Environmentally coupled gestures as a communicative resource in the word explanation activity: A multimodal analysis of interaction in social VR

Heidi Spets

Viittausohje:

Spets, H. (2023). Environmentally coupled gestures as a communicative resource in the word explanation activity: A multimodal analysis of interaction in social VR [Ympäristöön yhdistyvät eleet kommunikatiivisena resurssina sanaselitysaktiviteetissa: Multimodaalinen analyysi vuorovaikutuksesta sosiaalisessa virtuaalitodellisuudessa]. *Prologi – Viestinnän ja vuorovaikutuksen tieteellinen aikakauslehti*, *20*(1), tulossa. https://doi.org/10.33352/prlg.120936

To cite this article:

Spets, H. (2023). Environmentally coupled gestures as a communicative resource in the word explanation activity: A multimodal analysis of interaction in social VR. *Prologi – Journal of Communication and Social Interaction*, *20*(1), forthcoming. https://doi.org/10.33352/prlg.120936

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)
forthcoming
https//doi.org/10.33352/prlg.120936

Environmentally coupled gestures as a communicative resource in the word explanation activity: A multimodal analysis of interaction in social VR

Heidi Spets
MA, Doctoral researcher
University of Oulu
heidi.spets@oulu.fi

received 11.10.2022 / accepted 15.9.2023 / published 30.11.2023

Abstract

This article contributes to our understanding of how participants use different resources to accomplish word explanations in social virtual reality (VR). The article draws on conversation analysis to examine audio-visual data of interaction on the Rec Room VR platform. A view of the physical space the participants inhabit has also been captured. There are twelve participants, and they have minimal experience with social VR. English is used as a lingua franca. The focus is on participants' use of environmentally coupled gestures (EnCGs) during a word explanation activity. The activity has two or more participants playing a word-guessing game, in which one participant explains a word using drawings and gesture as well as speech. Findings show that EnCGs that feature elements in the environment are more readily interpretable than EnCGs that feature elements over the avatar body. The latter can result in situations in which achieving the goal of a word explanation activity (correct guess) can be difficult. In addition, the explainer's orientation to their physical body and the recipient's orientation to the virtual body during the joint word explanation activity can create situations in which the gestures become difficult to interpret for the recipient. To conclude, the observations in this article reveal the importance of the alignment of virtual and physical gestures for the intelligibility of gesture in VR.

KEYWORDS: avatars, conversation analysis, gesture, interaction, virtual bodies, virtual reality

Introduction

Social virtual reality (VR) refers to immersive technologies which can be used to socialise and communicate with others while engaging in joint activities and gaming (Maloney et al., 2021; Maloney & Freeman, 2020). These technologies can use motion capture to transfer one's physical movements to VR. It is already being used to connect people from around the globe, for example, for multiplayer games and work meetings. In the future, VR will probably be used even more for the purposes of distance education (see, e.g., Davidsen et al., 2022; Pirker et al., 2020; Pirker & Dengel, 2021), as well as remote work and collaboration (see, e.g., Li et al., 2021), both of which have become increasingly vital recently. Effective new ways to communicate in virtual teams are needed for both work and leisure, which is why the potential of immersive social VR needs to be studied (Li et al., 2021).

Social VR is connected to social gaming, one of the main activities it has been developed for (see, e.g., Gunkel et al., 2018). Many of the major platforms include gaming as an activity (e.g., Rec Room[1], ALTspaceVR, and VRChat). Gameplay is an everyday activity in which people of all ages and backgrounds can come together and play (Baldauf-Quilliatre & Colón de Carvajal, 2021). It is a structured activity that is organised around a game and its rules and therefore differs from everyday family interaction, for example. Achieving gameplay requires the participants to actively organise and coordinate their actions (Hofstetter, 2021). Social VR offers a new platform for examining gameplay as an interactional activity. As collaboration is at the core of gameplay, an analysis of the activity illustrates how participants achieve teamwork in social VR.

VR makes co-present real-time interaction possible, whatever the physical location of each participant. Although VR or any other form of mediated interaction is not intrinsically deficient, the affordances it provides differ from face-to-face interaction (Arminen et al., 2016). It is therefore important to examine how people engage in and manage different activities using the affordances of a specific VR platform. Furthermore, VR is a new kind of environment to be coupled with embodied action. Environmentally coupled gestures (EnCGs) are gestures which utilise the environment in meaning-making (Goodwin, 2007). Without the environment, the gesture might lose a crucial aspect of its meaning. This element could be a drawing or something in the landscape; indeed, the environment itself can be that *element*.

This article aims to examine how participants use different resources to accomplish word explanations in social VR. The focus is on participants' use of EnCGs. As social VR platforms are becoming more popular, it is important to examine their use, and what participants' actions within such spaces can reveal about human interaction. Video recordings of interaction on the Rec Room VR platform are used to examine how the participants progress the word explanation activity. The method is multimodal conversation analysis (CA), which allows a detailed analysis of interaction as it occurs moment by moment.

Social virtual reality

As a setting, social VR is a technological configuration in which users interact through virtual bodies. VR offers unique opportunities to examine participants' actions in social interaction as, for example, the participants operate with two bodies (one physical and one virtual) and

[1]Recroom.com

in two spaces (Kohonen-Aho & Haddington, 2023), sometimes even across realities (see, e.g., Olbertz-Siitonen et al., 2021; Paulsen et al., 2022). Additionally, the use of screen captures can show us participants' views of the virtual space (incl. possible orientations) (see, e.g., Paulsen et al., 2022). Through avatars, the users can interact with both the environment and each other using speech and gesture (see, e.g., Blackwell et al., 2019; Zhang et al., 2017). The avatars' movements in immersive VR are based on motion capture, meaning the participants' gestures and other embodied actions are a representation of the movement of their physical bodies (Maloney et al., 2021). This allows a greater level of interactivity and immersion than other more traditional media such as TVs or desktop computers (Fox et al., 2009).

Although this paper focuses on immersive VR, there is already a significant body of research on social interaction in virtual worlds that are accessed through desktop computers (such as Second Life), especially in the field of computer-mediated communication (CMC) (see, e.g., Antonijevic, 2008; Schultze & Brooks, 2019; Sivunen & Nordbäck, 2015). While not immersive, interaction in virtual worlds is also mediated through avatars. In both virtual worlds and immersive VR, the participants have access to the same image and sounds – either through a screen (virtual worlds) or a head-mounted display (immersive VR). The participants also share access to each other's actions and orientations in both environments. In virtual worlds, the participants observe and control their avatars through their screen with a mouse and keyboard; in immersive VR, the avatars' movements are based on the participants' physical bodies' movements (Mills et al., 2022). For example, research on virtual world interaction has examined how participants establish a transition to an encounter via embodied pre-beginnings (Kohonen-Aho & Vatanen, 2021), as well as the creation of interactional spaces and the negotiation of space (Berger et al., 2016; Locher et al., 2015).

The use of avatars as virtual proxies in interaction in VR creates a situation in which participants inhabit two bodies at once: one that is physical, and one that is virtual (Kohonen-Aho & Haddington, 2023). As their avatars share the virtual space, users can have common points of reference. They can orient to the same things and recognise where the other is looking, or what they are seeing. Of course, the features of the technology, such as the accuracy of motion capture, can affect how feasible that is, and sometimes the animation or graphics can be sufficiently crude so that mutual orientation becomes difficult.

In VR, interaction is mediated, meaning it is mediated through technology (Jones, 2012). The environment is not as fully available to all participants in VR as it would be in face-to-face interaction due to issues with the field of view or the sense of another's physical presence (Haddington et al., 2023; Hindmarsh et al., 2006; Kohonen-Aho & Haddington, in press; Luff et al., 2003; Spets, 2023), for example. When using head-mounted displays, the field of view is limited. The horizontal field of view in the head-mounted displays used in this study is around 90 degrees, whereas humans have a horizontal field of view of around 120 degrees (without considering limitations such as glasses). This, combined with the lack of a physical sense of presence, can cause issues in trying to point out an object or establish and maintain mutual orientation, for example (Hindmarsh et al., 2006).

Furthermore, showing objects or referring to them can be a complex action in a mediated

setting, and it may require more interactional work from the participants (Hindmarsh et al., 2000; 2006; Melander & Svahn, 2020). When pointing at something in the physical world, one can simply turn from the person pointing to the indicated object. In VR, it may require the pointer to be located before where they are pointing at can be seen (Hindmarsh et al., 2000; 2006).

Participants adapt and use the interactional resources available to them whenever – and wherever – they interact. This extends not only to *traditional* resources such as talk and gesture that are shaped and repurposed in situ (Mondada, 2016) but also to technological means. Such means can be utilised in ways beyond their pre-designed purpose (Olbertz-Siitonen & Piirainen-Marsh, 2021). For example, mouse cursor movements can be used as pointing *gestures* (Melander Bowden & Svahn, 2020; Olbertz-Siitonen & Piirainen-Marsh, 2021).

