# Interaction in immersive virtual reality:

## Breakdowns as trouble-sources in co-present VR interaction

Veikko Kaisto
Master's Thesis
English Philology
Faculty of Humanities
University of Oulu
Autumn 2020

**Table of Contents**

## 1. Introduction

Picture two people locked in a conversation. They have been at it for some time, and the topic happens to be on the more serious side of the spectrum – enough so to demand a degree of focus from both participants involved – only for the recognizable buzz of a smartphone's vibration alert to suddenly spring to life, cutting through the voices of the individuals involved. The conversation stops, mid-sentence, followed by a moment of hasty scrambling as the phone's owner retrieves the device from their pocket to either silence it and attempt at a return to the discussion, or break the interaction by taking the call. (An unexpected disruption has just occurred, forcing the participants to react. They are hardly an uncommon aspect of everyday interaction – everyone knows what it feels to be distracted by something unexpected or to lose their train of thought mid-sentence, after all – and we have by and large grown used to working through and around them.

Now picture the same initial scene, but this time the disruption comes in the form of motion suddenly freezing in one participant's point of view, or an object flying directly at another's head and passing through it with seemingly no resistance nor reaction, or perhaps an unexpected third individual suddenly appearing out of nowhere, invading one participant's personal space to such a degree that they are, effectively, phasing into their body. Such scenarios are quite impossible in our physical reality, of course, but within a *virtual* reality – even a very convincing one – the same rules which govern the real world do not always apply.

The very fact that virtual reality (VR) is being discussed as a technology which may see the beginnings of mainstream adoption in the near future is a testament to how far it has come in public eye over the course of the past decade. To an average western consumer of media, the concept likely brought to mind something akin to the titular computer simulation of *The Matrix* (1999) for the longest time, whereas those who had delved deeper into the science fiction genre may have drawn comparisons to the global information networks and cyberspace of works such as *Neuromancer* (1984) or the holodecks depicted in the various iterations of *Star Trek*. VR has not existed exclusively within the realm of genre literature, of course. However, while research into the technology reaches back decades, commercially available consumer-grade virtual reality devices have only entered the public discourse in the 2010s. In the years following the successful crowdfunding campaign of the Oculus Rift in 2012, head-mounted displays (HMDs) aimed at consumers – enthusiast or otherwise – have steadily entered the market throughout the latter half of the 2010s, bringing the subject of virtual reality into general discourse along with it. It was the announcement of the Oculus which ignited my personal interest in VR as a tangible, real-world technology available to the general populace instead of merely a recurring theme of science fiction or something exclusive to the aforementioned applications of research and industry.

Beyond purposes of entertainment, VR offers opportunities for fields such as long-distance collaboration, design and prototyping, scientific visualization, and psychiatric treatment (Bowman, Doug, Kruijff, LaViola, & Poupyrev, 2004; Geszten et al., 2015). Between single-user experiences to ones supporting dozens of participants, from hardware ranging from room-sized Cave Automatic Virtual Environments (CAVEs) to the contemporary head-mounted displays (HMDs) and motion controllers which have become increasingly available to the general public over the course of the last decade, the potential applications of the technology are unquestionably broad. Among the key draws of virtual reality are the heightened degrees of immersion and presence which the technology affords. Rather than interacting with a virtual environment via a two-dimensional interface – such as a traditional mouse, keyboard, and monitor of a personal computer, the touchscreen of a smartphone, or a contemporary video game console's controller – contemporary virtual reality devices aim to make the user's interfacing with the virtual environment as seamless and natural as possible; a person wearing a HMD can observe their environment by merely moving their head as they would in the ”real” world, for instance, as opposed to utilizing devices such as a mouse or a joystick to adjust their view. However, while new technologies offer new opportunities and potential applications, they also bring fresh challenges for designers and developers to overcome without the wealth of prior documentation and best practices to rely on which older, more mature technologies enjoy. While still a ways off from being a truly ubiquitous piece of mainstream technology, VR's ability to enable users to experience virtual spaces in a radically different manner to ”traditional” methods of interacting with them makes it both an intriguing field to follow, and a worthwhile topic to study. But it is not only a virtual world which users can interact with, but the other users within it.

Thus, we finally arrive to the general topic of this paper: What happens when unexpected issues arise during co-present interactions within virtual reality, such as when a real-world gesture fails to translate to VR due to an input device issue, or when the mechanical rules of the virtual environment are suddenly in conflict with the user's own, intuitive expectations? Does the interaction carry on with little disruption or come to a screeching halt as the issue is remediated or worked around, or do the users even attempt to resolve the issues in lieu of ignoring or otherwise circumventing them? And do certain kinds of disruptions appear to consistently elicit certain kinds of responses from affected users? These are some of the questions which I aim to investigate over the course of this thesis. To do so, I will analyse footage of in-game interactions within the social VR game *Rec Room* (Rec Room Inc., 2006) through an application of conversation analysis (CA) focused on identifying and examining such breakdowns.

The following section of the paper shall explore past research done on the topic of interaction in VR, while also establishing relevant key concepts and their respective definitions. Afterwards, the discussion will turn to the data being analysed in this thesis, the way it was collected, and how the collected data has been edited and otherwise handled in the process of using it for this paper. This will be followed by a section in which the research methodology and related theory will be presented, after which we shall proceed to the analysis itself, a discussion of the findings, and a summary of any conclusions drawn.

## 2. Background: Past research on interaction in virtual reality

In this section, we shall define and discuss key concepts related to the topic of this paper while providing a literature review of prior research relating to the study of interaction in virtual reality. Beyond VR itself, we will touch upon the role of avatars in virtual environments, the concepts of immersion and presence, and the various communications affordances which games and other forms of software have made to facilitate co-present interaction in virtual environments.

### *2.1. Virtual realities and virtual environments*

As a central concept of this thesis, it is only reasonable to begin by focusing on the topic of virtual reality and other related terms – namely, mixed reality (MR), augmented reality (AR), virtual environments (VE). At the time of writing, the colloquial use of VR has shifted towards using software applications – often ones focused on entertainment – by interfacing with them through head-mounted displays (HMDs) such as the HTC Vive, Oculus Rift, Valve Index, and various smartphone-based devices. This is by and large a development resulting from the growing availability of HMDs in the consumer market in recent years, and the attention this particular kind of VR hardware and software has gained in the media and amongst the general public as result. Neither the concept of virtual reality or that of a HMD are recent inventions, however, with the earliest examples of the latter having been developed all the way back in the late 1960s and the term itself being popularized circa the late 1980s (Bowman, Doug et al., 2004; Muhanna, 2015). Furthermore, while HMDs are perhaps the most archetypical type of modern immersive VR hardware, other varieties such as the Cave Automatic Virtual Environment (CAVE) also remain in use. In other words, a definition for VR which is predicated on a very specific type of hardware would likely be too restrictive to be considered accurate. At its broadest, on the other hand, VR can be used to refer to any computer-generated world (Pan & Hamilton, Antonia F. de C., 2018) which is general to the point of being of little use to us. Thus, the following question becomes pertinent: What kind of more specific definitions for VR have been offered in past research, and what recurring key characteristics can we identify from them which fit the purposes of this study?

Let us begin by exploring where VR stands in contrast to other similar technologies, most notably augmented reality. Both VR and AR are considered to exist on a spectrum of mixed realities, where a purely virtual reality exists at one of the extremes, with the purely physical reality at the other. In between the extremes are varying degrees of computer-generated information being injected into the real world such as AR, in which such information is added to the environment when it is viewed through a device such as a smartphone or Microsoft HoloLens (Bowman, Doug et al., 2004; Tham, 2018). Rather than AR, which adds information to the physical world we already inhabit, VR presents a completely synthetic, computer-generated world which may or may not be based on real-world locations. This characteristic is consistent across all the other definitions we are about to consider. Head-mounted displays, for instance, function by rendering the view of a virtual reality the dominant visual experience by completely covering the wearer's vision while blocking all view of the physical world (Thomas & Glowacki, 2018).

As a term, virtual reality is often connected to virtual environments (VE), but the distinction between VR and VE has, at times, been vague. Bowman (2004), for instance, uses the terms interchangeably while offering the following definition: "A synthetic, spatial (usually 3D) world seen from a first-person point of view. The view in VE is under the real-time control of the user (p. 7)." The concept of an *immersive* virtual reality also began gaining traction over the course of the 2000s. While we will further discuss immersion in section 2.3., let us examine following definition of immersive virtual environments (IVE) as offered by Bailenson et al. (2008):

> An *immersive virtual environment* is one that perceptually surrounds the user, increasing his or her sense of presence or actually being within it. Consider a child's video game; playing that game using a joystick and a television set is a VE. However, if the child were to have special equipment that allowed him or her to take on the actual point of view of the main character of the video game, that is, to control that character's movements with his or her own movements such that the child were actually inside the video game, then the child would be in an IVE. In other words, in an IVE, the sensory information of the VE is more psychologically prominent and engaging than the sensory information of the outside physical world (p. 104, source emphasis).

The authors here identify several typical characteristics of IVEs: the users' bodies are tracked as they interact with the software, in turn enabling head orientation and body position to be reflected within the VE, while also keeping sensory information from the physical world to a minimum. HMDs and CAVEs are given as examples of systems which support IVEs with their ability to track users' bodies and become the dominant perceptual experience in terms of both sight and sound. The VR set-up utilising HTC Vive HMDs from which data for this thesis was gathered also fits this definition. As with Bowman's (2004) interchangeable use of VE and VR, the concepts of immersive VE and immersive VR appear to be similarly familiar to one-another based on the above definition of Bailenson et al. (2008) with the key distinction being on the use of immersive devices such as HMDs. Another subset of VEs is the collaborative virtual environment (CVE), also defined by Bailenson et al. (2008) in the following manner:

> Collaborative virtual environments (CVEs) involve more than a single user. CVE users interact via avatars. For example, while in a CVE, as Person A communicates verbally and nonverbally in one location, the CVE technology can nearly instantaneously track his or her movements, gestures, expressions, and sounds. Person B, in another location, sees and hears Person A's avatar exhibiting these behaviors in his or her own version of the CVE when it is networked to Person A's CVE. Person B's CVE system then sends all of the tracking information relevant to his or her own communications over the network to Person A's system, which then renders all of those movements via Person B's avatar, which Person A can see and hear (p. 105).

This definition closely matches one offered by Chen et al. (2014) in their study of face-to-face interactions between users and their virtual avatars via a CAVE VR system. It is within these kinds of multi-user VEs where a sense of co-presence can be created. The game within which the interactions this thesis will focus on analysing were recorded, Rec Room, also fits this definition of a CVE.

### *2.2. The avatar and environment in VR interaction*

One recurring topic of the past studies done on the subject has been the role of a user's avatar. An avatar is a virtual representation of a user within a virtual environment, "a character who is fully controlled by a human being" (Pan & Hamilton, Antonia F. de C., 2018,

p. 402), which provides practical information for other participants (user's location in virtual space, what they are doing, the direction they're facing, etc.). As such, they are often considered a key feature of CVEs, and contribute to establishing a sense of co-presence between multiple users (Bailenson, 2006; Bosch-Sijtsema & Haapamäki, 2014; Chen, 2014). They can also be considered a part of the interface through which an user interacts with a given virtual environment and its other participants (Berger, 2016). Depending on the software and hardware used, avatars may be manipulated via key-based inputs, motion capture of the user's body or motion-sensitive controllers, vocal input, or a mixture of the aforementioned options (Chen, 2014). It is worth noting that beyond human-controlled avatars, virtual environments may also be populated with computer-controlled characters, referred to as virtual agents or non-player characters (NPCs) in the context of video games. The focus of this thesis, however, is interaction between human users as opposed to human-NPC interaction.

VR users have also been shown to experience a degree of body ownership illusion (BOI) with their virtual avatars. A virtual body left in an uncomfortable posture, for instance, has been shown to both increase a user's own level of discomfort and make it harder to maintain their concentration (Bergström, 2016), with similar results also having been found when placing the user and their avatar in a threatening situation within VR (Schulze, Pence, Irvine, & Guinn, 2019). The avatar's body being in a different posture to that of the user was not found to preclude one from experiencing BOI, though uncomfortable postures did appear to lessen the degree of it more than comfortable ones. This, in turn, may relate to a person's body-awareness increasing "in times of illness, tiredness, or when there is difficulty in accomplishing a practical activity" (Murray & Gordon, 2001, p. 366). Personal space can also become a factor as players grow immersed with a virtual environment, with users unexpectedly finding themselves in close proximity to objects and characters being a potential cause of discomfort (Fox, J., Arena, & Bailenson, 2009; Iachini, Coello, Frassinetti, & Ruggiero, 2014). This may be further complicated if players' reflexive responses to such an event, for instance attempting to move their own avatar or pushing the offending object or character away from them, does not bring about the expected result due to limitations of the virtual world.

