, S. (2003). Fractured ecologies: Creating environments for collaboration. *Human-Computer Interaction, 18*(1–2), 51–84. https://doi.org/10.1207/S15327051HCI1812_3

Mondada, L. (2012). Coordinating action and talk in interaction in and out of video games. In R. Ayaß, & C. Gerhardt (Eds.), *The Appropriation of Media in Everyday Life* (pp. 231–270). Benjamins.

Mondada, L. (2013). Interactional space and the study of embodied talk-in-interaction. In P. Auer, M. Hilpert, A. Stukenbrock, & B. Szmrecsanyi (Eds.), *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives* (pp. 247–275). De Gruyter.

Mondada, L. (2014). Requesting immediate action in the surgical operating room: Time, embodied resources and praxeological embeddedness. In P. Drew, & E. Couper-Kuhlen (Eds.), *Requesting in social interaction* (pp. 269–302). Benjamins.

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics, 20*(3), 336–366. https://doi.org/10.1111/josl.1_12177

Mondada, L. (2018). Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on Language and Social Interaction, 51*(1), 85–106. https://doi.org/10.1080/08351813.2018.1413878

Mondada, L. (2019). Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics, 145*, 47–62. https://doi.org/10.1016/j.pragma.2019.01.016

Moore, R. J., Hankinson Gathman, E. C., Ducheneaut, N., & Nickell, E. (2007). Coordinating joint activity in avatar-mediated interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 21–30. https://doi.org/10.1145/1240624.1240628

Nishizaka, A. (2006). What to learn: The embodied structure of the environment. *Research on Language and Social Interaction, 39*(2), 119–154. https://doi.org/10.1207/s15327973rlsi3902_1

Nishizaka, A. (2007). Hand touching hand: Referential practice at a Japanese midwife house. *Human Studies, 30*(3), 199–217. https://doi.org/10.1007/s10746-007-9059-4

Nishizaka, A. (2011). Touch without vision: Referential practice in a non-technological environment. *Journal of Pragmatics, 43*(2), 504–520. https://doi.org/10.1016/j.pragma.2009.07.015

Olbertz-Siitonen, M., Piirainen-Marsh, A., & Siitonen, M. (2021). Constructing co-presence through shared VR gameplay. *Journal for Media Linguistics, 4*(2), 85–122. https://doi.org/10.21248/jfml.2021.31

Raymond, G., & Lerner, G. H. (2014). A body and its involvements: Adjusting actions for dual involvements. In P. Haddington, T. Keisanen, L. Mondada, & M. Nevile (Eds.), *Multiactivity in Social Interaction. Beyond Multitasking* (pp. 227-246). John Benjamins Publishing Company.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica, VIII*(4), 289–327. https://doi.org/10.1515/semi.1973.8.4.289

Schmidt, A., & Marx, K. (2021). Staging tele-presence by embodying avatars: evidence from Let's Play Videos. *Journal für Medienlinguistik, 4*(2), 52–84. https://doi.org/10.21248/jfml.2021.35

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.

Strong, J. (2020). Immersive virtual reality and persons with dementia: A literature review. *Journal of Gerontological Social Work, 63*(3), 209–226. https://doi.org/10.1080/01634372.2020.1733726

Tekin, B. (2021) Quasi-instructions: Orienting to the projectable trajectories of imminent bodily movements with instruction-like utterances. *Journal of Pragmatics, 186*, 341–357. https://doi.org/10.1016/j.pragma.2021.10.018

Tekin, B. S., & Reeves, S. (2017). Ways of spectating: Unravelling spectator participation in Kinect play. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '17)*, 1558–1570. https://doi.org/10.1145/3025453.3025813

Tuncer, S., & Haddington, P. (2019). Object transfers: An embodied resource to progress joint activities and build relative agency. *Language in Society, 49*, 61–87. https://doi.org/10.1017/S004740451900071X

Zemel, A., & Koschmann, T. (2014). 'Put your fingers right in here': Learnability and instructed experience. *Discourse Studies, 16*(2), 163–183.

OTSIKKO JA ASIASANAT SUOMEKSI:

Fyysisen ja virtuaalisen toiminnan ekologioiden yhdistäminen: ohjeiden antaminen ja seuraaminen VR-pelisessioissa

ASIASANAT: manuaalinen toiminnan ohjaaminen, murtunut ekologia, multimodaalinen keskustelunanalyysi, ohjeet, ohjeistettu toiminta, virtuaalitodellisuus

Appendix

Transcription conventions:

sign	meaning	sign	meaning
.	falling intonation contour	>joo<	increased speech rate
,	level intonation contour	<joo>	decreased speech rate
¿	slightly rising intonation contour	.joo	word produced with inhalation
?	rising intonation contour	.h	audible inhalation
↑	sharp rise in pitch	h	audible aspiration
↓	sharp fall in pitch	( )	uncertain hearing
minä	emphasis	£nih£	smiley voice
JOA	strong emphasis		
[	beginning of simultaneous talk	*	embodied actions by Max
]	end of simultaneous talk	+	embodied actions by Ari
(.)	micropause	^	embodied actions by Jan
(0.5)	silences in tens of a second	¤	embodied actions by Elo
(( ))	transcriber's comments, descriptions of nonverbal actions	•	embodied actions by Tea
:	preceding sound is stretched	ø	actions by the VR
se-	glottal stop or cut off	#	figures
°joo°	whispered talk		
=	latches between words or turns	*--->	the embodied action continues across subsequent lines…
		--->*	…until the same symbol is reached.
		--->>	the embodied action continues after the excerpt's end