While research on VR interaction is a growing field, there is a need to examine gesture in VR from a conversation analytic perspective. The context in previous research has been largely desktop-accessed virtual worlds, not the immersive VR discussed in this article. Research is emerging only now because the equipment has not previously been sufficiently affordable for everyday use where CA materials are typically gathered. Social VR technologies have been developed to a point where it is also easier for researchers to build lab settings and experiments. As the use of social VR spreads (popular social VR includes platforms such as Rec Room, AltspaceVR, VRchat and Sansar), it becomes critical to understand VR interaction. Fine-tuned interaction mechanisms such as embodied action formation and ascription are less known in a VR context. Multimodal CA is well suited for this task, as it analyses interaction as it occurs moment by moment between participants, and how participants use different multimodal resources to perform social actions (Mortensen, 2012).

One of the foci in CA is the progressivity of interaction (see, e.g., Stivers & Robinson, 2006). When designing avatars, developers have made choices regarding the appearance and functionality of the avatars (as well as other features of the platform, see, e.g., Kolesnichenko et al., 2019; McVeigh-Schultz et al., 2018). For example, the avatars in Rec Room lack full human bodies with all limbs and joints as well as finely articulated hands. These choices may now affect the progressivity of interaction. If the resources required to perform an action are unavailable to a participant, their ability to progress an activity may be affected. Research like this can reveal how the activity of a word guessing game is performed in a VR setting, and what kind of resources the participants use.

Environmentally coupled gestures

Multimodality is intrinsic to social interaction (Mortensen, 2012). In social interaction, participants build their actions by mobilising various multimodal resources such as speech, gesture, gaze and body orientation (Mondada, 2016). These resources are combined and used in establishing, negotiating, and repairing intelligibility and meaningfulness in interaction (Mondada, 2014; 2019). Although such resources, particularly gesture, have been examined in detail in CA research, the virtual body and the use of its features as a resource in VR interaction have not.

When someone holds an object in their hand and uses it as an element of the gesture, they are making an environmentally coupled gesture

(EnCG). For example, one could trace a crack in the coffee cup they are holding as they talk about it. Such gestures feature elements that are not part of one's body. These environmental elements are crucial to such gestures, as they represent an important element of the gesture's meaning. Without the element, a gesture may become completely meaningless (Goodwin, 2007). EnCGs are complex combinations of physical elements (objects, the environment, other participants), as well as the sequential environment (previous turns and larger contexts), that can be coupled with gesture and speech. They are thus communicative events that are designed for the recipient to see as a result of systematic work by the participant making the gesture (Goodwin, 2007).

Gestures are co-expressive with speech: neither is redundant, and they express different aspects of a shared meaning (see, e.g., Kendon, 2004; McNeill, 1992; 2005). Speech and gesture are bound together, and disrupting one can affect the other (McNeill, 2015). Gestures feature in the co-construction of intersubjectivity through their part in turn-taking, for example (Mortensen, 2012). Gestures can be categorised in various ways. These categories of gesture are not strict, and a gesture can feature elements of multiple categories (McNeill, 2005). In this article, deictic and iconic gestures are of interest. Iconic gestures' form or manner of movement presents an image of an object or action (McNeill, 2005). For example, one might be talking about an object and use gesture to illustrate its shape or size. Deictic gestures point to something. It can be a tangible entity, or the gesture can be used metaphorically (McNeill, 2005). The gesture is not necessarily made with the hand, but another extensible body part can also be used. Although the extracts chosen for this article have mostly deictic gestures, the collection also features iconic gestures.

EnCGs can also leave a trace and become inscriptions (Goodwin, 2007). Inscriptions might "fall beyond the boundaries of gesture" (Goodwin, 2007, p. 207), but there is a similarity between the two actions of using gesture to highlight something and inscribing something. This "family resemblance", as Goodwin (2007, p. 207) calls it, shows that this act of drawing in the environment does not necessarily fall beyond the boundaries of gesture. The main difference is that although the two may share the same or similar movements, one leaves a trace, while the other does not. As an example of the similarity between gesture and inscription, inscriptions as actions can function similarly to pointing gestures. Just as a pointing gesture refers to something in the environment, inscriptions can refer to something in whatever it is marking.

The use of gestures, especially EnCGs, remains somewhat unknown in an immersive VR context. Hand gestures in Rec Room are rather crude, as the participants lack control of their fingers (apart from their thumbs), and the hand movements are based on controller movements, not their physical hands. Although the users' capability of using gesture in VR has been examined in detail (Li et al., 2019), it is less known how participants use gesture in interaction in immersive VR. It is very different from forming gestures consciously via a keyboard, where you select gestures and body movements for an avatar from a predefined library. 3D Charades, a word explanation activity in Rec Room, provides a new functional context for examining EnCGs, as well as iconic and deictic gestures.

Materials and method

The VR platform used in this study was Rec Room, a social VR platform. It provided the

users with a large virtual space in which they could freely interact with both the environment and other users. The users could participate in various activities, ranging from charades to basketball to simply being with other users from all over the world in the form of computer-generated avatars. Some of the activities in Rec Room included Paintball, Disc Golf, and 3D Charades. The participants used HTC Vive to access the VR platform. Depending on their use of headphones, some of the participants could hear each other, both through the VR space and in the physical space they shared.

The participants interact in Rec Room as avatars (see Figure 1). In Rec Room, the avatars' appearance is pseudo-humanoid. They have certain humanlike features like upright posture, some facial expressions (e.g., smiling, frowning), and hands with opposable thumbs. The avatars lack certain features of the human body such as arms connecting their hands to their bodies, as well as a lower body. The avatars are also rather crude, and they do not differ in body size. The only feature of the avatar bodies that is determined by the user's physical body is their height, which is determined by tracking the elevation of the head-mounted display. The multimodal resources available to the participants are limited in Rec Room. They cannot use facial expressions to interact, as the avatars' expressions are automated ("predefined" in Antonijevic, 2008) and have little to do with the participants' actions. They can use the movements of their avatar bodies and its head, as well as head and body orientation, and hand gestures to interact in VR.

Figure 1. Avatars

Note. (c) Rec Room Inc.

The materials[2] used in this study were collected from Rec Room by six university student groups, each recording around an hour of audio-visual material. The recording sessions were part of a course on interactional linguistics. The course was not compulsory for the students, and they received credits upon completion of the course. There were twelve participants, all of whom were novice VR users, and they used English as a lingua franca. The participants gave their voluntary and informed consent to participate in the study and agreed to the use of frame grabs and pictures in scientific publications. They are referred to with pseudonyms. The data have been anonymised as agreed with the participants. Extracts have been transcribed in accordance with the conventions in Mondada (2018; see the appendix for a condensed list

[2] The materials were collected in 2016 with "tethered" head-mounted displays, meaning they needed to be connected to PCs through a cable connection. The HTC Vives that were used also used separate infra-red beacons to track the users' movements. In comparison, more modern head-mounted displays are often wireless and have sensors within the head-mounted display and the controllers to track users' movements.

of transcription conventions). The transcribed embodied actions are those of the participants' avatars.

The participants received few if any instructions about how they should interact during the recording sessions. Before the session, they were given information leaflets that described the study in general terms so as not to influence their actions. The participants were told that the students were interested in interaction in VR, even if the students had a more specific feature of interaction or an activity in mind. The participants were recruited by students as part of their course assignment. They were not paid or otherwise compensated for their time. The students recruited mostly people they knew. Of the six pairs of participants, five engaged in the 3D Charades activity.

The participants interacted in Rec Room in pairs. They were not instructed to do anything specific, as long as they interacted with each other, although some pairs were guided to the 3D Charades activity at some point during the recording session. While the 3D Charades' rules (displayed on a clipboard in the virtual space) guided the participants to use only a tool called the 3D pen (plus the commonly understood Charades rules of no speaking), the participants mostly chose to speak, nonetheless. The rules were not put in place or reinforced by the researchers. The aim in guiding some participants to the 3D Charades activity was that it was – based on previous experiences – an activity that got the participants to interact with each other in a way that provided potential materials for a multimodal analysis of social interaction.

The data were collected at the LeaF infrastructure at the University of Oulu. It had two sets of equipment, which were used to capture data. This allowed the unique possibility of recording two participants interacting in VR and capturing both participants' views of the situation through a screen capture from their headsets. However, it is important to consider that screen captures can differ from the experience of being in VR (Paulsen et al., 2022), and it is difficult to ascertain what the participants are looking at. During the recording session, there was a 360 camera in the ceiling for a view of the physical space (see Figure 2 for the set-up). Combining the three streams provides a more complete view of the situation compared to recording a single participant interacting with others (see Figure 3 for the edited video). Recording the views from both the virtual and physical world made it possible to see the participants' actions unfold simultaneously in both the physical and the virtual, as the participants inhabited both their physical and virtual bodies (Kohonen-Aho & Haddington, 2023). Using such parallel videos of physical and virtual spaces enables observations such as realising that a gesture made in the physical world does not appear in VR. Having two participants act together also raises the odds of recording them interacting with each other because random encounters can be rare inside the game. The benefit of having two participants in the same physical space, as well as the same virtual space, is that one can record real-time co-present interaction.