Avatars in contemporary VR software are not perfect recreations of user's bodies, thus users inhabiting virtual avatars "will always be operating from the boundaries of a body that is different to his/her own" (Thomas & Glowacki, 2018, p. 152). This is not necessarily a major detriment to a user's experience in regards to immersion, as avatars of different bodily dimensions and genders to that of the one's own have a negligible effect on the levels of immersion and body ownership experienced by a given user within VR (Slater, Spanlang, Sanchez-Vives, & Blanke, 2010). User avatars also pose their own challenges for interaction, notably regarding nonverbal aspects of multimodal communication. In a study researching face-to-face interactions between avatars in a virtual environment by comparing verbal-only, gesture-only, and verbal-and-gesture forms of communication, Chen (2014) found that verbal-only communication made for the most efficient interactions: "Due to technology limitations or high cost, humanoid avatars do not commonly support facial expression and other subtle social cues" (p. 414). The overall fidelity of VEs can also be an undermining influence for the overall user experience. Not only do high-fidelity VEs demand both stronger hardware on the user's behalf and more development time on the developer's part, they may also increase the probability of users experiencing simulation sickness and negatively impact an user's perceptual experience (Kapralos, Collins, & Uribe-Quevedo, 2017). Perceiving and becoming aware of other users can also be challenging, as is determining whether or not someone else is available for interaction with limited cues available (Berger, 2016).

Users and their respective avatars do not, however, exist within a vacuum in most VR applications. As in within the real world, the environment itself also plays a factor in communication. Just as real spaces are designed and built for specific purposes – offices, classrooms, auditoriums, et cetera – so too are virtual spaces, which often try and replicate real-world locations: "The concepts of language and space are intimately connected. Communication happens in spatial contexts, and the spatial context has repercussions on the communication" (Berger, 2016, p. 85). The features of a virtual space determine how many avatars can exist within it, how close to each other they must be, and what kind of items may be available for the users to interact with. Consider, for instance, the room featuring the 3D Charades game-mode in the social VR game Rec Room. It is presented as a space akin to a theatre stage, with the person drawing the shapes which others must try and guess up on the stage (where a paint tool which enables a user to draw shapes in the virtual environment is also located), with the other players being placed in the audience portion of the room by default. In this case, the layout of the space itself aims to guide players into their roles during a given round of Charades. Players can, however, move about freely within the space once a round of 3D Charades starts, which can lead to unexpected actions such as switching out roles on the fly during an ongoing round.

While Berger (2016) made his findings in the context of *Second Life*, a title predominantly played with a mouse and keyboard on a standard computer monitor in contrast to a HMD, many of their observations also apply to environments experienced in immersive VR. There are, however, noteworthy differences: In *Second Life* and most other MMOs, the player controls their character from a third-person perspective, and the avatar's body and head predominantly face in the same direction (most often matching the facing of the third-person camera, i.e. the player's view). In contrast not only is the preferred perspective in immersive VR a first-person one, but the player's avatar's head and arms can usually move independently from the rest of the body. These factors can make it easier for other users to determine which way another user is looking and enables rudimentary gestures with motion-tracking controllers. However, the first-person perspective also offers a lower degree of environmental awareness than a third-person one such as the one seen in Second Life, where players can freely pan the camera perspective around and adjust its distance from their avatar typically by using a computer mouse. Such a camera enables players not only to easily see the entirety of their avatar in relation to their surroundings, but also to view more of the environment at once by zooming the camera out or adjusting the viewing angle. Contrast this with immersive VR, where the user's point of view is often rooted on their avatar's head.

### *2.5. Presence and immersion in VR*

The senses of presence and immersion are both key recurring topics in research relating to virtual environments. Neither concept has a clear-cut commonly accepted definition and are sometimes used interchangeably. While presence and immersion are not primary focuses of this study, discussing them is worthwhile as we analyse co-present interaction within a VR setting. Furthermore, the focus is on specific instances of interaction where intuitive and reflexive responses are likely to occur. It stands to reason that the degree of presence felt by the users in the dataset of this study has played a part in informing their reactions within VR, and while we do not have the means to assess the degree of presence felt by the users, it is important to be cognizant of the part it may play. In this section, we will settle upon a definition for both along with discussing their relevance to VR interaction.

Presence is considered a defining aspect of VR experiences (Bosch-Sijtsema & Haapamäki, 2014; Tham, 2018), with some going so far as to name it a sensation which VR is uniquely capable of producing (Abrash, 2014). Virtual environments have also been shown to be more effective at accomplishing their purpose (entertainment, education, etc.) when instilling a greater sense of presence in their users (Bowman, Doug et al., 2004; Lin, Duh, Parker, Abi-Rached, & Furness, 2002). A common thread of definitions given to the term is a sensation of being in a space; Tham (2018), for instance, characterizes presence as "a psychological space of existing within an environment" (p. 180). Other authors offer more specific definitions: Slater (2018) defines presence as an illusion of being in a virtual world even while knowing that one is not, emphasizing both the artificial nature of the environment and believing oneself to be there despite being aware of the fact, which is also echoed a description of presence laid out by Abrash (2014, p. 16): "[Presence is] more than just looking at someplace interesting; it's flipping the switch that makes you believe, deep in your lizard brain, that you are someplace interesting." This belief that one is both present within the artificial environment and that events occurring there are actually taking place increases the likelihood of users exhibiting response-as-if-real (RAIR) reactions to stimuli (Slater, 2009). With similar definitions also echoed by Bowman (2007), Bates-Brkljac (2012), and Marini et al. (2012), we can arrive to the following definition for the purposes of this paper: Presence is a user's psychological response to experiencing virtual reality, where they believe themselves to be inside the virtual environment despite being aware of its artificial nature.

We shall also distinguish the oft-conflated concepts of presence and immersion from one another for the purposes of this paper. To make this distinction, we can refer to one offered by Abrash (2014): "[Immersion] means that you feel surrounded by the image of the virtual world; presence means you're in the virtual world" (p.16). Bowman (2007), meanwhile, defines immersion as "the objective level of sensory fidelity a VR system provides" (p. 38). For our purposes, we shall lean towards a definition of immersion which refers to the degree of sensory feedback provided by a VR system to the user.

Multiple factors can contribute to enhancing a sense of presence within VR. The ability to move and affect the virtual environment are important contributors to creating a sense of presence (Bates-Brkljac, 2012; Thomas & Glowacki, 2018), as is a sense of spatial awareness affected by aspects such as the field of view (FOV) offered by the hardware being used (Lin et al., 2002). Aiming for photorealism in a VE's presentation is not necessarily the most efficient method of achieving a sense of presence and verisimilitude; rather, the VE should "be faithful to its spatial representation" (Barricelli, 2016, p. 880) and focus on ensuring that events which occur within VR come across as plausible to users witnessing them (Bates-Brkljac, 2012). In their study among students using various VR devices, Tham (2018) found that rich sensory stimulation, positive emotions, and a sense of having agency served to enhance users' sense of presence, and that it was not hindered by users controlling avatars which did not match their own physical traits. Conversely, the sense of presence and immersion can also be undermined. VR applications can be more physically taxing and disorientating for users than traditional forms of software (Geszten et al., 2015), and technical issues and the stimuli, such as glitches in the HMD, latency, performance issues, background noise not related to the VE, can bring an user out of their immersed experience (Thomas & Glowacki, 2018). Issues relating to the VR devices, thus, stand a risk of breaking a player's sense of presence.

### *2.4. Communication affordances in VR*

What kinds of methods of communication do users have to interact with each other within virtual environments? The options available to users may differ drastically between different applications based on input methods and the core mechanics of a given piece of software itself. In the case of Second Life, Berger, Jucker & Locher (2016) identify several different *affordances* the game makes in order to offer players options to communicate with each other. These are further broken down into language- and avatar-based affordances. An in-game text-based chatbox, instant messaging (IM), voice over IP (VOIP, aka. voice communication/voice chat), notecards, action scripts, and signs in the game environment are identified as language-based affordances, while avatar-based affordances include the appearance of the avatar itself, their movements as the player behind a given avatar moves them within the virtual world, and various gestures in the form of emotes – laughing, nodding, clapping, et cetera. Of these, the authors name open chat (global channel in in-game chat), VOIP, and IM as the primary communication methods. Emotes – physical gestures performed by a player's avatar – can be triggered via typing specific commands or activating them through a dedicated menu, at which point the player avatar plays a specific animation and possibly produces certain sounds.

In regards to nonverbal communication in particular, Antonijevic (2008) broke nonverbal cues down into four distinct categories: predefined (actions generated by the software's own systems, such as idle animation loops), user-defined (actions performed by the user, such as moving their avatar or adjusting its facing), blended (actions selected by the user which make use of predefined assets, such as emotes), and missing cues (actions which cannot be executed within the virtual environment). While these categories are based on the communication options provided by *Second Life*, they are also applicable to VR titles. In the context of *Rec Room*, for instance, a predefined action would be the changes which occur on the drawn face of an user's avatar as voice communication is used, while fine finger motions or accurate depictions of gaze would constitute as missing cues. A lack of subtle gestures and expressions on user avatars is one of the challenges VR interaction faces (Marini et al., 2012).

While these affordances are all mainstays of contemporary online video games - particularly massively multiplayer titles such as *Second Life* and *World of Warcraft* - the affordances offered to a user within VR often differ. VOIP is the dominant form of inter-user communication, as while in-game text chats do exist in certain games, the use of a microphone is much easier considering the realities of HMDs and their respective controllers. An even greater difference applies to the topic of gestures or emotes: with player avatars' hand- and head-motions following those of the player themselves when using a VR set-up such as those used in the dataset of this paper, 'pre-packaged' emotes are much less common in VR applications. Instead, the act of gesturing and emoting is left for the player themselves to carry out within the constraints of both the VR application being used, and the hardware they are utilizing. In some cases users develop rituals for ‘bowing in’ when multiple players simultaneously enter or otherwise encounter one another within the same VE, in the form of participants taking turns to acknowledge the presence of others through specific, simple gestures (Thomas & Glowacki, 2018).

In the next section, we shall discuss the data gathered for this thesis, the data collection process, and how the data has been handled.

## 3. Data overview and the data collection process

The data which this thesis's analysis is based upon consists of audio-visual recordings of pairs of individuals interacting within VR. The materials used were recorded in October 2016, as a part of a course on interactional linguistics with a focus on multimodal interaction within Virtual Reality. They were gathered over the course of recording sessions carried out by six groups of students partaking in the course, each recording approximately one hour of audio-visual footage during their respective sessions. I was personally a member of one of these groups carrying out the recordings. The sessions themselves were carried out at the LeaF facilities at the University of Oulu. During these sessions, two individuals recruited by each group (henceforth referred to as *participants*) made use of HTC Vive headsets and wireless controllers as their VR hardware, and Rec Room (2016) by Rec Room Inc. as the VR software used during all recording sessions. The video footage gathered during each session was recorded from three points of view: the first-person viewpoints of both Vive HMDs worn by the participants, and a ceiling camera which captured a 360-degree view of the recording environment. Each camera also captured its own audio track. The recordings of each point of view were edited into a single synchronized video for the purposes of the analysis itself.

The rest of this chapter elaborates on the data collection process, the technologies used, and participants involved in further detail. Challenges faced throughout the process will also be discussed in their respective sections.

### *3.1. The VR hardware: HTC Vive*

The Virtual Reality hardware used by the participants during the data collection sessions was the HTC Vive virtual reality system. The system consists of a Vive HMD worn by the user, a wireless controller for both hands, and a pair of base stations (also referred to as *lighthouses*) which track the user's location along with head- and hand-motions in the space they have been set up in. Gyroscopes, accelerometers, and other sensors are used to track the user's position and movements. The base stations enable the system to support 'room-scale' experiences, allowing the user's movements within a designated area in the real world to be reflected within the virtual environment. Room-scale functionality was utilized during the data collection sessions.

The Vive HMD offers a maximum field of view of approximately 110° (Vlachos, 2015), with a total display resolution of 2160x1200 pixels between two 1080x1200 OLED screens. It obscures the user's view completely; while the headset itself sports a front-facing camera through which the user can see the real space they are moving around within, the study participants did not make active use of the functionality. The headset also sports an integrated pair of headphones for audio playback and a directional microphone to capture the user's voice for either recording purposes or for in-game communication via VOIP.


Figure 1. Example a point-of-view (POV) screenshot from a Vive HMD, playing Rec Room. The teleportation pointer can also be seen (green line ending in a square).

As the Vive is not a wireless system, cable management – notably those of the HMDs themselves, as the base stations can (and are recommended to) be set up beyond the boundaries of the room-scale area in which user(s) are moving within – can become a concern, especially if multiple users are using the devices within the same room. While the Vive does alert users about reaching the boundaries of the space that has been designated for them to move within with a built-in Chaperone system, it cannot detect the location of cables or other Vive users within the same physical space. These factors did cause some difficulties during the recording sessions themselves, which will be elaborated on as we discuss the technical set-up of the sessions.