In conversation analysis (CA) (Sacks, 1992; Sidnell, 2013), interaction is studied on a moment-by-moment basis as it unfolds over time. Social activities are examined as participants accomplish them as situated sequentially organised turns of action. In recent years, CA has adopted a more holistic approach to interaction and moved from talk-in-interaction to talk, em-

Figure 2. The recording set-up

Note. The HMDs and hand controllers, as well as the 360 camera, can be seen here.

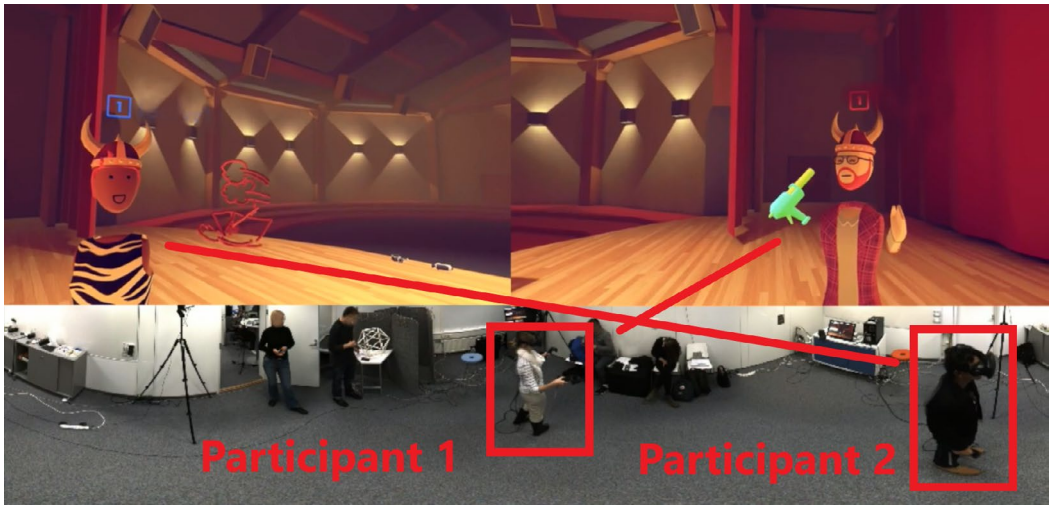

Figure 3. The edited view of all three recorded streams

Note. A screen recording from participant 1 is on the upper left, and their avatar is on the upper right. Meanwhile, a screen recording from participant 2 is on the upper right, and their avatar is on the upper left. The lower image is from a 360-degree camera recording the participants (highlighted) in the physical space ((c) Rec Room Inc.).

bodiment, and materiality in interaction. This is called multimodal CA (Goodwin, 2000; Mondada, 2016), and it will be the method used to analyse the collected data. Participants use different interactional resources such as gesture, talk, posture, and body movement to organise their actions (Mondada 2014; 2019; Mortensen, 2012). The data used in CA are recordings of naturally occurring social interaction that is interfered with minimally. Here, the data were quasi-experimental, as the participants were recorded in " [...] situations that simulate naturally occurring interactions and situations" (Due, 2015, p. 154). The participants used VR equipment at a research site and would not have interacted in VR without the study. However, the recorded interaction depended entirely on the participants, with little to no input from the researchers present in the physical space.

Results: Use of communicative resources during the word explanation activity

This article's focus is on the resources with which the participants achieve gameplay in social VR interaction. The Rec Room VR platform gives the participants an ability to create new objects by using a 3D pen, a glue pistol like *pen*0 they can use to draw 3D shapes in the air. These new objects are three-dimensional drawings that remain where they are drawn, often in the air. The participants use these drawings to structure their explanations in the 3D Charades activity. They are used as set-off points for the explanation and referred to with gesture and speech. The choice of activity was informed by its nature as a game that facilitated interaction between participants.

This study's focus activity was the game 3D Charades. It is a word guessing game in which one player draws a card with a word on it and then proceeds to pantomime or draw the word so that the other player(s) can guess it without the explainer using the word itself. The traditional rules of the Charades game explicitly state that the players should not use speech; that the words or phrases should be *acted out*. In the case of 3D Charades in Rec Room, acting out is replaced by drawing with the 3D pen. It is unclear how many of the participants were aware of these rules. Some participants acknowledged the rules, and one participant even said "I shouldn't even say anything while doing this, but it's hard to show" at one point. Even these participants did not strictly adhere to the rules. In any case, some if not most participants spoke in their attempts to describe a word.

In 3D Charades, the participants play in a variety of ways, from free play (no timer, started by simply moving to the stage and picking a card) to initiating a round (timer, started by pressing *play* in the game menu), and with or without the 3D pen. The participants take one of two roles in the activity: the *explainer* and the *recipient*. The explainer designs their turns to elicit guesses from the recipient, often building on previous turns during a longer explaining and guessing period. The recipient's guesses can inform the explainer of the recipient's understanding of the explanation, as well as the recipient's access to the explainer's embodied conduct.

The 3D Charades activity was chosen as the focus, as it provided a context in which the participants interacted with each other and looked at each other. As there was no other source of participants' embodied action in the virtual space apart from the views captured from their head-mounted displays, these properties became crucial for conducting multimodal CA. The word explanation activity that forms the basis of the game is also a fruitful context for

the analysis of gestures, as it promotes the use of drawings and embodiment rather than speech. As embodiment and use of material resources in interaction have yet to be examined in detail in social VR, examining the 3D Charades game provides new insights into social VR interaction.

The following analysis illustrates how participants use different resources, especially EnCGs, to progress the word explanation activity. Three extracts have been chosen from a collection of examples to illustrate the different resources used by participants. The collection consists of nine instances of the word explanation activity, most of which contain multiple EnCGs and/or inscriptions. The extracts illustrate situations in which the explainer is pointing at drawings in the environment and drawings on the avatar body.

Pointing at a drawing in the environment

Two participants, Sami and Jutta, have been playing 3D Charades for a few rounds. Sami is explaining the word *Matrix* to Jutta. He uses drawings, talk, and gesture in his explanation. Jutta guesses correctly in the end. Some of the pauses in Sami's explanation are due to him fixing the position of his HMD and headphones.

Extract 1a (Are those numbers 7:10)

```
1          *(2.0)^(2.0)^(3.8)
 =>sami:   *draws numbers, keeps drawing to l. 4-->
   jutta:        ^tilts her head to the left^
2  Jutta:  ↑are those ¤numbers.¤
                     ¤teleports closer to sami¤
3          ^(0.7)^
           ^turns to sami^
4  Sami:   ¤@y:e-@*¤ ^yeah.^
   jutta:  ¤teleports closer to sami¤
   sami        -->*
   jutta:             ^turns to sami^
5          (1.9)
6  Sami:   and we're- (.) we are both inside these?
7          %(0.7)%(1.7)
   sami:   %teleports%
8  Jutta:  inside?=
9  Sami:   =like,
10         (3.9)
11         would be if this was a horror film,
12         (1.0)
13 Jutta:  £oh£.
14         (1.8)
15         is it like a game↑ or:,
16         (0.6)
17 Sami:   uhm,
18         (1.9)
19         uhh hhh
20         ↑I ↑think ↑you ↑would ↑get ↑this a lot faster *if the#se* (0.8)
 =>sami:                                                  *points at numbers*
   fig                                                          #fig3
21 Sami:   uhh numbers were ^not red?
   jutta:                  ^tilts head to the right, holds position to l. 24-->
22 Sami:   but they we:re (0.4) .hh %green.%
   sami:                           %teleport%
23         (3.9)
```

Sami starts by drawing numbers (ones and zeros) on two parallel lines. Jutta asks, "are those numbers" (l. 2), which Sami confirms (l. 4). As Jutta is finishing her question, she teleports closer to Sami to see the drawing from the correct angle (l. 2), as from her previous position, the numbers were mirrored. After Sami has finished drawing the numbers, he continues his explanation by saying, "we are both inside these?" (l. 6). Jutta says, "inside" in an understanding check (l. 8). She also tilts her head (l. 1 and 21; see also fig. 5, l. 29, Extract 1b), using it as a resource to understand the drawing and show that she is orienting to thinking about her next guess. These head movements are also visible in the physical world.

One of Jutta's turns – "is it like a game or" (l. 15) – receives a delayed response from Sami. He uses the turn-initial particles "uhh" and "uhm" (l. 17 and 19) to delay, giving Jutta some time to reformulate her incorrect guess (Pillet-Shore, 2017). Sami then moves on to talk about how the colour of his drawing could account for why it is not recognisable as what Sami is referring to. It is a hint to the recipient regarding the word she is trying to guess: the movie *Matrix* uses green numbers as a prominent visual effect. It is at this point that Sami uses an EnCG. Sami waves his right arm horizontally, pointing at the numbers he has drawn (Figure 4). As he gestures, he says, "if these" (l. 20), referring to the numbers.

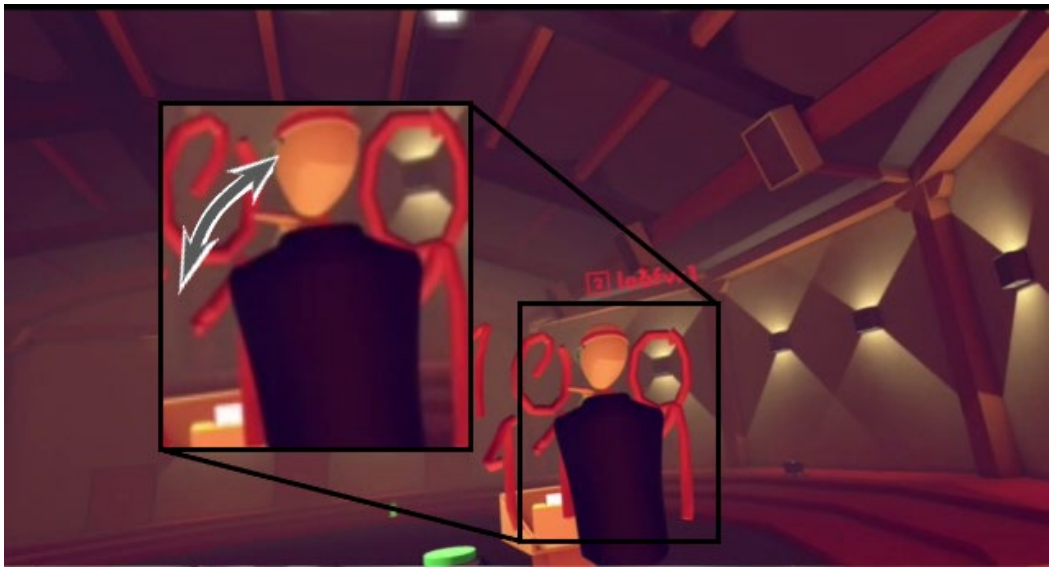

Figure 4. Sami gesturing at the numbers he has drawn (l. 20)

Extract 1b

```
24 Jutta:  green?^
             -->^
25         uhh %hh%
   sami:       %teleports to the table%
26         (2.9)
27 Jutta:  horror %game% with green numbers.
   sami:          %teleports%
28         (2.4)
29 Jutta:  ^that we're inside of.#
           ^tilts head back, turns it slowly right until she's
           looking forward again-->
   fig                             #fig4
30         (3.2)^
   jutta:    -->^
31 Jutta:  .hhh
32         hhhh it's ↑probably super easy w- when I know the *answer bu:t,
 =>sami:                                                    *starts drawing-->
33         (1.6)
34 Jutta:  uhmm
35         (5.8)
36 Sami:   actually this- (.) ∘(yeah) well∘.*
                                        -->*
37         (6.1)
38 Sami:   ∘(actually) I should do the-∘
39         %(2.6)%*(5.0)
   sami:   %teleports%
                *starts drawing-->
40 Sami:   (and I like)- wow I can (0.2) draw* in three dimensions.
                                          -->*
41         that's cool.
42         (0.3)
43 Jutta:  mmh?
44         (3.1)%(1.3)
   sami:        %teleports%
45 Sami:   I'll just-
46         (2.3)
47         *>nope,<
 =>        *writes NEO in the air-->
48         (1.6)*(6.4)*(0.9)
   sami:    -->*     *writes an M-->
49 Jutta:  .hh OH* is it matrix.
   sami:      -->*
50         (0.4)
51 Sami:   ye:s ri- that's right.
```

At several points, Jutta indicates that she is searching for the word and orienting to the activity at hand with body movement and speech (Goodwin & Goodwin, 1986; Heller, 2021). Jutta tilts her head to the side and turns her gaze to the middle distance (l. 29, Figure 5) to indicate that she is thinking about and searching for the word. Jutta also summarises her understanding of the explanation so far (l. 27 and 29).

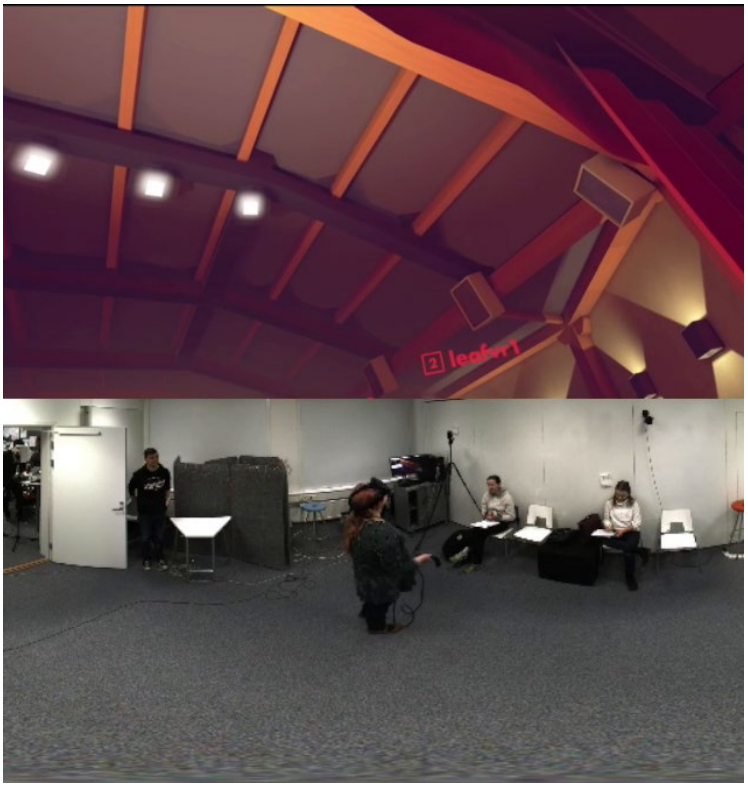

Figure 5. Jutta tilting her head (l. 29)

Jutta makes the correct guess on line 49 after Sami writes "NEO" in the air (l. 47), referring to the main character in the *Matrix* movies. Sami has just started writing an "M" in the air as Jutta makes the correct guess. Writing in the air using the 3D pen is one of the prominent resources the explainers use during word explanation in the whole collection.

Sami uses the following resources in his explanation: drawings (incl. writing) in the air and elaborative speech, as well as a gesture related to his drawing. The gesture is accompanied by speech, both referring to the drawn numbers. The explanation is a back-and-forth between the participants. Sami uses his turns to explain and elaborate on his previous turns in response to Jutta's requests for more information and her understanding checks. Jutta also repositions

her avatar by teleporting at the beginning. Her previous position was such that the drawings appeared as mirror images to her. By teleporting, Jutta gains access to Sami's drawings from the correct angle, allowing the two participants to share a point of view. In addition to teleporting, Jutta also uses head tilts as she thinks about her answers.

The EnCG Sami uses in his explanation is deictic, and the environmental element is a floating drawing. The EnCG refers to a drawing located in the environment, and it is intelligible to the recipient. However, the following two extracts illustrate two cases in which the explainer uses deictic EnCGs that refer to a drawing located over their virtual body. These EnCGs are less readily understandable to the recipient.

Pointing and orienting to different bodies

As has already been mentioned, the avatars in Rec Room are humanoids that are missing some body parts such as a lower body and arms connecting the hands to the torso. The participants also orient to different bodies – the explainer to their own physical body, and the recipient to the explainer's virtual body. This can make EnCGs that refer to specific body parts challenging to the recipient. The extract illustrates how the appearance of the avatar and participants' orientations to different bodies feature in the use of EnCGs in VR interaction. Here, Heikki is explaining the word *rib* to Pertti. The participants have been playing 3D Charades for some time at this point.