### *3.2. The VR software: Rec Room (Rec Room Inc., 2016)*

Rec Room is an online multiplayer video game developed and published by Rec Room Inc. (formerly known as Against Gravity), originally released for Microsoft Windows in July 2016. While predominantly marketed as a VR experience, the title is not exclusive to VR devices, and supports more traditional video game control methods (e.g. controllers, mouse-and-keyboard combinations) and 2D displays. The game places its players within a virtual world reminiscent of a recreational centre, from which various activities (game modes) or the players' personal rooms may be accessed. Each player has an avatar consisting of a torso, hands, and a head, with hand- and head-motions of users wearing HMDs and using motion controllers being replicated by the avatar. The avatar's head sports a simple drawn face, the expression of which changes as the player uses the in-game voice communication. The player has no direct control over their avatar's expression. The voice communication uses a microphone built into the HMD, and the broadcast of the speaker's voice takes into account the 3D environment they are within; in other words, factors such as the speaker's distance from other players in the virtual environment affects the volume and clarity of their voice as heard by any listeners.

In Rec Room, player engagement with the environment consists of moving within the space via teleportation and manipulating various interactable assets (also referred to as *props* or *items*) such as buttons or objects which the player can pick up with their hands. Some items enable the player to perform special actions: a paint gun enables one to draw shapes, while a camera allows a player to record clips of their environment in-game. The teleportation movement system has the player use one of their controllers to designate the location where they wish to move to by aiming at it with one of their controllers (see figure 1). This destination is visible to other players while aiming, and moving to it is accompanied by a sound of footsteps which others can also hear. Not all of the in-game audio is directly related to actions of the players themselves, as certain environments and activities are accompanied by auxiliary background sounds and music.

It must also be noted that the version of Rec Room that was used in the autumn of 2016 may not match the contemporary up-to-date version of the game. This is due to post-launch patches and content updates having altered the game's mechanics, content, and presentation since the title's original 2016 launch.

### *3.3. The participants*

A total of twelve participants took part in the recording sessions, each having been recruited by the six student groups participating in the course. Each group recruited two participants, all of whom were studying at the university. The participants were between 19 and 34 years of age, with the average age being 24,6. Gender distribution among the participants was predominantly male, with a total of 9 male participants and 3 female participants. Three of the participants held a nationality other than Finnish, and three of the participants spoke a native language other than Finnish; the others were native speakers of English, Spanish, and Indonesian. English was the language used during the recording sessions, as the university course for which the data was gathered on focused on interactions carried out in that language. Thus, most of the participants were speaking a second language. None of the participants proclaimed to have had prior experience with VR.

While the participants had received some general information regarding the course the material was being gathered for, they were not offered specific details about the subject(s) each group was interested in researching. Each participant signed a consent form for their participation in the study and filled out speaker description form with basic information. Some groups also carried out further interviews with their participants. These interviews were not, however, a mandatory feature of the sessions.

*3.4. The recording environment*

The recording sessions were carried out within the LeaF Forum at the University of Oulu. Each session took place within the same space, set up with two HTC Vive headsets and their respective controllers and base stations. Each Vive was connected to its own computer, running a separate copy of the Rec Room video game and capturing audio-visual footage via Open Broadcaster Software (OBS). Furthermore, a ceiling-mounted camera capturing a 360° view of the room was also used to record the session participants in the real world. The student groups themselves were not responsible for the technical set-up of the necessary equipment. Instead, this was carried out by LeaF laboratory manager Antti Siipo.

Figure 2. Example of footage captured by the ceiling-mounted camera. A blur filter has been applied over uncovered faces.

It is worth mentioning that the recording environment which was used for the data gathering session was not a typical environment for the use of VR devices. The usual intended use of room-scale VR is to have only a single person – the one wearing the VR device – within the designated room-scale boundaries, with no other persons present. In the recording sessions, however, the two participants shared the same room-scale environment, in addition to which other individuals (members of the student groups carrying out the recording, university personnel, etc.) were also present. This, combined with the cables connecting the HMDs to their respective computers, meant that the environment contained potentially hazardous clutter which had to be taken into account during the recording sessions themselves. It fell upon the non-participants present to ensure that those wearing the HMDs did not collide with each other or objects in the real world.

### *3.5. The recording sessions*

The individual sessions themselves were largely freeform in structure. While members of the student groups were present in the same space within which the participants were recorded, most opted to only try and offer minimal guidance for the individuals wearing the HMDs and engaging with the virtual environment. As the HMDs block the wearer's vision of the real world completely, the participants could not see others present in the recording environment. Any advice given by the students had to be offered verbally due to this limitation of the participants' vision. The goal was to record interactions with as little 'scripting' on the part of the students as possible, instead focusing on events which arose naturally within the virtual environment. Thus, the participants were allowed to engage with Rec Room freely over the course of the recording session, save for occasional guidance regarding, for instance, where the participants could find certain activities within the game environment. Some of these activities (or game modes or 'minigames', as they are sometimes referred to) ultimately proved better at facilitating interaction between the participants than others.

The students could not remain completely hands-off during most sessions, however. As the participants had no prior experience with either VR or Rec Room, guidance from the students was occasionally needed; the students themselves, meanwhile, by and large also only had limited experience with VR and the software being used, and could only offer a limited amount of assistance. Guiding the participants towards some of the game's activities was a common form of advice given, beyond matters relating to basic game mechanics. In addition, beyond the game itself, the students had to ensure that the two HMD-wearing participants did not collide with one another or objects within the environment. Moving the cables of the HMDs around to avoid creating a tripping hazard also proved a pertinent measure, for instance. They would also find themselves becoming involved with matters of technical troubleshooting in several instances.

The two participants recruited for the recording sessions were not the only users present within the virtual environment. Courtesy of Rec Room's nature as an online game and the fact that the game was being played in a public session, the participants could both see and, occasionally, interact with users who had been automatically placed into the same multiplayer session. These random, unscripted interactions ranged from brief greetings to conversations and more involved interactions in the context of some of the game's activities.

### *3.6. Handling and editing of the recorded data*

The recording sessions produced two kinds of data from three different sources: audio and video from the two HTC Vive head-mounted displays and from the 360-degree camera mounted on the roof of the room in which the recordings took place. In order to make the process of transcribing and analysing interesting sequences within the footage, the video footage from each source was synchronized and edited into a single video file; this had already been done for a portion of the footage (approx. two hours total) following the original 2016 recording sessions. While the video footage from all three sources could be edited into a single file, only audio from the HMDs was used in the combined footage; the audio recorded by the roof-mounted camera had also picked up other sounds in the recording environment, which would have obfuscated in-game sounds if combined with the HMD-sourced audio. The individual audio tracks of each HMD and the roof-mounted camera remained available after the editing, and could be examined independently should the need arise.


Figure 3. An example of edited (combined) footage. Leafvr2's POV is on the top-left, leafvr1's on the top-right, ceiling camera view on the bottom half. All instances of combined footage follow this positioning pattern.

Beyond the merged files, I also made use of the original recordings from each individual source during the transcription process in several instances. While the combined view was useful for observing motions, gesture, and the users' spatial orientation, the individual audio tracks from each source provided more clarity especially in instances where numerous in-game sound effects were playing. The built-in microphone of the roof-mounted camera proved particularly useful in transcribing the participants' speech without other audio interfering in instances where in-game audio made it difficult to ascertain what was being said, and when the participants had to interact with the people observing them. The high resolution of the source footage also proved useful for observing specific visual elements when viewed independently.

In the next section, we move on to discussing the research methodology applied in the analytical section of this thesis, and the relevant theory behind the phenomena we are interested in.

## 4. Methodology and theory: Conversation analysis, breakdowns, and repair in interaction

This section focuses on laying out the theoretical groundwork for the methods which we will use to carry out the data analysis itself. We shall begin by discussing conversation analysis (CA), which the analysis of this thesis is based on, in general terms and its applicability to analysing multimodal interaction. We then explore the concept of *breakdowns* in the context of human interaction with technology and computers, how they manifest, and their role as possible trouble-sources in VE/VR interactions. Finally, we discuss the phenomenon of repair in interaction and how repair manifests in the context of HCI.

### *4.1. Methodology: Conversation analysis*

Conversation analysis (CA), the methodology which we shall apply to analysing the data examined in this thesis, can be characterized as "the study of social action as achieved through the medium of talk in interaction" (Alasuutari et al., 2008, p. 437). It is a form of study which, in practice, begins by identifying some interesting kind of behaviour in naturally occurring social interaction (observation), followed by seeking other instances and exploring the nature of the phenomenon in depth (Sidnell, 2013). With a sufficient number of instances collected, the researcher can "begin to describe the practice or phenomenon in terms of its generic, context-independent properties, moving away from the particularities of any single case" (Sidnell, 2013, p. 78). The emphasis on naturally occurring interactions renders audio and video recordings of such interactions the preferred kind of material for CA study, in contrast with data gathered via means such as controlled sociological experiments or interviews (Mondada, 2013).

Among the key concepts which underpin CA study are turns-at-talk and turn-taking, turn design, social action, and sequence organization (Drew, 2005). Taking turns is "the most basic form of organization for conversation" (Drew, 2005, p. 80), while turns themselves are highly variable: their length, the order in which participants in conversation speak, and the purpose for which they are used are all malleable and vary from turn to turn (Sacks, Schegloff, & Jefferson, 1974). Turns themselves are constructed from smaller components, Turn-Construction Units or TCUs, which in turn can be sentential, clausal, phrasal, or lexical in nature (Drew, 2005; Sacks et al., 1974). Turn design refers to the speaker of the turn deciding what purpose the turn is to serve, and what it shall consist of in order to serve this desired purpose (Drew, 2005).

Social actions are the actions – asking, telling, correcting, etc. – which participants in conversation are performing while partaking in it (Kitzinger, 2012; Sidnell, 2013). Sequence organization in turn refers to how sequences of these social actions appear and progress in conversation (Drew, 2005; Kitzinger, 2012). Adjacency pairs are a common form of sequence organization, in which the current speaker performs an action, to which the recipient is expected to respond in a fitting manner; when posing a question (a first-pair), for instance, an answer (second-pair) is the expected response, or meeting a greeting (first-pair with another greeting (second-pair) (Kitzinger, 2012). Sequences may also be expanded on before their initiation or after their completion. A pre-expansion prepares an oncoming sequence (do you know what I'm thinking about? etc.); an insert expansion expands on either the first-pair or second-pair; a post-expansion occurs following the second-pair part, which can be minimal (does not trigger a new sequence, such as "okay" or "alright") or non-minimal (triggers a new sequence, such as by asking "really?" after receiving an answer – thus implying that more information is still desired by the recipient). Preference is also a factor in sequence organization, in the form of certain turns expecting specific responses from the turn's recipient (Sidnell, 2013). Positive responses are generally preferred over negative ones, for instance, and negative responses are often paired with attempts to mitigate them (Drew, 2005).

Interaction between individuals is not, however, all talk. Gaze, gesture, and physical artifacts within the environment in which the interaction takes place can all be relevant to said interaction; in other words, social interaction is *multimodal* in nature (Mortensen, 2012). This multimodality has been a subject of interest in the study of human-computer interaction, for instance, and is the focus of the methodological framework devised by Norris (2004) to study the multimodal nature of interactivity with a focus on the *embodied* (speech, gesture, gaze, etc.) and *disembodied* (environmental cues, furniture, signs, etc.) modes of interaction. But while Norris' framework – multimodal interaction analysis – considers these two modes independently, a CA approach to studying multimodality focuses on the interplay between multiple modes of interaction: “Multimodal analysis, from a CA perspective, attempts to describe how talk, visual resources (predominantly gesture, gaze and body posture), the use of physical artifacts in the participants’ surroundings, and the surroundings themselves are jointly used to perform coherent social action” (Mortensen, 2012, p. 2).

Whereas recordings of social interactions are the main type of data which CA studies, transcripts made of this data are a core tool of the analytical process of examining a given phenomenon. These transcripts “need to be detailed enough to facilitate the analyst’s quest to discover and describe orderly practices of social action in interaction” (Hepburn & Bolden, 2013, p. 58), and enable the researcher to observe turns, actions, and sequences play out within the interaction. Among the most common transcription conventions utilized in CA research are those laid out by Jefferson (2004), which we shall be applying to the data being analysed in this thesis. In addition, we will also be utilizing the multimodal transcription conventions by Mondada (2016) to better account for gestures, gaze, and embodied actions which occur during the analysis sequences.

It is to be noted that, while transcriptions are a common tool of CA analysis, the transcripts themselves are neither the subject of study, nor a complete log of *everything* that transpires over the course of a transcribed sequence of interaction. They are a tool with which the phenomena of interest itself is examined in depth, and are a result of a transcriber’s own subjective decisions on what aspects of interaction are relevant enough for this phenomena to include in the transcript (Wagner, 2012).