Extract 2 (Sort of like here 00:25)

```
1  HEIKKI: ¤(1.0)¤ %(sort of) like here.
           ¤glances at pertti¤lowers gaze towards his right side-->
                    %body tilts with the head-->
2          *(1.0)¤%(1.2)*¤
           *draws a filled-in circle over the right side of his body*
              -->¤lifts gaze to pertti¤
               -->%
3  PERTTI: *seventeen seconds.
 =>heikki: *starts repeatedly pointing at the drawing he has just made-->
4          (1.6)
5  HEIKKI: ¤can you see this, can you ¤%see this?*
           ¤lowers gaze a little.....¤lowers gaze to his side-->
                                      %body moves away from the drawing-->
                                             -->*post-stroke hold-->
6          (0.5)%¤
             -->%
              -->¤
7  PERTTI: ¤yeah I can see¤ a (0.4) hand- hand sort of looking thing*
   heikki: ¤lifts gaze to pertti¤
                                                                  -->*
8  PERTTI: *.hhh¤*
 =>heikki: *points at his side*
                ¤looks at the word card-->
9          (2.6)¤
             -->¤
10 HEIKKI: ¤.hh it's¤ sort of like *a part of-*
           ¤lifts gaze to pertti¤
                                  *single wave with both hands at waist level*
```

```
11 HEIKKI: (.) uhh, *the part of a human body*, *¤%and then.%¤*
 =>                *waves hands near shoulders**.....*draws a circle-->
                                                ¤glances down¤
                                                 %body moves%

12 PERTTI: (1.2) h:and?*
   heikki:           -->*
13 HEIKKI: ¤n:o.
           ¤shifts gaze to the left-->>
```

Heikki starts by saying, "(sort of) like here" (l. 1) and then draws a circle over his avatar's ribs, filling it out in the process (l. 2). Both his actions and his speech bring the attention to his virtual body. The beginning of the explanation can be seen in Figure 6. Pertti does not guess or respond to the explanation at this point, stating the remaining time instead (l. 3). Heikki then asks if Pertti can see the drawing, pointing at the drawing over his avatar's right at waist level (l. 5), to which Pertti responds that he sees a handlike shape (l. 7). The participants orient to different bodies: Heikki to his physical body; Pertti to Heikki's virtual body, as that is accessible to him. Heikki then points at his side again (l. 8). Heikki elaborates his explanation by saying "it's sort of like the part of a human body, and then" (l. 10–11), pointing at his previous drawing (l. 10–11), and then drawing over the previous drawing (l. 11–12). This time Pertti responds by guessing (l. 12), but he hesitates, and soon afterwards, the time runs out.

Figure 6. Heikki's first inscription (l. 2)

Note. Upper right, bottom centre: Heikki making the first inscription at the beginning.

The two inscriptions (l. 2 and 10–11) are almost identical. Heikki draws two filled-in circles over his avatar's right-hand side, approximately where his ribs are on his physical body. On the avatar, the drawings are further away from the body compared to where Heikki's hand is in relation to his physical body (Figure 6). They are not drawn *on* the virtual body, but *over* it in

the air. The location of the drawings also differs somewhat between Heikki's physical and virtual body, as they appear closer to the avatar's chest than its side.

The participants have asymmetric access to the communicative resources central to the explanation. Heikki is using a location on his physical body as the basis of his explanation. However, Pertti has access only to Heikki's virtual body. This, combined with a gesture pointing at a specific location on Heikki's physical body, makes the gesture unintelligible to Pertti. Additionally, while the avatars in Rec Room are humanoid, the lack of certain parts such as arms and a lower body can be misleading when trying to indicate a specific body part on the virtual body. When pointing at one's own ribs, one would have to extend one's elbow outward and bend the whole arm at an angle. This creates a noticeable visual cue, as the whole arm needs to be moved. However, this cue does not translate fully into the avatar's movements due to the features of the avatar's build. As the avatar has no arms connecting its hands to its body, there is no extended elbow. The difference between the original EnCG and the one visible in VR can be seen in Figure 7, the former on the left and the latter on the right.

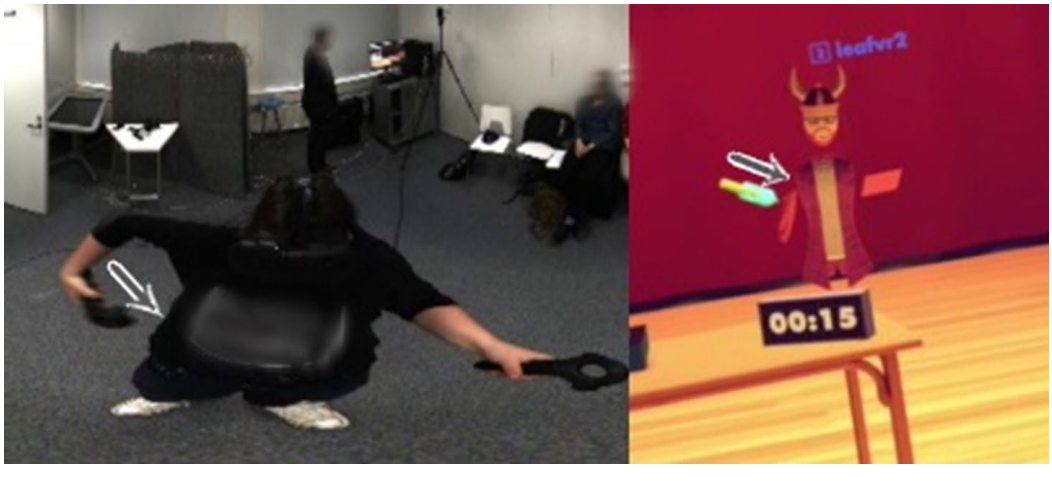

Figure 7. Difference between EnCGs
Note. Left: Heikki's physical body. Right: Heikki's avatar.

To summarise, when an inscription or an EnCG is coupled with a specific location on the human body, this coupling does not translate into VR as intended by the participant. This is especially the case when the avatar is not one-to-one with a human body or differs from the body of the person gesturing. The participants also have asymmetric access to each other's bodies and are unable to orient to each other's physical bodies, as they have no access to them.

Pointing and sensory mismatches

In the physical world, most of us can see and feel our bodies. We have a sense of our bodies' movement, as well as their location in space

and in relation to each other (Sheets-Johnstone, 2011). On the Rec Room virtual platform, this is not the case. When the participants gaze down on their virtual bodies, they see little to nothing there. They do not have the same sense of their virtual body and its dimensions as they do of their physical body (somatosensory mismatch, see e.g. Mills et al., 2022). If the participant's perception of their avatar does not meet the reality of their avatar, there are few opportunities to correct that perception. As gestures are organised with reference to the embodied configuration at hand (Goodwin, 2000), the interpretation of gestures in VR depends on the participants' sense of their virtual bodies.

In the previous extract, Heikki often gazes down on his virtual body and the inscriptions he has made during the explanation. From his perspective, the inscriptions seem to be exactly where they are supposed to be. However, the direction of his gaze – and the field of view available to him – do not show Heikki the entirety of his avatar. Additionally, it is his physical body's dimensions that he is using as the basis of his gestures (see Figure 7 for a comparison of the avatar and Heikki's physical body).

The next extract focuses on how the difference between Heikki's perception and the reality of his virtual body features in his use of EnCGs. Heikki is explaining the word *pocket*. As is the case with *rib* in the previous extract, Heikki's explanation is based on drawing shapes over the sides of his virtual body. However, Heikki orients to his physical body when explaining the word *pocket*, which affects his ability to design his gestures in a manner that is intelligible in VR.

Extract 3 (So you see 05:41)

```
1  HEIKKI: ^so you^ see,
   pertti: ^teleports closer to heikki^
2  HEIKKI: *(0.4) +here's+* (0.3) *here's me.*
           *waves hands at head level*
   pertti:        +focuses gaze on heikki+
   heikki:                        *waves his hands again*
3  HEIKKI: ¤(1.0)
           ¤turns gaze down-->
4  PERTTI: %yeah.*¤
   heikki: %left hand stays up until l. 17-->
                *right hand moves to left side-->
              -->¤gaze at his left side-->
5  HEIKKI: (0.6) *and# (0.7) here¤ are *¤you know like,¤
 =>           -->*draws circle over left side*
                              -->¤gaze at his right side¤
 =>                                    *draws circle over right side-->
   fig                #fig8
6  HEIKKI: ¤(1.0)*¤
           ¤lifts gaze up¤
              -->*
7  PERTTI: .hhh
8  HEIKKI: (0.3) state of# the ¤art¤,
                               ¤drops gaze towards the left drawing¤
   fig                     #fig9&10
9  HEIKKI: ¤(1.0)¤
           ¤lifts gaze to pertti¤
10 PERTTI: ( ) state of the art.
```

```
11 HEIKKI: *¤he he [he.]
 =>        *draws over the drawing on the left-->
            ¤gaze at his left side-->
12 PERTTI:          [oh,]*¤
   heikki:            -->*
                       -->¤
13 PERTTI: ribs?=
14 HEIKKI: =*¤you know,
 =>         *draws over the drawing on the right-->
             ¤keeps gaze on pertti-->
15 HEIKKI: (.)y-y-[you-(.)*¤you put¤  ](0.3)
16 PERTTI:        [I know (it's) ribs.]
   heikki:             -->*
                        -->¤looks down briefly¤
17 HEIKKI: ¤you know like %(0.3) things like (0.4)
           ¤looks towards pertti-->>
            lower handL-->%
```

Heikki has started by turning to Pertti and then bringing the attention to his virtual body by saying "here's me" and gesturing at himself (l. 1–2). The explanation itself starts with Heikki drawing two circles over both of his avatar's sides to represent pockets (l. 5–6, Figure 8). These inscriptions occur during "and here are you know like" (l. 5), and the first can be seen in Figure 8, with both visible in Figure 9 (on the right). Heikki fills in the drawings as he continues his explanation (l. 11 and 14).

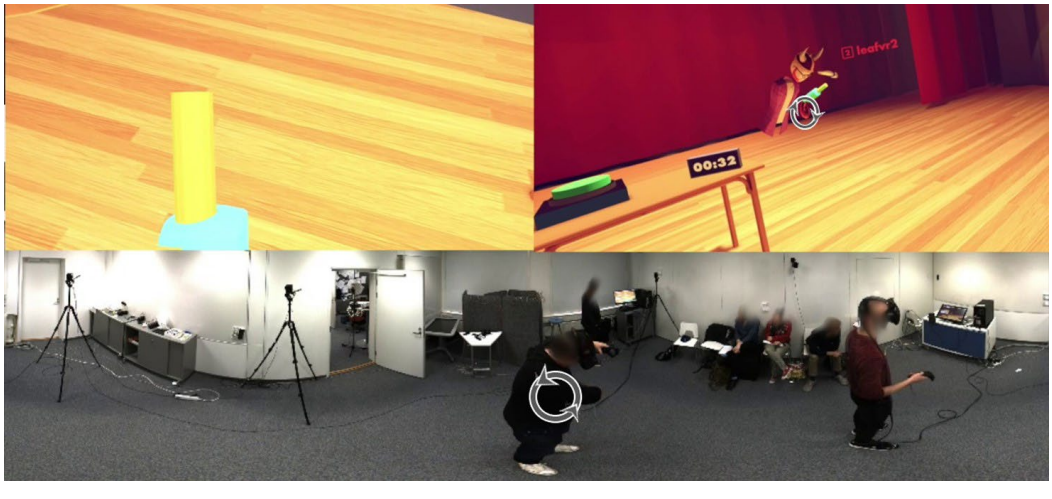

Figure 8. The beginning of Heikki's explanation

When examined through the physical world, the drawings are positioned over what would seem to be Heikki's hoodie's front pockets. In VR, the drawings seem to be roughly in the same position as the drawings Heikki made to represent ribs in Extract 2. Heikki's actions are also based on his physical body, making the intended form of the gesture unavailable in VR. However, the appearance of the avatar is only vaguely humanoid, and it lacks a lower body, making Heikki's gestures largely unintelligible because he is orienting and gesturing in relation to his physical body.

Heikki's approach to explaining the word *pocket* is similar to his approach to explaining *rib*, with him drawing circular inscriptions over his avatar's sides. Pertti's responds to the explanation by saying first "ribs" and then "I know it's ribs" (l. 13 and 16), displaying his orientation to the similarity of Heikki's inscriptions. Figure 9 shows just how similar the inscriptions are. In the context of Heikki's previous explanation, the inscriptions over that position are easy to mistake for a repeat of the inscriptions for ribs. The placement of the inscriptions does not correspond with anything pocketlike on the avatar. There are no visual cues that would orient the recipient to trousers as the avatar has no trousers, or even a lower body, to which to refer. Additionally, the virtual representation of the participants' embodied conduct lacks the finesse that is present in their conduct in relation to the physical body. The Rec Room environment is capable of replicating only so much of their movement, and the avatar's movements are often less precise than the participants intend. As the participants cannot see or feel their virtual bodies' movements, there is little to let them know of the differences between the movements they can sense themselves doing and the movements their avatar is doing.

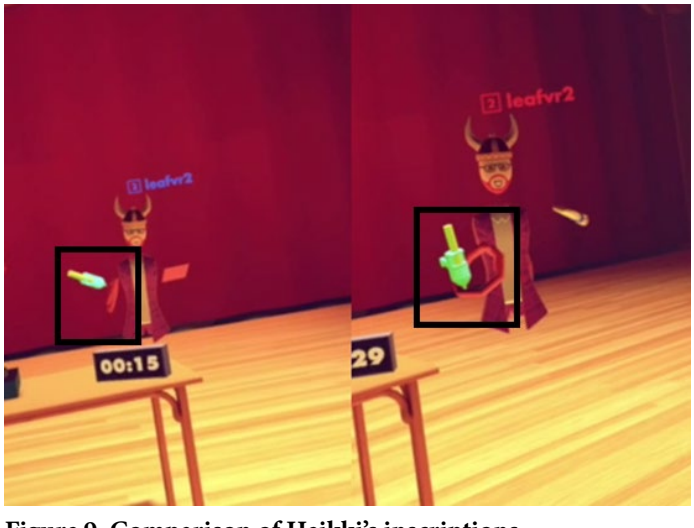

Figure 9. Comparison of Heikki's inscriptions

Note. Rib on the left, pocket on the right.

In addition, Heikki's movements cause the avatar to shift around. This leads to the inscriptions disappearing at times as they are stationary, and the avatar body covers them whenever they overlap (Figure 10). The participants can only sense their physical body, not their proxies in the virtual space, the avatars. In addition, the relative positions of the avatar and the inscription change each time the avatar moves. Whenever this happens, the intended meaning becomes more difficult to interpret, as the inscriptions are no longer available to the recipient. One cannot sense an object in VR in the same way as one can in the physical world (Mills et al., 2022). If there was something close to the side of your physical body, there would be ways for you to know it was there. You might see it from the corner of your eye or feel its closeness. You would have a general idea of the object's location in relation to your body. As these physical cues are missing in VR, it is more difficult for Heikki to be aware of the location of his inscriptions.

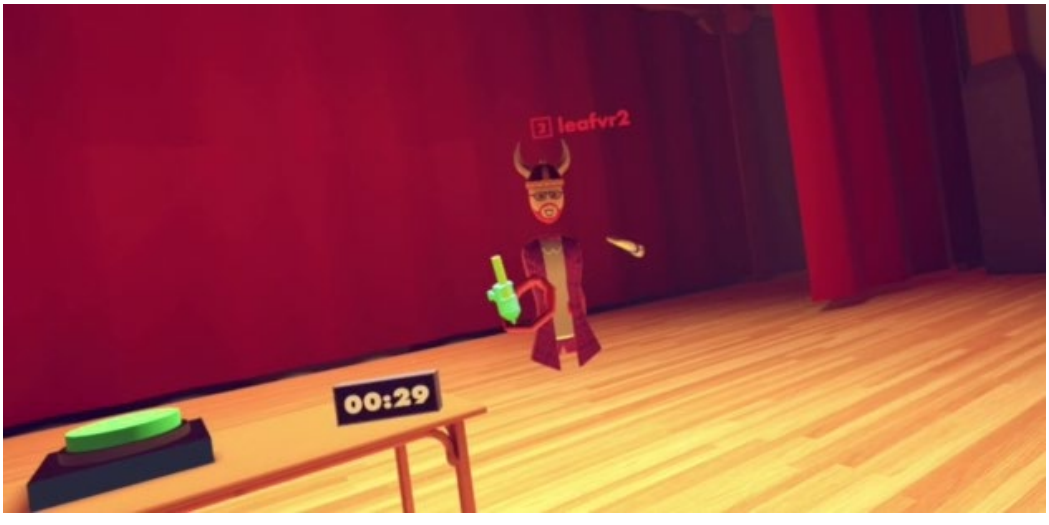

Figure 10. How the inscriptions made by Heikki in lines 5–6 appear in VR

Note. The inscription on Heikki's left has disappeared within the avatar.

Conclusion

This paper's main conclusion is that the alignment of virtual and physical gestures is important for the intelligibility of gesture in VR. Although some resources such as writing, as well as inscriptions and drawings in the environment, can promote the word-guessing activity, other resources such as inscriptions and drawings over the avatar body can hinder it.

As the analysis above revealed, the Rec Room virtual platform adds its own unique resources and challenges to playing a game of charades and achieving its goal. During 3D Charades, the participants are not only faced with the challenges of explaining a word without using the word but also the novelty of playing the game in a VR environment with its unique virtual affordances. In all the instances analysed in this article, 3D drawings are used as a resource for structuring the activity. The participants can

introduce these drawings as objects in space that serve as set-off points for the explanations. These objects are then described using talk and gesture, and they are ascribed certain features and enriched with additional information. The drawings are treated as reference points once in the world, and they provide a fruitful environment for EnCGs, as the participants use speech and gesture to refer to them. This means they are coupling their utterances with the environment by making the drawings the environmental element. At times, the participants make the drawings as part of an EnCG, making them into inscriptions that leave a trace in the environment.