### *4.2. Breakdowns in human-computer interaction*

Problems inevitably arise when users interact with computers and software, be they caused by difficulties with software design, an unintuitive interface, or misplaced expectations on the behalf of the user themselves. Studying such problems has been a common topic of HCI research, and one approach to studying the topic has been to focus on *breakdowns*. Breakdowns have been studied in numerous HCI contexts, from issues users have experienced during computer-mediated communication (CMC) (Sharples, 1993), mobile learning (Sharples, 2009), how users adapt to working on IT-enabled tasks (Nicolajsen, 2019), evaluating the design of virtual environments (Marsh, 2001; Spagnolli, 2002) and interactive entertainment (Ryan & Sigel, 2009), and their occurrence and role in the context of video game design and gameplay (Barr, 2007; Iacovides, Woods, & Scanlon, 2011; Iacovides, Cox, McAndrew, Aczel, & Scanlon, 2015; Pelletier & Oliver, 2006). They can be used to study aspects of both the technology being used (design issues, general user experience, etc.) and the users who utilize it understand and use the technology for their own purposes (Sharples, 1993).

Sharples (2009) describes breakdowns as "observable critical incidents where a learner is struggling with the technology, is asking for help, or appears to be labouring a clear misunderstanding" (p. 32). While this definition focuses on *learners* due to the subject of the original paper, in the context of this thesis it would be more appropriate to apply it to *users* in a broader sense. In practical terms, breakdowns can be anything which interrupt the ongoing flow of work or gameplay or other use of the technology at hand, and can be triggered by causes ranging from technological failures to unintuitive design on the part of the software or input device(s), input mistakes on the part of the user, or the user missing critical information or being left unsure of where to go or what to do in the context of a video game (Iacovides et al., 2015; Sharples, 1993). Several breakdowns can both occur concurrently, and lead to others (Iacovides et al., 2011; Spagnolli, 2002). In regards to their impact on user interaction with a given technology, breakdowns force users to adjust their current course of action in order to overcome or circumvent them, such as by adopting new strategies in gameplay (Ryan & Sigel, 2009). They can relate to many aspects of a user's experience with a given technology, including their actions, understanding, and sense of involvement during use (Iacovides et al., 2011).

Breakdowns do not only impact how users use a device at a given moment, but also their sense of immersion and involvement in their task(s). When users are in a state of flow and immersed in what they are doing, the technological device they are using acts as an extension of the user themselves without them having to focus on the device itself (Ryan & Sigel, 2009). When this flow is interrupted, users can begin shifting their focus from what they are doing with the technological device to the device itself; in the context of breakdowns occurring in virtual environments, breakdowns can lead to the user's focus being allocated away from their interaction with the VE (Marsh, 2001). Users do usually exhibit a degree of tolerance to typical breakdowns internal to VEs (graphical/audio glitches, minor input issues, a degree of uncertainty regarding the VE's internal logic) without losing their immersion despite being aware of trouble of some kind having occurred (Iacovides et al., 2015; Marsh, 2001). However, an inability to overcome a major breakdown over an extended period of time can result in the user growing bored or frustrated (Iacovides et al., 2015; Ryan & Sigel, 2009). Breakdowns which completely block the user's progress at a task they are performing are particularly severe, and risk leading to disengagement and a loss of immersion. A diminished sense of agency can also lead to a lack of involvement and immersion on the part of the user, which can result from input-related issues, and actions taken by the user resulting in outcomes perceived as unexpected and/or undesirable (Iacovides et al., 2015). Breakdowns can also be examined from the perspective of their impact on a user's relationship with the technology they are using. If a device is considered ready-at-hand (RaH) by a user, they treat it as an extension of themselves while performing activities with it, i.e. they are not focused on it; if a device is present-at-hand (PaH), the user is consciously aware of the device and puts at least some of their focus towards it (Ryan & Sigel, 2009). When user actions are viewed from this perspective, transitions from RaH to PaH can be considered breakdowns.

While observing the breakdown itself is an important component of diagnosing the event, it is only the first part of the process. Identifying the triggers behind the difficulties during the actions leading up to the breakdown, trying to ascertain what the users are trying to accomplish when the breakdown occurs, and the users' responses to overcoming the breakdown are other important parts of the process (Iacovides et al., 2015; Nicolajsen, 2019; Spagnolli, 2002). Thus, identifying the breakdown event itself is but the first component of breaking down a breakdown – what is happening, where is it happening, when is it happening. The most reliable method of identifying a breakdown event is by observing the user: Sudden changes in the user's present course of action are a common sign of a breakdown event occurring, alongside vocalizations of surprise and confusion, repetitions of actions leading up to the breakdown, and the user engaging in trial-and-error behaviour (Ryan & Sigel, 2009; Spagnolli, 2002). We may use the act of one VR user trying to push the avatar of another user with their hands and failing to move them as an example of a breakdown event. In this example, one user cannot affect the position of another player's avatar, and the user attempting to push the other repeats the action several times upon failing to achieve the intended effect. Their repeat actions may also be paired with utterances hinting at confusion.

With the breakdown event itself having been identified, the second step is to try and ascertain what the user is trying to achieve with the actions they are attempting to perform when a breakdown occurs – i.e. the user-motivation component of why they are performing the breakdown action itself (Spagnolli, 2002). Our information on user motivations can never be completely free of presumption should the user themselves not explain their motivations at any given moment, but observation on the context of the action and various cues exhibited can lend these presumptions a degree of plausibility. If a VR user is making pushing motions with their hand towards another user's avatar, we can presume that they are attempting to affect the position of said avatar within the virtual space they inhabit. The exact reasons for why they are attempting to do so we must try and infer via whatever clues are provided both during the breakdown event and the actions immediately preceding it: Is the other player uncomfortably close? Perhaps their avatar is blocking access to something the user under observation wants? Is the user simply experimenting to find out whether they can move another player's avatar in the environment by pushing them? And so on.

Finally, there is the resolution of the breakdown – or lack thereof. How the user ultimately resolves the trouble they are facing (or fails to do so) may hint at knowledge they have gained of the system they are working with, or suggest unintuitive design in the system should the user fail to resolve the situation and carry through with their desired action (or, at least, learn why the action cannot be performed). It is at this stage in the breakdown process where we can expect to witness successful repairs-in-interaction. To carry on with the example of pushing another user's avatar, a successful breakdown resolution would see the user attempting the push realizing that they cannot affect the position of another's avatar through such means, and adjust their actions accordingly depending on their original motivation for desiring to move the avatar in question. If they were simply experimenting whether or not they could move another player's avatar, the user may simply move on to performing a new action. If the other player's avatar was in the user's way or simply at an uncomfortably close proximity, the user may opt to move their own avatar to a new, more preferable location.

Another concept worth briefly discussing in the light of breakdowns are *breakthroughs*. Whereas breakdowns refer to problems encountered by users, breakthroughs refer to the successful resolution of a breakdown in a manner which leads to both progress in the task the user is trying to accomplish, and learning something about the technology they are utilizing in the process (Iacovides et al., 2015). In contrast with his definition of breakdowns, Sharples (2009) describes breakthroughs as "observable critical incidents which appear to be initiating productive new forms of learning or important conceptual change" (p.33), with Iacovides et al. (2011) adjusting his definition to fit video game players successfully carrying out "a set of actions." (p. 6). Other kinds of breakthroughs are also described by the latter authors, namely those relating to involvement (players entering a flow-state) and understanding (players attaining success through coming up with new strategies).

While breakthrough events themselves are not a key focus of this thesis, it is reasonable to presume that a degree of overlap exists between successful repairs in interaction and breakthroughs. With that in mind, let us next proceed to exploring the topic of repair in interaction.

### *4.3. An overview of repair in interaction*

In the context of conversation analysis (CA), repair is "the cover term for a range of practices (…) by which trouble of all sorts is managed" (Fox, B., Benjamin, & Mazeland, 2012, p. 1) over the course of human interaction. It is a two-part process began by a *repair initiation* and followed by a *repair proper*, and involves two participants: a speaker and a recipient of the talk (i.e. other participant(s) in the interaction). Either the speaker or the recipient can perform both the initiation and the repair proper, with certain combinations being more frequent than others (Fox, B. et al., 2012). A clearly perceivable error does not have to occur in order for a repair to be initiated; a repair in which motivation for the repair cannot be ascertained is common to self-initiated self-repair in particular. Furthermore, self- and other-initiated repairs can exist within the same fragment of talk. From the point of view of turn organization, repair always affects the progression from one TCU to another, either by pausing progression or adjusting the manner in which social actions progress within a given sequence of interaction (Kitzinger, 2012).

For repair to occur, something which needs repairing must first appear over the course of an interaction. This *trouble-source*, also called a *repairable*, is the issue in interaction which the act of repair seeks to resolve (Kitzinger, 2012; Schegloff, Jefferson, & Sacks, 1977). No specific criteria for what is and is not a trouble-source/repairable beyond their ability to trigger a repair in interaction exists; in other words, beyond trouble-sources relating to verbal forms of interaction, gaze, gesture, environment, and features of the technology being used in mediated forms of communication can all produce trouble-sources (Arminen, 2013; Frohlich, Drew, & Monk, 1994; Greiffenhagen & Watson, 2009). As summarised by Schegloff, Jefferson, & Sacks (1977): ”It appears that nothing is, in principle, excludable from the class ‘repairable’” (p. 363). As turn-taking is an intrinsic component of human conversations, resolving trouble related to it is one of the common functions of repair in conversation: “Repair mechanisms exist for dealing with turn-taking errors and violations; e.g., if two parties find themselves talking at the same time, one of them will stop prematurely, thus repairing the trouble” (Sacks et al., 1974, p. 701). Interruptions (Who, me? Excuse me?), false starts, repeats, and premature stoppages are examples of repair methods which are often deployed in face of trouble sources which relate to turn order and turn-taking within conversation.

Beyond the question of who initiates the repair – the speaker or a participant – the sequential location of the initiation itself is a key factor in repair organization (Arminen, 2013). Repairs are not consistently initiated at a certain specific point during interaction. Instead, “repair-initiation opportunity spaces” (Fox et al, 2012, p. 2) can occur in various parts of an interaction: they may occur in the same TCU where the trouble-source itself is located, in the transition space between turns, in the turn which immediately follows the trouble-source turn, or in the so-called third position or within the third turn.

Self-initiated repairs are repair actions which are initiated by the speaker of the TCU which contains a trouble-source. Self-initiated repairs carried out within the same TCU are the most common type of repair in English-language interactions (Kitzinger, 2012). An entire self-repair operation usually begins with the speaker cutting their current speech short, producing a repair solution for the trouble at hand, and carrying on with the interrupted speech if able. These interruptions are most commonly carried out via a cutoff, uh/um, or silence (Fox, B., 2012; Kitzinger, 2012). The repair actions themselves often take the forms of recycling (repeating sections of preceding talk in the trouble-source TCU) or performing a replacement (substituting a section of preceding talk in the trouble-source TCU with something else), or combinations thereof. Insertions – adding new information to a prior statement – can also be found, as can abandonments, in which “the speaker begins a TCU and then abandons it, beginning instead a new TCU” (Fox, B., 2012, p.3). While self-initiated repairs are most commonly performed within the same TCU, they can also be performed in the transition space between turns, the third turn, and in the third position (Kitzinger, 2012).

Other-initiated repair refers to repair which is initiated by a recipient of the turn containing a trouble-source, (Kitzinger, 2012). It can lead to the repair proper being carried out by the speaker of the trouble-containing TCU, or by the a recipient – either the repair initiator themselves, or another ‘other’ present in the interaction (Benjamin & Mazeland, 2012; Kitzinger, 2012). While self-initiated repair pauses the progression of a turn, other-initiated repair affects the progression of the interaction sequence at play (Kitzinger, 2012). The most common form of other-initiated repair is other-initiated self-repair, where the recipient in the conversation indicates that a problem exists, but leaves the repair proper for the speaker to carry out: “the most common way in which others (e.g. recipients) deal with trouble in speaking, hearing, or understanding is to initiate the process of repair (…) but to leave the opportunity to provide the repair solution to the speaker of the trouble-source” (Kitzinger, 2012, p. 249). When an other-initiated repair is initiated, it is often up to the initiator to select the next speaker who can take a turn (Bolden, 2018).

Recipients have many options for initiating repair, depending on how they identify the trouble-source, and how accurately they assess the nature of the trouble (such as whether it is related to hearing, understanding, or accepting of the trouble-source). Nonverbal cues, such as furrowing of brows, can also play a part in the initiation, in addition to which regular turn-taking practices can be overridden during repair interactions (Benjamin & Mazeland, 2012). The latter is typical when a participant can claim to be an authority on the trouble-source (Bolden, 2018). Initiations positioned adjacently to the trouble-source aid in the source's identification, whereas in non-adjacent initiations, recipients can offer further aid in helping the speaker identify the trouble-source such as by asking questions (Benjamin & Mazeland, 2012). Other repair operations typical of other-initiated repair are repeating the trouble-source TU, requesting elaboration on the trouble-source, or a corrective replacement or a confirmation when the initiation is a candidate understanding. The initiators may also be repeated, often with increased strength from the initial utterance (Kitzinger, 2012). Gestures can also other-initiate a repair without a verbal element (Arminen, 2013; Kitzinger, 2012).