Whenever the virtual and physical gestures are aligned, they seem understandable, such as in situations in which the actions of the physical body translate well into the actions of the virtual body. Writing would also appear to be a resource that functions well as long as the participants view it from the right direction. All in all, inscriptions and drawings, as well as pointing at them, seem to function as a resource that promotes the word-guessing activity in situations in which they are not drawn in relation to the explainer's body.

However, some resources seem to hinder the sensibility of the explanation and thus the recipient's ability to guess. Drawings made on and over the body, as well as pointing at one's body, can become unintelligible when combined with the mismatch between one's virtual and physical bodies. The participants can also orient to different bodies during the explanation, as the explainer can use their physical body as the basis of a drawing or gesture. The materials used in this study provided the opportunity to see the mismatch between the bodies and orientations. In the view captured from the physical world, one can see the participants' actions as they do them. The screen capture from VR then provides a view of what the recipient can perceive.

In new technological contexts, existing practices are not reproduced as is, and technology can reconfigure the use of resources (Due, 2015). This article yields insights into how avatar bodies function as a resource in VR interaction. Two of the examined instances illustrate how the participants' virtual bodies feature in their actions that are visible in VR through their avatar bodies. The participants' avatars display a few nonverbal cues regarding the recipient's understanding of the explanation, and it may be difficult for the explainer to design their own actions accordingly. For example, there are no facial expressions that can function as indications of trouble (Lilja & Piirainen-Marsh, 2019; Pajo & Laakso, 2020).

Some of the gestures examined in this article featured the participants' virtual bodies as an element of meaning. The analysis illustrated that the appearance of the avatar and the participants' inability to perceive their virtual bodies affected the intelligibility of EnCGs. The virtual body is limited as a resource for action in VR due to the mediated nature of VR interaction. Not all the movements a user makes are transferred to the movements of their virtual body, and the transfer is inaccurate, resulting in less precise movements. Technology and features of the specific virtual environment affect the users' ability to use their avatars as a resource for action. The avatars have been designed, as has the environment, in accordance with the developers' plans: in Rec Room, they are both highly stylised. The participants' ways of perceiving the environment, objects, and each other depend on the features of the VR equipment (e.g. the field of view, limited haptics, with only some vibrations in the controllers).

The lack of a physical sense of the body in VR combined with a field of view that is narrow compared to a human's usual field of view can result in an inability to perceive one's virtual body. The participants' actions show that their perceptions differ from the reality of their virtual bodies. From what is visible to us, their actions reveal how they orient to their physical bodies despite the use of a virtual body. Where the explainer orients to their physical body, the recipient has access to, and therefore can only orient to, their virtual body. At times, an important element of meaning – such as a feature of the human body – was invisible on the participant's virtual body, making the gesture unintelligible to the recipient. One explanation could be that the participants' inexperience with VR is a contributor in such situations. It has been shown that experience affects depth of interaction in VR, as an experienced user spends less time learning how the world works (Yilmaz et al., 2015). However, the participants in the present study are still in the process of building the reflexive awareness (Goodwin, 2000) needed to interact proficiently in a virtual environment.

The focus in this article was on EnCGs that incorporated deictic gestures. Possible future research could focus on examining aspects of iconicity in EnCGs, as such gestures were also part of the collection. It would be worth investigating how gestures depicting the shape, size, or movement of an object are used as a resource during word explanations. Further research that focuses on interactional challenges in VR may illustrate participants' skilful ways of overcoming them by adapting to the use of virtual interactional resources (Arminen et al., 2016). A longitudinal analysis of VR interaction could show how such adaptation occurs over time. As hybrid and mediated communication settings become increasingly widespread, users will have to adjust to fractured ecologies and new affordances. Examining interaction as it occurs in such spaces will be vital for our understanding of how these technologies are used in communication, and how people adapt to new interactional settings.

Acknowledgements

The article is based on my doctoral research, supervised by Prof. Tiina Keisanen and Dr. Laura Kohonen-Aho.

References

Antonijevic, S. (2008). From text to gesture online: A microethnographic analysis of nonverbal communication in the second life virtual environment. *Information, Communication & Society, 11*(2), 221–238. http://dx.doi.org/10.1080/13691180801937290

Arminen, I., Licoppe, C., & Spagnolli, A. (2016). Respecifying mediated interaction. *Research on Language and Social Interaction, 49*(4), 290–309. https://doi.org/10.1080/08351813.2016.1234614

Baldauf-Quilliatre, H., & Colón de Carvajal, I. (2021). Co-constructing presence between players and non-players in videogame interactions: Introduction to the Special Issue. *Journal for Media Linguistics, 4*(2), 1–13. https://doi.org/10.21248/jfml.2021.46

Berger, M., Jucker, A. H., & Locher, M. A. (2016). Interaction and space in the virtual world of Second Life. *Journal of Pragmatics, 101*, 83–100. https://doi.org/10.1016/j.pragma.2016.05.009

Blackwell, L., Ellison, N. B., Elliott-Deflo, N., & Schwartz, R. (2019). Harassment in social virtual reality: Challenges for platform governance. *Proceedings of the ACM on Human-Computer Interaction, 3*, 1–25. https://doi.org/10.1145/3359202

Davidsen, J., Larsen, D. V., Paulsen, L., & Rasmussen, S. (2022). 360VR PBL: A new format of digital cases in clinical medicine. *Journal of Problem Based Learning in Higher Education, 10*(1), 101–112. https://doi.org/10.54337/ojs.jpblhe.v10i1.7097

Due, B. (2015). The social construction of a Glasshole: Google Glass and multiactivity in social interaction. *PsychNology Journal, 13*(2–3), 149–178.

Fox, J., Arena, D., & Bailenson, J. N. (2009). Virtual reality. A survival guide for the social scientist. *Journal of Media Psychology: Theories, Methods, and Applications, 21*(3), 95–113. https://doi.org/10.1027/1864-1105.21.3.95

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

Goodwin, C. (2007). Environmentally coupled gestures. In S. D. Duncan, J. Cassell, & E. T. Levy (Eds.), *Gesture and the Dynamic Dimension of Language: Essays in Honor of David McNeill*. (pp. 195–212). John Benjamins. https://doi.org/10.1075/gs.1.18goo

Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica, 62*(1), 51–75. http://dx.doi.org/10.1515/semi.1986.62.1-2.51

Gunkel, S., Stokking, H., Prins, M., Niamut, O., Siahaan, E., & Cesar, P. (2018). Experiencing virtual reality together: Social VR use case study. *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video*, 233–238. https://doi.org/10.1145/3210825.3213566

Haddington, P., Kohonen-Aho, L., Tuncer, S., & Spets, H. (2023). Openings of interactions in immersive virtual reality: Identifying and recognising prospective participants. In P. Haddington, T. Eilittä, A. Kamunen, L. Kohonen-Aho, I. Rautiainen and A. Vatanen (Eds.), *Complexity of Interaction: Studies in Multimodal Conversation Analysis* (pp. 423–456). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-30727-0_12

Heller, V. (2021). Embodied Displays of "Doing Thinking." Epistemic and Interactive Functions of Thinking Displays in Children's Argumentative Activities. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.636671

Hindmarsh, J., Fraser, M., Heath, C., Benford, S., & Greenhalgh, C. (2000). Object-focused interaction in collaborative virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI), 7*(4), 477–509. https://doi.org/10.1145/365058.365088

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%2Fj.1467-954X.2006.00672.x

Hofstetter, E. (2021). Achieving preallocation: Turn transition practices in board games. *Discourse processes, 58*(2), 113–133. https://doi.org/10.1080/0163853X.2020.1816401

Jones, R. H. (2012). Analysis of mediated interaction. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Wiley. https://doi.org/10.1002/9781405198431.wbeal0024

Kendon, A. (2004). *Gesture. Visible action as utterance*. Cambridge University Press. https://doi.org/10.1017/CBO9780511807572

Kohonen-Aho, L. & Haddington, P. (2023). From distributed ecologies to distributed bodies in interaction: Capturing and analysing 'dual embodiment' in virtual environments. In P. Haddington, T. Eilittä, A. Kamunen, L. Kohonen-Aho, T. Oittinen, I. Rautiainen, & A. Vatanen (Eds.), *Ethnomethodological Conversation Analysis in Motion: Emerging Methods and Technologies* (pp. 111–131). Routledge. https://doi.org/10.4324/9781003424888-8