The less common form of other-initiated repair, where the other also carries out the repair proper, often manifests in the form of a participant in a conversation repairing an item in the talk of others – i.e. correcting them. Such an interaction "involves a claim of greater authority or better access to the trouble-source" (Benjamin & Mazeland, 2012, p.4) on the behalf of the individual conducting the repair, and the speaker of the trouble-source turn can either contest or accept the correction being made. In instructional settings, those being corrected may verbally confirm their acknowledgement of the correction after receiving it and adjusting their actions accordingly (Levin et al., 2017).

Third-position repair, also known as repair-after-next-turn, is a repair method to try and remedy a breakdown of intersubjectivity (i.e. mutual understanding) between the participants in a conversation (Schegloff, 1992). It is a type of repair which occurs when the aforementioned breakdown becomes evident during the second turn of the sequence, prompting the speaker of the first turn – which contains the trouble-source – to enact repair in the third position. This repair can take the form of a straightforward correction, reformulation of the trouble-source, adding more specificity to the trouble-source, elaborating on the purpose of the turn containing the trouble-source, and characterizing the trouble-source as something else (e.g. 'I was just joking…') followed by withdrawing from the topic. A back-and-forth between the speaker who produced the trouble-source turn and the other party in the conversation is a common trait of third-position repair, which often follows the sequence a repair initiation, an agreement, a rejection, and the repair proper (Fox, B. et al., 2012).

*4.4. Repair in human-computer interaction*

Human-machine interactions "seem prone to troubles that have to be remedied for [an] action to continue" (Arminen, 2013, p. 23), and computers are no exception to this observation. The phenomenon of repair has been a subject of study within the field of HCI, be it in the form of interactions between human and computer, and interaction between humans using a computer or mediating their interactions through a computer. In the former context, the interaction takes the form of an input-output sequence, wherein the user inputs a command or request, and the computer responds by following through on the request and producing a desired result. A 1994 study by Frochlich, Drew, & Monk is an example of past CA research into the first kind of HCI interaction, where the commands input by the user were considered the user's means of interaction, while user interface elements – text boxes, error messages, etc. – were considered the computer's means of interaction. An error message could act as an initiator for other-initiated self-repair, for instance. In regards to differences found between HCI and regular human-human interaction, the authors found that "getting the computer to recognize what you mean is considerably more difficult than the comparable activity in conversation" and "whereas participants can be assumed to share the same rules and procedures of interpretation borrowed from ordinary conversation, they cannot in HCI. Users employ rules of interpretation borrowed from ordinary conversation or learned from interactions with other computers" (Frohlich et al., 1994, p. 415). The latter in particular was identified as a factor prone to undermining intersubjectivity. The authors also argue that repair had two distinct functions in the interaction which they studied: beyond being a means of correcting flawed inputs, it also served the function of allowing a user to "learn the system's action grammar for interpreting inputs" (p. 416). Furthermore, difficulties arose in the way of failed repair attempts, failures to identify trouble-sources, and repair actions themselves leading to further problems.

An example of study into repair in human-human interactions using a computer is offered by Greiffenhagen & Watson (2008). Their study is a piece of CA research in the context of HCI focused on *visual repair*: “The identification and correction of things that perceivedly ‘go wrong’ for participants on the computer screen” (Greiffenhagen & Watson, 2009, p. 2). This was the focus on their study done on collaborative work in front of a computer. When trouble arose during collaborative tasks, participants often formed into user-helper pairs where one person carried out the desired actions, while the other observed and offered advice verbally. Some, however, opted for a cooperative approach, where both participants focused on completing the desired action at the same time. Embodied conduct can produce trouble-sources just as the talk within a conversation can, “in particular” in HCI contexts (Greiffenhagen & Watson, 2009, p. 9).

As with regular conversations, individuals adjust their speaking in an effort to make their message clearer when interacting with computers, with repair being one amongst the means of achieving this (Raudaskoski, 1990). Much of repair in the context of HCI “tends to be connected with misunderstanding and the computer system not working ‘intelligibly’” (Raudaskoski, 1990, p. 155), with users not understanding the system they are working with and the instructions it provides. This, in turn, can lead to users providing the system with commands it does not understand, leading to output which does not match the user’s expectations or no output whatsoever, i.e. unresponsiveness. Both are commonly taken as signs of trouble having occurred when operating machines (Arminen, 2013). The occurrence of such troubles can lead to a *remedy of action* rather a repair of talk, wherein the act of repairing something – often an environmentally-coupled action – occurs without a problematic utterance, or any utterance whatsoever, and can be viewed as a form of environmentally-coupled repair (Arminen, 2013; Nevile, 2007).

Computer-mediated communication (CMC) can also feature additional challenges for effective interaction. In their CA study on a conversational lesson carried out via SCMC (synchronous computer-mediated communication) where the participants had access to both text- and voice-chat as communication options, Nguyen (2017) found that the two different forms of communication could sometimes compete for the participants’ attention.

Repair cues and repairs themselves could go unnoticed, for instance, when they were communicated via one means of communication while the recipients' attention was on another. Regardless, the participants did not treat the different means of communication "as separate semiotic means; rather, their affordances can be employed holistically and dynamically in the contingent unfolding of social actions" (Nguyen, 2017, p. 113).

Next, we shall move on to analysing the effects of breakdowns on VR interactions, and how users respond to and attempt to remedy them.

## 5. Analysis: Breakdowns as trouble-sources in VR interaction

This chapter will be broken into two sections based on the nature of the sequences being analysed. The first section examines breakdowns which relate to Rec Room's internal mechanics and rules, while the second examines breakdowns which relate to the VR devices being used – i.e. the HTC Vive HMD and the Vive controllers. The clips chosen for transcription were picked from footage featuring participants from Groups 2, 4, and 5. The participants will be referred to as *leafvr1* (top-right view in figures unless otherwise stated) and *leafvr2* (top-left view in figures unless otherwise stated) based on the in-game usernames attached to the two VR systems in use. Other users shall be referred to with either their own in-game usernames or pseudonyms on case-by-case basis. Finally, if members of the student groups observing the participants in the real world become involved in the interaction, they will be labelled as *observers* (observer1, observer2, etc. if several become involved within the same excerpt). The findings will be discussed at the end of each section.

### *5.1. Game-related breakdowns*

In this section we examine three sequences where the breakdowns experienced by the participants are caused by factors internal to the game they are playing, such as unclear game rules and mechanics. The first sub-section focuses on a pair of sequences where unclear game rules result in confusion, while the second sub-section examines an instance where mechanical limitations and other users' avatars cause trouble to the participants. Each sequence takes place within Rec Room's 3D Charades game and are sourced from the recording sessions of Group 2.

#### *5.1.1. Breakdowns related to understanding of game rules*

First, some relevant context regarding the rules and mechanics of the game mode the users in the following excerpts are playing within Rec Room. The game in question is a VR interpretation of the Charades word-guessing game, referred to as 3D Charades within Rec

Room, where the goal of one person is to explain a given word through nonverbal means within a limited amount of time. In 3D Charades, the person whose turn it is to carry out the explaining is given access to a paint gun with which to draw shapes which either depict the word being explained or offer hints towards it. The explaining player picks randomly selected words to explain from a box within the environment at the beginning of the round, and once the correct guess has been made, hits a button which ends the round in a success. If no correct guess is made, the round ends as the allotted time runs out.

By and large the participants who engaged with the 3D Charades game mode understood the game's various mechanics rather intuitively. However, some of the game's underlying systems were not as clearly communicated to the player. The environment in which the 3D Charades take place in is modelled after a theatre, with a raised stage with the aforementioned paint gun, word box, and the button which is used to end the round. At the beginning of each round, one player is placed (or 'spawned') on-stage as the designated explainer, behind a desk holding items relevant to their role in the game. Only the user who spawns on-stage as the explainer can read the words drawn from the box, while the guessers see ”No Peeking!“ instead should they attempt to read the cards – a built-in anti-cheating measure, in other words.


Figure 5: A card viewed by the designated explainer in 3D Charades (left), and a card viewed by a guesser displaying the 'No Peeking!' message.

This measure would become a trouble-source and a breakdown trigger for some participants playing 3D Charades. As they were not made aware of the mechanic, users casually switched explainer-guesser roles at the beginning of a new round, unaware that all other users save for the one who had originally spawned on-stage would be unable to read the words which they were to draw and explain to others. With all this in mind, let us examine the following interaction between leafvr1 and leafvr2 as they begin a new round of 3D Charades:

(1) No Peeking! (G2C2)

```
                ((A new round has just begun, and leafvr2 addresses leafvr1))
LEAFVR2:        you wanna do it? .hh
                (0.8)
LEAFVR1:        yeah, yeah (0.9) just drop it over there
                ju*st- there we go
leafvr2           *drops the paint gun*
                (5.9)
leafvr1         +retrieves the paint gun, relocates behind the table, draws a
                "No Peeking!" card+
LEAFVR1:        not this again, no
                +discards card 1, draws another, identical one, pauses+
leafvr2         *films leafvr1, drops camera he's holding, takes another*
                (2.6)
LEAFVR1:        whaaat?
                +discards card 2, draws card 3+
LEAFVR1:        (0.8) ohhh, I see (.) hah (0.8)
                +discards card 3, teleports away from the table+
                you should do it, this, because every time I pick up
                like, uh, (0.6) a thing from the box it just says no peeking
LEAFVR2:        no peeking?
LEAFVR1:        it's your turn, like, th- the game thinks it's your turn
                (.).hh daaaamn
```

Shortly after spawning as the designated explainer, leafvr2 asks leafvr1 whether he wishes to handle the explaining instead (line 2). The latter agrees, moving to the explainer's position while acquiring the paint gun from leafvr2 in the process. leafvr1 then retrieves a No Peeking!  card from the word box, mistakes it for a repeat from a previous round of the game as exemplified by his "not this again, no" (line 9), and summarily discards it. It is only upon drawing the second card where VR1 realizes that something is not working the way he has been expecting it to. A pause occurs, in turn followed by the discarding of the second card while voicing his confusion about the situation with a drawn-out "what?" (line 14). He then realizes what has happened (line 17) and proceeds to inform the other player (lines 19-12).

The breakdown faced by the users (leafvr1 in particular, with him being in the role of the explainer while leafvr2's focus is elsewhere for much of the sequence) ultimately occurs between lines 11-14 in the above excerpt. It is encountered following his drawing of the second identical card, resulting in a pause, voiced confusion, and repetition of the same action. The underlying cause is a lack of understanding or awareness of the game's subtler mechanics on his behalf, as the game has not provided either of the participants with any reason to think that such a limitation could be in play. In the breakdown patterns of Ryan & Siegel (2009), this could be considered an example of a breakdown in the developing strategy, more specifically event-triggering: leafvr1 is expecting the cards he draws from the box to be randomized regardless of who draws them (as a real world boxful of random word cards would presumably work), but this proves not to be the case. This is despite leafvr1 having already encountered the No Peeking! card during a prior round, as exemplified by his remark on line 9. The anti-cheating mechanic at work has thus already interfered with leafvr1's and leafvr2's 3D Charades experience and has gone undetected, with them only becoming cognizant of it over the course of this sequence.

The disruption caused by this breakdown does not last for an extended length of time, however. After drawing the third No Peeking card in a row (line 15), leafvr1 understands that the repeated cards are not a result of chance, but a mechanic of the game itself at work (line 17). With the realization of what is happening (reached through repeating the action several times and drawing a conclusion based on it), the user proceeds to correct the situation by remedying their action, getting leafvr2's attention and explaining the problem to them. Him teleporting away from the table (line 18) also serves at least two functions: It brings leafvr1 closer to the other participant as he tries to get his attention, while also freeing up the explainer's designated spot next to the word box for leafvr2 to claim. In this way, leafvr1 is not only remedying their own conduct in the interaction, but also sharing information to others in order to keep them from the same trouble-source repeating itself in the future. It is during this explanation where we can also observe minor instances of self-initiated self-repair in leafvr1's speech: note his utterance of "uh" as he attempts to find the correct term for the word cards (line 19), ultimately settling for a substitute term ("a thing from the box"), and the manner in which he performs a cutoff before completing his explanation regarding what he thinks the game wants the users to do (line 21). The explanation on line 21 is also a further elaboration on leafvr1's prior turn, prompted (other-initiated) by leafvr2's response (line 20), who has apparently been left confused about leafvr1's initial explanation of the problem on lines 16-19.

As leafvr1 and leafvr2 have also gained a better understanding of the game's internal workings in the process of resolving the trouble, the act also fits the definition of a breakthrough. This knowledge that players can enjoy different privileges depending on their role as assigned by the game proves useful information to have when troubleshooting future troubles, such as in another breakdown which would emerge but a few minutes later. Leafvr1 and leafvr2 have been joined by another user with the username *subcomandante* (referred to as *SC* in the transcript), and the users carry on playing 3D Charades as a larger group. A fourth player, TotalVR, briefly enters the space but does not interact with the others in this sequence. At the end of a round, subcomandante correctly guesses a word drawn by leafvr1, who in turn instructs subcomandante to push the Got It! button. This button is used to trigger the end of the turn and add a point to the game's scoreboard. Things do not play out as the participants expected, however, and this leads to the following sequence of events:

(2) "Push the button." (G2C3)

((sc has just successfully guessed leafvr1's word))
LEAFVR1: ya got it, push the button
(2.1)
I shouldn't even say anything while doing this but it(h)'s
it's *hard to (.) to show (2.7) *
sc *pushes the Got It! button, nothing happens*
LEAFVR2: +there's a hole, and-
+gesturing at leafvr1's drawing+
LEAFVR1: did you- did you push it? (( to SC ))
*#(2.1) *
sc *pushes the Got It! button several times to no effect*
fig #fig6
LEAFVR1: okay (0.7) it doesn't work (.) why?
SC: [haah ]
LEAFVR1: [the g]ot it button
(2.4)
LEAFVR2: because ^you-
leafvr1 ^leafvr1 pushes the button, round ends^
oh I had to push it
(( SFX: 'Game over!' ))
SC: ahh
LEAFVR2: that was it
(1.7)

```
LEAFVR1:     oh (.) this- the mechanics they have in this, um (.6) game is
             pretty (0.4) int(h)eresting
```


Figure 6. subcomandante pushing the Got it! button (green) to no effect. Note the green portion of the button depressing under his hand in response – it is only the end-of-round trigger which fails to activate.

Once again the game has placed restrictions upon the players based on their roles during the round (explainer vs. guesser): As with the drawing of word cards, only the explainer can trigger the round-ending Got It! button, a rule which the users appear to be unaware of. Just as with the No Peeking! cards, the root of the breakdown (which is triggered at line 9) lies in a lack of communication of the game's internal rules to the players. This, in turn, leads to a number of issues, including confusion in the interaction between users (leafvr1 asking subcomandante whether he had pushed the button just after he had done so), an unsuccessful attempt at remedying the issue (subcomandante pushing the button again several times after leafvr1's question), and drawing wrong conclusions about the game's functional state (leafvr1 presuming the button simply not working as intended based on their prior experience with 3D Charades). Had they failed to resolve the issue and been left under the impression that they had just experienced a fault in the software, the users' trust in the polish and internal consistency of the virtual world may have been undermined

(Iacovides et al., 2015). This may, in turn, have led to a diminishing of the users' sense of engagement and presence. Even as they ultimately become aware of the game mechanic at play, at least some of the users appear unsure of the logic behind them. This is exemplified by leafvr1 characterizing the game's mechanics as "interesting" in a laughing tone (likely referring to both the button-related confusion *and* the No Peeking! cards from before).

Despite initially thinking that they may have experienced a mechanic rendered inoperative by a software glitch, the users manage to successfully overcome the breakdown fairly quickly after it occurs. The actual resolution of the problem is handled by leafvr1, who proceeds to push the button themselves and successfully triggering the end of the round (line 17). While this happens without external prompting on his part, leafvr2 appears to begin voicing a suggestion for him to do just that at the very moment leafvr1 moves to push the button, cutting leafvr2 short in the process. If this was the case, then both of the study participants came to a similar conclusion at nearly the same moment and proceeded to initiate repair. This may have been influenced by knowledge gained during the No Peeking! sequence which took place but minutes prior. Leafvr1 also performs self-repair between lines 13-15, partially concurrently with subcomandante's utterance; having just remarked on how "it" does not work, leafvr1 quickly clarifies further by specifying that he is referring to the Got It! Button. Instances of self-initiated self-repairs in the form of word searches also seem to appear between lines 24-25: leafvr1 first appears to have trouble finding the word "game" based on him uttering an "um" followed by a brief silence before settling on the word, in turn followed by another pause before he characterizes the game's mechanics as "interesting." The former could be a sign of uncertainty on his part as to what a game within a game such as Rec Room's 3D Charades should be called. Leafvr1 asking subcomandante whether he had pushed the button or not on line 9 could also be seen as an instance of other-initiated repair, as it is his question which indicates to subcomandante that trouble has occurred, prompting him to repeat his prior action of pushing the button.

### *5.1.2. Breakdowns related to avatar interaction*

As is the case in real life, other individuals can be disruptive to interactions carried out within a virtual world, be it through means such as interrupting a speaker mid-turn or performing actions which prove distracting. In VR, users can feel a high degree of ownership towards their avatar alongside a strong sense of presence within the virtual world; this can lead to users responding intuitively to events which affect their avatar as if it were their own body, and may also develop expectations on how they can interact with avatars of other users (and their appropriate reactions to it).

Earlier, leafvr1 and leafvr2 were joined by a third player with a username consisting of a random string of numbers (henceforth referred to as *Numbers*). This user appeared to be of a very young age and/or unable or unwilling to interact with the two study participants, who carried on playing 3D Charades. Both leafvr1 and leafvr2 seem to presume the former to be the case, as they refer to Numbers as a “kid” several times in the footage (including line 18 below). As a new round of the game is triggered, the following sequence of events takes place:

(3) Personal space (G2C1)

```
leafvr2      *spawns on-stage as a new round starts*
numbers      +teleports to leafvr2+
             +waves an object at leafvr2---> 1.24
LEAFVR1:     *what is he doing?                  *
leafvr2      *attempts to wave numbers away from him*
NUMBERS:     .hh
LEAFVR1:     *ha ha #ha                                      *
leafvr2      *attempts to shove numbers’ avatar away from himself*
fig                 #fig4
NUMBERS:     .hhhh
             +the waving of the object at VR2 continues+
leafvr2      *turns to look back at VR1 briefly*
LEAFVR1:     heh heh heh
LEAFVR2:     .hh
             (0.6)
LEAFVR1:     he- he wants it real bad (.) *just give it to him      *
leafvr2                                   *keeps pushing----------->*
             (.) just- let the kid have it (.) .hhh hah hah heh .hh
             what is happ(h)ening?
LEAFVR2:     I- I *tried to push him away but-               *
                  *turns to face leavr1, turns back to numbers*
             (.) y’know it d(h)oesn’t w(h)ork
LEAFVR1:     I think it’s his turn anyways
             *so just teleport here, let him have it*
leafvr2      *grabs numbers’ object, throws it      *
numbers      -->+teleports after the object+
             (0.9)
leafvr2      *teleports off-stage*
             (1.5)
LEAFVR1      let him have his +own- (.) he’s taking both of the cameras  +
```

31 numbers +retrieves camera items from the ground-->>+

Instead of being able to play the round out as normal, leafvr2 is immediately side-tracked by the unexpected presence of Numbers' avatar at a very close proximity, forcing a shift of focus from the Charades game to the other user. In this instance, the user is encountering a breakdown with multiple components. While the most obvious breakdown is one of disconnect between what the user *wants* to do and what the game *allows* him to (to push back the offending player's avatar vs. the game mechanics preventing users from affecting the position of other players' avatars) which is triggered on line 8, the sudden appearance of Numbers in leafvr2's personal space and their close proximity (line 2), and the visual feedback of the object clipping through leafvr2 and his view repeatedly as it was being waved at him (line 11) were all likely to affect leafvr2's initial reaction to a greater or lesser extent. The first-person viewpoint and the clipping of an object through leafvr2's vision could be considered an example of a breakdown in *perceiving the environment* in the breakdown patterns of Ryan & Siegel (2009).

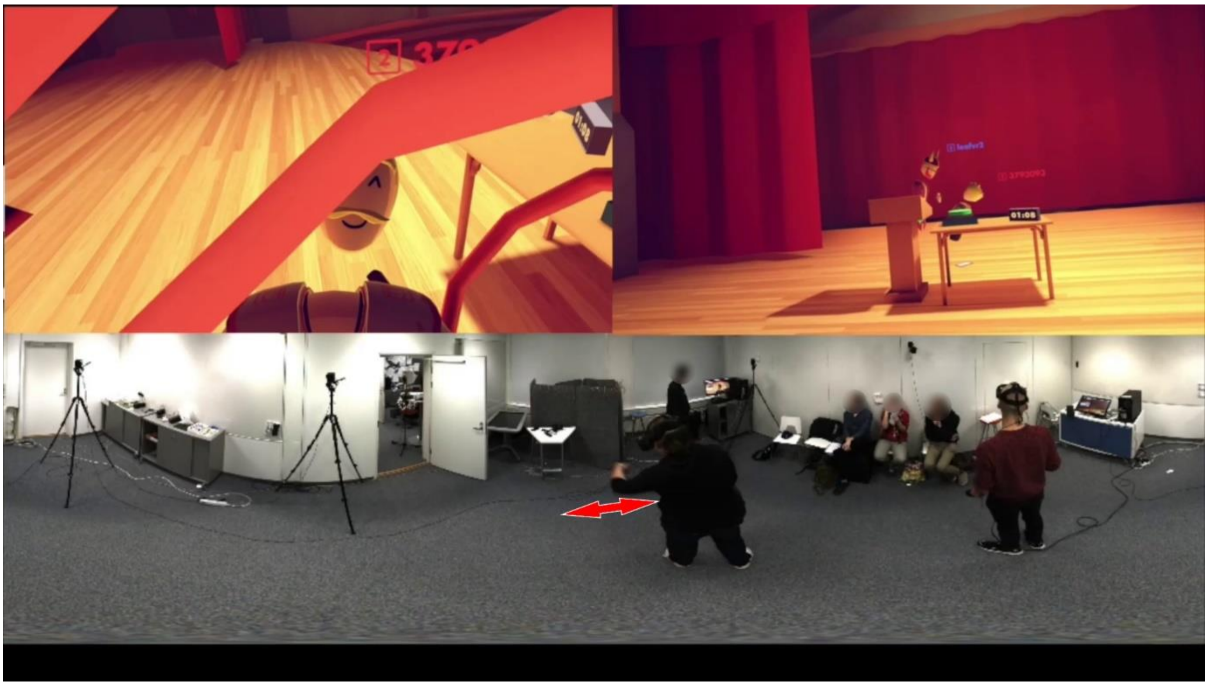

Figure 4. Leafvr2 attempting to push Numbers away. The arrow in the roof-mounted camera view (bottom) indicates the repeated pushing motions of his left hand.

Between lines 16 and 18, we can observe leafvr1 offer leafvr2 advice on what he should do to resolve the situation he has found himself in, doing so without verbal cues from leafvr2. Granted, the manner which leafvr1 proceeds to offer advice is not particularly clear; he keeps telling leafvr2 to give “it” to Numbers (lines 16 & 18) without specifying what he is referring to. It is possible that leafvr1 presumes that Numbers wants the paint gun leafvr2 is holding, or leafvr1 has misread the situation and thinks that it is leafvr1 who is waving the object around. Regardless, in this instance leafvr1 adopts a role somewhat akin to an instructor helping the other participant resolve the trouble he faces, carrying on all the way to the end of the excerpt (line 31). This places the two participants into a relationship reminiscent of a instructor-student pair seen in Levin et al. (2017) or a user-helper pair observed by Greiffenhangen & Watson (2008), albeit one where leafvr1 is responsible for most of the speech as leafvr2 is preoccupied dealing with Numbers save for his explanation on him trying to unsuccessfully push the offending user away. Leafvr2 only appears to follow leafvr1’s advice after he has successfully removed Numbers from his personal space, in the form of teleporting away from the table on line 30. The actual resolution of the trouble was carried out by leafvr2 himself, who adjusts his approach by successfully grasping the object held by Numbers and throwing it away, prompting the latter to chase after it (lines 24-26).

It is also worth noting that leafvr2 either did not think to or chose not to take certain alternative courses of action available to him. He did not attempt to verbally instruct Numbers to move away from him, for instance, and nor did he move his own avatar away from theirs despite having access (and having already grown familiar to) Rec Room’s default form of locomotion, teleportation. In fact, leafvr2 does not utilize the teleportation function at all in the excerpt until leafvr1 verbally prompts him to (line 27). It could be argued that the nature of leafvr2’s reaction is an example of response-as-if-real (Slater, 2009), with the user reacting in a manner which should have instinctually worked in the real world instead of using tools such as the aforementioned movement option to remove his avatar from the situation. It also suggests that leafvr2 is experiencing at least a degree of body ownership illusion, considering his response to Numbers’ extremely close proximity and that of the object they held to his own VR avatar (Bergström, 2016; Iachini et al., 2014). All this would also be in line with observations made by Iacovides et al. (2015), who found that breakdowns in action (attempting to push another avatar away, in this case) do not necessarily diminish the level of involvement (and, in the context of VR, presence) a user feels while engaging with a game so long as said breakdowns do not prove too disruptive to the experience.