Kohonen-Aho, L., & Vatanen, A. (2021). (Re-) Opening an encounter in the virtual world of Second Life: On types of joint presence in avatar interaction. *Journal for Media Linguistics, 4*(2), 14–51. https://doi.org/10.21248/jfml.2021.30

Kolesnichenko, A., McVeigh-Schultz, J., & Isbister, K. (2019). Understanding emerging design practices for avatar systems in the commercial Social VR ecology. *Proceedings of the 2019 on Designing Interactive Systems Conference*, 241–252. https://doi.org/10.1145/3322276.3322352

Li, J., Vinayagamoorthy, V., Williamson, J., Shamma, D. A., & Cesar, P. (2021). Social VR: A new medium for remote communication and collaboration. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3411763.3441346

Li, Y., Huang, J., Tian, F., Wang, H., & Dai, G. (2019). Gesture interaction in virtual reality. *Virtual Reality & Intelligent Hardware, 1*(1), 84–112. https://doi.org/10.3724/SP.J.2096-5796.2018.0006

Lilja, N., & Piirainen-Marsh, A. (2019) How hand gestures contribute to action ascription. *Research on Language and Social Interaction, 52*(4), 343–364. https://doi.org/10.1080/08351813.2019.1657275

Locher, M. A., Jucker, A. H., & Berger, M. (2015). Negotiation of space in Second Life newbie interaction. *Discourse, Context & Media, 9*, 34–45. https://doi.org/10.1016/j.dcm.2015.06.002

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

Maloney, D., & Freeman, G. (2020). Falling asleep together: What makes activities in social virtual reality meaningful to users. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 510–521. https://doi.org/10.1145/3410404.3414266

Maloney, D., Freeman, G., & Robb, A. (2021). Stay connected in an immersive world: Why teenagers engage in social virtual reality. *Interaction Design and Children*, 69–79. https://doi.org/10.1145/3459990.3460703

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226514642.001.0001

McNeill, D. (2015). *Gesture in linguistics. International Encyclopedia of the Social & Behavioural Sciences* (2nd ed.), 10, 109–120. Elsevier. http://dx.doi.org/10.1016/B978-0-08-097086-8.53050-5

McVeigh-Schultz, J., Márquez Segura, E., Merrill, N., & Isbister, K. (2018). What's it mean to "be social" in VR?: Mapping the social VR design ecology. *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*, 289–294. https://doi.org/10.1145/3197391.3205451

Melander Bowden, H., & Svahn, J. (2020). Collaborative work on an online platform in video-mediated homework support. *Social Interaction. Video-Based Studies of Human Sociality, 3*(3). https://doi.org/10.7146/si.v3i3.122600

Mills, K. A., Scholes, L., & Brown, A. (2022). Virtual reality and embodiment in multimodal meaning making. *Written Communication, 39*(3), 335–369. https://doi.org/10.1177/07410883221083517

Mondada, L. (2014). The local constitution of multimodal resources for social interaction. *Journal of Pragmatics, 65*, 137–156. https://doi.org/10.1016/j.pragma.2014.04.004

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics, 20*, 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://www.lorenzamondada.net/multimodal-transcription

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

Mortensen, K. (2012). Conversation analysis and multimodality. In C. A. Chapelle (Ed.), *Conversation Analysis and Applied Linguistics: The Encyclopedia of Applied Linguistics*. Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0212

Olbertz-Siitonen, M., & Piirainen-Marsh, A. (2021). Coordinating action in technology-supported shared tasks: Virtual pointing as a situated practice for mobilizing a response. *Language & Communication 79*, 1–21. https://doi.org/10.1016/j.langcom.2021.03.005

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

Pajo, K., & Laakso, M. (2020). Other-initiation of repair by speakers with mild to severe hearing impairment. *Clinical Linguistics & Phonetics, 34*(10–11), 998–1017. https://doi.org/10.1080/02699206.2020.1724335

Paulsen, L., Davidsen, J. G., & Steier, R. (2022). "Do you see what we see?" – Perspective-taking across realities. In A. Weinberger, W. Chen, D. Hernández-Leo, & B. Chen (Eds.), *15th International Conference on Computer-Supported Collaborative Learning (CSCL)* (pp. 300–303). International Society of the Learning Sciences (ISLS). Computer-Supported Collaborative Learning Conference, CSCL. https://hdl.handle.net/11250/3065837

Pillet-Shore, D. (2017). *Preference organisation*. Oxford Research Encyclopedia of Communication. https://doi.org/10.1093/acrefore/9780190228613.013.132

Pirker, J., & Dengel, A. (2021). The potential of 360° virtual reality videos and real VR for education—A literature review. *IEEE Computer Graphics and Applications, 41*(4), 76–89. https://doi.org/10.1109/MCG.2021.3067999

Pirker, J., Lesjak, I., Kopf, J., Kainz, A., & Dini, A. (2020). Immersive learning in real VR. In M. Magnor & A. Sorkine-Hornung (Eds.), *Real VR – Immersive Digital Reality* (pp. 321–336). Springer. https://doi.org/10.1007/978-3-030-41816-8_14

Sacks, H. (1992). *Lectures on conversation*. Blackwell Publishers.

Schultze, U., & Brooks, J. A. M. (2019). An interactional view of social presence: Making the virtual other "real". *Information Systems Journal 29*(3), 707–737. https://doi.org/10.1111/isj.12230

Sheets-Johnstone, M. (2011). *Primacy of movement: Expanded second edition*. John Benjamins. https://doi.org/10.1075/aicr.82

Sidnell, J. (2013). Basic conversation analytic methods. In J. Sidnell, & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 77–99). Wiley-Blackwell. https://doi.org/10.1002/9781118325001.ch5

Sivunen, A., & Nordbäck, E. (2015). Social presence as a multi-dimensional group construct in 3D virtual environments. *Journal of Computer-Mediated Communication 20*(1), 19–36. https://doi.org/10.1111/jcc4.12090

Spets, H. (2023). Intersubjective interaction during the word explanation activity in social virtual reality. In P. Haddington, T. Eilittä, A. Kamunen, L. Kohonen-Aho, I. Rautiainen, & A. Vatanen (Eds.), *Complexity of Interaction: Studies in Multimodal Conversation Analysis* (pp. 145–174). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-30727-0_5

Stivers, T., & Robinson, J. (2006). A preference for progressivity in interaction. *Language in Society, 35*(3), 367–392. https://doi.org/10.1017/S0047404506060179

Yilmaz, R., Baydaz, O., Karakus, T., & Goktas, Y. (2015). An examination of interactions in a three-dimensional virtual world. *Computers & Education, 88*, 256–267. https://doi.org/10.1016/j.compedu.2015.06.002

Zhang, L., Sun, L., Wang, W., & Liu, J. (2017). Unlocking the door to mobile social VR: Architecture, experiments and challenges. *IEEE Network, 32*(1), 160–165. https://doi.org/10.1109/MNET.2017.1700014

OTSIKKO JA ASIASANAT SUOMEKSI:

Ympäristöön yhdistyvät eleet kommunikatiivisena resurssina sanaselitysaktiviteetissa: Multimodaalinen analyysi vuorovaikutuksesta sosiaalisessa virtuaalitodellisuudessa

ASIASANAT: avatarit, eleet, keskustelunanalyysi, virtuaaliset kehot, virtuaalitodellisuus, vuorovaikutus

Appendix: Transcription conventions (condensed from Mondada, 2019)

sign	meaning
(.)	A micropause, hearable but too short to measure.
(1.4)	Numbers in parentheses represent silence in tenths of a second.
wo-o-	Hyphens mark a cut-off of the preceding sound.
wo::rd	Colon indicates prolonged vowel or consonant.
hhh	Outbreath, proportionally marked
.hhh	Inbreath, proportionally marked
(words)	Unclear section.
((words))	Additional comments from the transcriber, e.g. about features of context or delivery.
Word,	'Continuation' marker, speaker has not finished; marked by fall-rise or weak rising intonation, as when delivering a list.
word?	Question marks signal stronger, 'questioning' intonation, irrespective of grammar.
Word.	Full stops mark falling, stopping intonation ('final contour'), irrespective of grammar, and not necessarily followed by a pause.
=	End of one TCU and beginning of next begin with no gap/pause in between (sometimes a slight overlap if there is speaker change).
[ ]	Square brackets mark the start and end of overlapping speech. They are aligned to mark the precise position of overlap as in the example below.
he he he	Pulses of laughter.
* *	Descriptions of embodied actions are delimited between
+ +	two identical symbols (one symbol per participant and per type of action)
Δ Δ	that are synchronized with correspondent stretches of talk or time indications.
*--->	The action described continues across subsequent lines
---->*	until the same symbol is reached.
--->>	The action described continues after the excerpt's end.
.....	Action's preparation.
fig	The exact moment at which a screen shot has been taken # is indicated with a sign (#) showing its position within the turn/a time measure.