*5.1.3. Discussion*

Before proceeding to device-related breakdowns, let us discuss some commonalities which can be observed in the three excerpts observed thus far (1, 2, and 3). Repetition of actions has been a recurring occurrence, be it in the form of repeatedly attempting to shove Numbers' avatar away in excerpt 3, the trio of cards drawn in 2, or subcomandante pushing the button several times under leafvr1's instruction. Users affected by the breakdowns also often voiced their confusion upon encountering a problem. The breakdowns encountered in all the excerpts were ultimately not major hindrances to the users' experience of the game, as they were resolved fairly quickly (within max. one minute of each breakdown occurring) and with minimal negative consequences for the users (major loss of progress in the game, etc.). These are typical of "breakdowns in action and understanding" as they are referred to as by Iacovides et al. (2015, p. 218). Their impact on the participants' level of engagement is left unclear; while user involvement in a game can *increase* as they overcome breakdowns (Iacovides et al., 2015), it is difficult to determine whether this was the case in excerpts 1, 2, and 3 in based on the recorded footage alone. It is also worth noting that in excerpts 2 and 3 multiple users involve themselves in the troubleshooting process once it becomes obvious that one of those present is facing difficulties with the game, or something is not working as expected.

As the participants of every research group were first-time VR users, macro-level player expectations based on prior experience with VR titles were unlikely to be a factor in the users' interactions within Rec Room (Iacovides et al., 2015). The participants did, however, bring with them their expectations of how things should work based on their experiences in the real world and with other computer software (Frohlich et al., 1994), and clearly developed expectations regarding the rules and internal mechanics of the game as they kept playing. In instances where these expectations were not met, the users were prone to reacting with confusion. Rec Room's conservative approach to offering users information with minimal UI elements such as text boxes or error messages seems to leave learning through trial and error as the most efficient method of exploring the game's systems, rules, and limitations. This lack of information provided by the game (and the lack of prior experience with it on the participants' part) perhaps contributed to the word searches which occurred when as leafvr1 was explaining a breakdown's cause (excerpt 1, line 19) or commenting on the game in broader terms (excerpt 2, lines 21-22); the user in question had not played the game long enough to come up with terms of his own, and the game had not provided him with in-game terminology to use. Of course, the participants being non-native speakers of English may have also been a factor in the observed instances of repair.

But how could these repeating issues regarding communication and feedback (or lack thereof) to the users on the game's behalf be addressed by changes in the game's design? Clearly differentiating the player picked as the explainer could be one course of action. Applying a special outline to their avatar or affixing a temporary title to their username could alleviate some confusion borne of unclear roles throughout the rounds of 3D Charades. A solution more specific to the No Peeking! cards would be to alter the card's text; instead of simply reading "No Peeking", the card could also include a string of text along the lines of "this card is only readable by (username)", with (username) replaced by that of the current explainer. Finally, regarding the sequence of events in excerpt 2, making the button completely unresponsive to the touch of anyone but the explainer would have avoided the mistaken conclusion drawn by leafvr1 of the button being broken. Had the button not been visually pushed down by subcomandante's attempts (perhaps flashing red in response as a further visual cue), the users would have had better odds of ascertaining that the source of the problem was not the button itself but the individual attempting to activate it. While such features might not have eliminated the trial-end-error behaviour and repetitions of prior actions completely from these sequences, they could have expediated the repair process.

### *5.2. Device-related breakdowns*

In this section, we examine five sequences where the breakdowns experienced by the participants are caused by technical issues relating to either the HTC Vive HMD or the Vive controllers. The first sub-section focuses on a pair of sequences relating to the former, while the second examines three sequences relating to the latter. These sequences take place within the Disc Golf, Paddleball, and Shield Soccer games within Rec Room, and are sourced from the recording sessions of Groups 4, 5, and 6.

### *5.2.1. Breakdowns related to VR display devices*

While the prior segments have focused on breakdowns related to the software the users were using alongside the actions of other users, the devices they use to interface with the virtual environment can also source of trouble. With a VR system like the HTC Vive, each of the main components (HMD, controllers, motion-tracking base stations) can potentially cause difficulties for users. Several instances of breakdowns relating to the HMDs and controllers took place within the data, with examples of the former being examined in this section, and examples of the latter in the next section.

We will begin with an excerpt wherein leafvr2 of Group 6 encounters issues relating to their HMD while playing Rec Room's disc golf. As was the case with the participants of Group 2 in the earlier sequences, the leafvr1 and leafvr2 have been joined by a third player with the username Neiltwo2. The three proceed with the game of golf, eventually arriving at the following interaction:

(4) "What happened?" (G6C2)

```
leafvr2      *rears hand back, throws their frisbee*
             (.) ((the frisbee falls into a hazard))
LEAFVR1:     noooo
LEAFVR2:     niiice
             ((leafvr2's footage experiences a sudden resolution drop))
             wh- what happened? (.) oh
             *lowers gaze down, looks at hands/controllers*
NEILTWO2:    it went over (.) I should've-
LEAFVR1:     yeah (.) it's at the water again
LEAFVR2:     I'm having some kind of- (1.1) thing, here
             (.6) (    ) (2.2)
LEAFVR1:     uh, no, I think- where did it go?
             (0.6)
LEAFVR2:     noo! ((footage from leafvr2's HMD cuts to black))
             ((UI prompt: "leafvr2 left the game"))
```

```
16  OBSERVER1:  is it working?
17  LEAFVR2:    I left the space
18  LEAFVR1:    oh yeah, yeah, I think it's just-
19  OBSERVER2:  +did it turn off the computer+
20              +moves to check on the computer in the recording room+
21  LEAFVR1:    oh okay
22  LEAFVR2:    I don't know
23  LEAFVR1:    I'll play this (.) until the end, I think
24  LEAFVR2:    yeah sure (0.3) I mean I don't know what I did, soo
```

In contrast to the game mechanic-related breakdowns we examined in the prior section, the breakdown faced by leafvr2 proves to be far more severe in nature. A sudden resolution drop occurs after the user performs a throwing motion (line 5), which leads to an expression of confusion and a shift of their gaze downwards towards their hands/controllers. The other users present, leafvr1 and Neiltwo2, presume that leafvr2's "what happened" on line 6 refers to the outcome of her frisbee throw, as opposed to the technical issues which have occurred. While the study participants and Neiltwo2 have taken more of a collaborative approach to their disc golf game thus far, cheering others on and offering advice, they still fail to correctly recognize the trouble faced by leafvr2 due to a lack of obvious external cues besides her own (misunderstood) vocal exclamation. Finding the right words to describe the issue she faces also proves problematic for leafvr2 as seen on line 10, resulting in a brief pause in her speech before settling on 'thing' as a descriptor of what is occurring.

Leafvr2 is then suddenly removed from the Rec Room game entirely (line 14). In this case, the affected user has no practical means of diagnosing the situation themselves without removing the VR devices they are using, never mind having the ability to resolve the problem; ultimately, the observing student members of Group 6 (observers1 and observer2) become involved in the interaction in order to return leafvr2 to the game. It is worth noting that while this event was a major disruption for the participants, it did not mark the end of the recording session; with instruction provided by the observers, leafvr2 managed to restart the game after approximately one minute of troubleshooting which took place after the excerpt, though her avatar was spawned back at the default starting location instead of returning her to the disc golf range. Furthermore, while leafvr2's issues were being resolved, leafvr1 and Neiltwo2 could and did carry on playing disc golf. The result of leafvr2's HMD issues could be characterized as the interaction being suddenly split into two from the original interaction taking place between leafvr1, leafvr2, and Neiltwo2. Leafvr2 shifted her focus to interacting with the observers outside the virtual environment, who took on a role akin to an instructor in helping leafvr2 troubleshoot and resolve the technical issues she was facing. Meanwhile, leafvr1 and Neiltwo2 kept interacting within VR as leafvr2's troubleshooting process took place.

Not all HMD-related issues within the research material resulted in equally notable disruptions to an ongoing interaction, however. Let us examine an excerpt involving leafvr2 of Group 4 as a point of contrast to the above:

(5) "What did I press?" (G4C1)

```
leafvr2      *rears back his hand for a throw, throws his frisbee*
             ((the resolution of leafvr2's recording suddenly drops))
LEAFVR2:     *what did I press?
             *falls still, looks down at controllers/hands*
             (1.2) it's-
leafvr1      +throws their frisbee, lands in a hazard+
LEAFVR1:     nooo
             ((leafvr1 remains focused on the disc golf))
leafvr2      *keeps looking around, down at the ground, at his hands*
LEAFVR2:     (2.1) *did I do something?
Leafvr2           *looks around, lifts right hand/controller up*
             (5.1)
             *teleports to his frisbee*
             *resumes playing*
```

In what is becoming a recurring theme of HMD-related breakdowns, the one experienced by leafvr2 occurs completely unexpectedly and leaves the user in a state of confusion. Leafvr2's immediate response is to stop what he's doing, look down at his hands (or the controllers he is holding), and assume that the sudden change in his vision has been caused by something he himself had done unintentionally (line 3). He appears to grow less confident of the cause being an action committed by himself on line 10, which is the last time he makes a verbal reference to something being amiss. Another inspection of his controllers follows (line 11), after which the trouble appears to pass and leafvr2 returns to playing out the round of disc golf. The initial occurrence of the problem faced by leafvr2 is very similar to the one faced by the users in the prior extract, with the key difference being that it did not escalate in severity to the point of removing leafvr2 from his Rec Room session.

The degree of disruption ultimately caused by the breakdown after it occurs around lines 2-3 is ultimately more limited than in the other excerpts we have examined thus far, particularly in comparison to the one featured in excerpt 4. The overall length of the sequence from beginning to end is no longer than 20 seconds, and the breakdown only really affects the experience of leafvr2 while having little impact on that of leafvr1 (who keeps on playing the game with little mind paid to leafvr2's troubles). This is in part down to the game mode the two players are engaged in; while 3D Charades revolves around constant interaction between the explainer and guesser(s) throughout the course of the round in an attempt to reach a common goal, in disc golf each player is primarily responsible for their own performance alone, thus encouraging players to prioritize their focus on their own performance.

While a clear breakdown occurs within the extract, a successful breakthrough does not. As leafvr1 is left unsure as to the cause of the trouble he is facing, he also cannot come up with a potential solution for it. Thus, his response is rather passive: inspecting his controllers, voicing his confusion, and simply waiting. While it does ultimately prove to be a successful strategy for overcoming the trouble in this instance, it is very difficult to ascertain *why* it worked.

### *5.2.2. Breakdowns related to VR input devices*

Beyond the HMDs, the data also held several instances of breakdowns caused by issues relating to the Vive motion controllers. Let us next examine one such instance involving the participants of Group 6, which took place several minutes before the events of excerpt 4. The participants have been joined by another player, Neiltwo2, while playing disc golf and have just reached a stage's final section, a small island in the middle of a water hazard. After all the players reach the island, leafvr1 begins experiencing issues when interacting with his own frisbee.

(6) “I’m flashing.” (G6C1)

```
leafvr1      *tries to teleport where leafvr2 is standing*
LEAFVR1:     uh (.) can you move a little bit? you’re on my spot
LEAFVR2:     which one? (.) me?
LEAFVR1:     yeah +(2.1)        + thanks
leafvr2           +teleports away+
leafvr1      *moves to throw his frisbee, it disappears*
LEAFVR1:     (.) wait *where’s my-            *
                      *looks around, behind him*
             lost my frisbee *(1.1)            there* it is
                             *pulls the frisbee back*
             *(1.2) it w(h)on’t stop gliding around*
             *frisbee drifts out of his grip       *
LEAFVR2:     ha ha ha (.) come baaaack
             it’s floating- *oh yeah                       *
leafvr1                     *frisbee returns to user’s hand*
LEAFVR1:                     now- now it’s on my hand
             (1.4)
             *throws frisbee, hits Neiltwo2*
             [ha ha ha]
LEAFVR2:     [ha ha ha]
LEAFVR1:     I just- I just realized you can throw into other players
             heh heh
NEILTWO2:    [ha ha ha]
LEAFVR1:     [heh heh heh]
NEILTWO2:    I’m sorry
leafvr1      *reaches for his frisbee, pauses, inspects hands*
LEAFVR1:     I’m (.) flashing some- for some reason
             I don’t know [if I’m losing power]
LEAFVR2:                  [are you being hazar]dous?
NEILTWO2:    [some- ]
LEAFVR2:     [I mean] [I was]
NEILTWO2:             [sometimes] your hands go retarded
             (1.9)
LEAFVR2:     you’re losing your hands I can’t see them
LEAFVR1:     *yeah, I-                               *
             *retrieves the frisbee with his left hand*
LEAFVR2:     [so]
LEAFVR1:     [I k]now *(.) oh        *
                      *throws frisbee*
NEILTWO2:    aw
LEAFVR2:     you tried so hard
             *#whoa ((leafvr1’s hand notably offset from his avatar))
leafvr1      *retrieves frisbee, tries to score, hands flicker-->*6.48
fig           #fig7
LEAFVR1:     I just- (.) put it- (.) put it into the- (.) where’s-
LEAFVR2:     just- *[yeah*, drop it
LEAFVR1:           *[oh*, there-]
                -->*   *drops the frisbee into basket, scores a point*
             (.) ugh
LEAFVR1:     yaaay
```

While leafvr1 begins encountering difficulties with his controller on line 6 (the frisbee moving erratically while in his grip likely being an in-game manifestation of a motion tracking issue), he does not make direct reference to it until line 27 where he describes what is happening as “flashing.” The frisbee keeps acting erratically each time he grasps it with his right hand, with both leafvr2 and Neiltwo2 remarking on the matter on lines 34 and 32 respectively. A particularly notable instance of desynchronization between leafvr1’s avatar and one of his hands occurs on line 42, prompting leafvr2 to react with a “whoa.” The unresponsive controller makes it extremely challenging for leafvr1 to score a point even while standing next to the basket. After repeated attempts and trying to switch from handling his frisbee with his left hand instead of the right, leafvr1 finally manages to finish the stage by scoring a point.

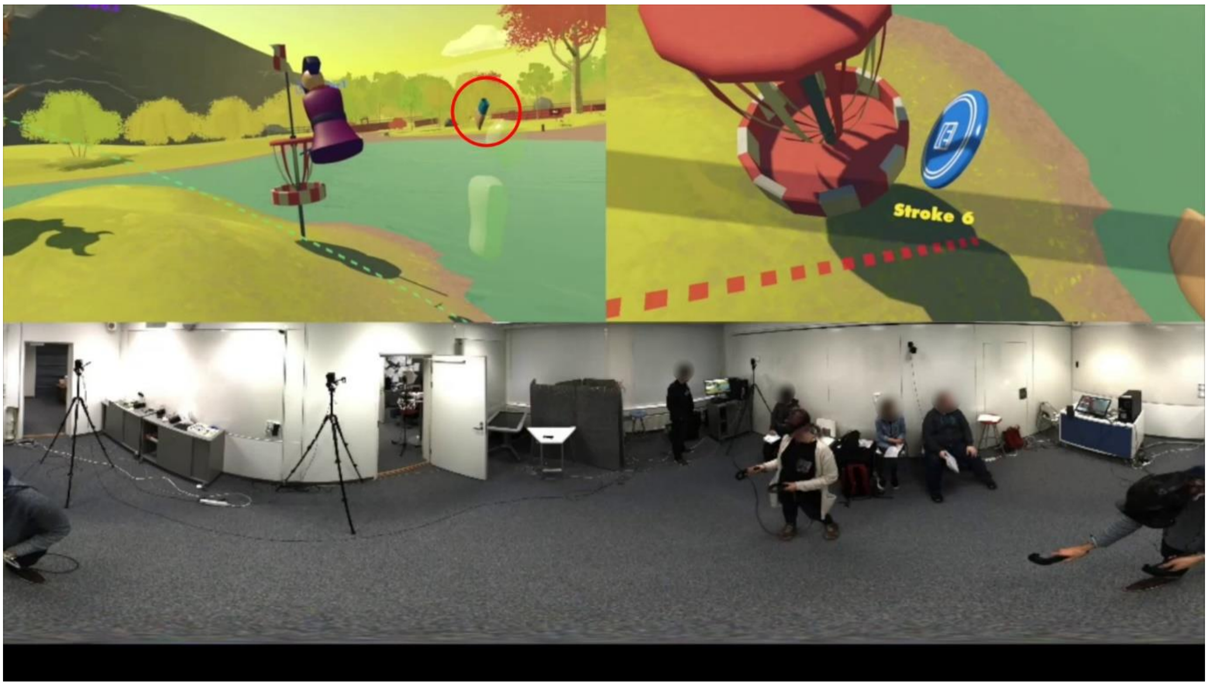


Figure 7. Leafvr1’s hand offset from their avatar, circled in red in leafvr2’s POV.

It is worth noting that the possibility of a technical fault being behind leafvr1’s control difficulties is only brought up once (line 28), with the other remarks referencing only the consequences of the issue at hand (the ‘flashing’, ‘losing’ one’s hands, etc.). Rather than try and ascertain the cause, leafvr1 instead opts to power through the issues in order the wrap up the disc golf course despite the challenges posed by the erratic behaviour of his hands in-game. This may relate to the difficulties the users faced in describing the trouble, using vague terms such as ‘flashing’ or the frisbee ‘gliding’ (line 11) to try and communicate the nature of the issue.

In the last two excerpts, we examine two sequences featuring the participants of Group 5. Leafvr1 has been suffering from minor hiccups with his right-hand controller for some time by this point in the session, and it is over the course of this first sequence where the issues grow more disruptive. This occurs as the two participants are playing a game of Rec Room's iteration of paddleball. As leafvr1 is attempting to swing his paddle, the motion fails to translate accurately into VR, triggering the following interaction:

(7) "Oh! My hand died." (G5C1)

```
LEAFVR1:    oh! oh! *(2.4)                                    *
                    *inspects his right hand, waves it around*
            my hand died (8.7) it's-
LEAFVR2:    (1.2) .hhh your controller's, like, floating o(h)ff
LEAFVR1:    yeah I dunno
LEAFVR2:    there it's back
LEAFVR1:    alright *(7.1)     * yeah my controller's not working
                    *waves hand*
leafvr2     +tries to teleport to leafvr1's side of the field+
LEAFVR2:    (2.2) can't teleport so far
LEAFVR1:    it's dying (2.9) could be-
            mm yeah (2.2) it catches and then it-
            it slowly drifts away (3.5) looks like a battery issue
            (3.4)
OBSERVER:   you ran out of battery?
LEAFVR1:    I think so *(.) there's one                  *
                       *holds a controller out for observer*
OBSERVER:   press the- try again
            it should come back after this (.) might have to recharge it
LEAFVR2:    .hhhh
LEAFVR1:    okay
OBSERVER:   is it-
LEAFVR1:    I have- yeah I have the-
            alright I think-
OBSERVER:   it has some power?
LEAFVR1:    yeah
OBSERVER:   okay
LEAFVR1:    alright, that's it, *it's- no, no                    *
                                *moves his right controller again*
            (.) it doesn't respond very fluently (.) it sticks
```

In this sequence, the user affected by the breakdown (leafvr1) is already well aware of the fact that something is wrong with his controller; it is the escalating severity of the issue which leads to a disruption in the ongoing interaction, and his focus turning to his controllers from the ongoing game of paddleball (lines 1-2). While his right-hand controller is not quite "dead" to the point of being completely non-functional, it is clearly experiencing rather severe tracking issues. These issues also become obvious to leafvr2, who can observe the paddleball racket in leafvr1's right hand desynchronizing from their avatar in a manner similar to what can be seen on figure 7. Witnessing this event also prompts leafvr2 to comment on it (line 4). After briefly waving his controllers around, leafvr1 declares that his controller is not working (line 7), followed by him describing the problem aloud over the course of the next few lines. He knows there's trouble but does not know how to fix it himself. Instead, he describes the symptoms caused by this problem aloud (lines 11-14). These explanations by leafvr1 feature numerous pauses in excess of two seconds in duration, during which the user is either focused on experimenting with the troublesome controller, or appears to be trying to come up with a suitable descriptor for the issues he is facing (e.g. 'it catches' on line 14 following a long pause, 'it sticks' on line 30 following a shorter pause). Leafvr2, meanwhile, has little further input to add, beyond remarking on his inability to teleport to leafvr1's half of the paddleball field.

This prompts one of the observers to involve themselves with the troubleshooting process, instructing leafvr1 to perform certain actions (possibly attempting to resynchronize the controller's tracking with the HMD and base stations) in an effort to try and remedy the problem (lines 15-24). As such, leafvr1's focus turns from his original interaction with leafvr2 to listening to the observer, the other participant falling to the wayside in the conversation for the remainder of the extract. Even with the remedial actions taken as per the instructions of the recording session observer, the problem ultimately persists (lines 27-29). The user carries on playing following the ultimately unsuccessful attempt at remedying the issue, while paying close attention to his controllers for the remainder of the recording session.

While the participants keep on playing after the end of the extract, eventually switching from paddleball to Shield Soccer, the controller issues ultimately lead to a premature end for Group 5's recording session:

(8) "We'll have to stop, then." (G5C2)

```
leafvr1      *attempts to hit the ball*
LEAFVR1:     oh (.) my controller's not responding
             *waves his hands around*
```

(6.8)
yeah both of my controllers are dying now
OBSERVER1: okay
OBSERVER2: we'll have to stop then
LEAFVR1: yeah I think so
(.4)
OBSERVER1: sorry about that
LEAFVR1: oh it happens
observer1 +moves to take controllers from leafvr1+

While the issues posed by the dying controllers in this instance were not as dramatic in their onset as the HMD issue of extract 5, the amount of disruption they caused to the experience of Group 5's participants was much greater: The trouble lingered for an extended period of time, and ultimately led to a premature end for the session courtesy of leafvr1's controller ultimately 'dying' completely during the Shield Soccer match. The controller woes faced by leafvr1 appear to lead to a lack in his sense of involvement with the game, as is typical of breakdowns which involve a loss of player agency (Iacovides et al., 2015). Instead of engaging with Rec Room's game modes or interacting with other players, he spends increasingly more time attempting to troubleshoot the issues he is facing, to the point of turning his interaction from the other participant he's engaging within VR to the observers present in the real world. While he does not remove the HMD until after the end of excerpt 8, it feels reasonable to presume that his sense of presence has been markedly diminished by the time his controllers are finally rendered completely unresponsive.

*5.2.3. Discussion*

The difficulties users experience in identifying the source and nature of hardware-related breakdowns is a recurring occurrence in the data, as are challenges related to communicating information about the breakdowns to other users. Even when the faulty component is recognized, such as the dying controllers in the case of Group 5's leafvr1, the exact cause is left unclear; while the user presumes the issue to be battery-related and the observers of the recording session appear to agree with the assessment, the user has no means to accurately diagnose the issue. An UI element alerting the user to controller-related tracking issues or a low battery charge may have gone a long way in avoiding the confusion felt by leafvr1 as he attempted to ascertain the source behind his controller troubles. Contrast this with personal computers running Windows operating systems which default to playing audio cues when input devices (keyboards, mice, etc.) are connected or disconnected, often also triggering a UI prompt to alert the user that something about the system configuration has changed, or network error icons displayed in many contemporary online multiplayer games when a user's internet connection experiences issues. Even if a given user might not understand the meaning of such a prompt, describing it to someone who does could expediate the troubleshooting process. Of course, UI prompts could be highly disruptive to the user's sense of presence in a VR environment if implemented carelessly.

On the subject of controllers, inspecting them in response to breakdowns which do not relate to the game mechanics or environment proved to be another recurring reaction demonstrated by the users. Doing so was among the initial responses users had in every excerpt of this section along with vocally indicating that something appeared to be amiss. This focus on the devices the users are using as opposed to the software they are interacting with is an example of a ready-at-hand (RaH) to present-at-hand (PaH) transition in the user's relationship to the device; instead of the VR system's components being extensions of the user as they perform various activities, the devices are brought to the forefront (Ryan & Sigel, 2009). This is particularly observable in extracts 7 and 8, where leafvr1 repeatedly draws attention to the Vive controllers and the issues they are causing. The controller-related issues appear to trigger breakdowns in action by the definitions of Iacovides et al. (2011), with unresponsive controls being a constant hindrance to affected users' actions within the game. Contrast this with the breakdowns examined in section 5.1., where no obvious RaH-to-PaH transitions could be observed.

But while Group 5's leafvr1 was able to draw conclusions regarding the hardware issue he was facing, the issues relating to the function of the HMDs themselves as experienced by the members of Group 4 and Group 6 proved more elusive when it came to determining their causes. While the individuals affected by the issues could tell that *something* was happening, they did not know *why* it was happening, and what they could possibly do to resolve it. As a result, the users' responses to the issues were uniformly passive in nature: they ceased their present course of action, indicated that something was wrong verbally (often in the form of being confused), and proceeded to stand still on the spot. In one case they were able to return to their paused activity after a rather brief disruption (excerpt 4), whereas in the other the visual glitches were followed by a more severe failure which removed the user out of the game entirely (excerpt 5). Leafvr1 of Group 6 also did not connect the “flashing” issues he faced to his controllers, despite the difficulties he experienced being similar to leafvr1 of Group 5 (unresponsive controllers, hands appearing offset from their avatar in others’ POV). Instead, he proved able to power through the handicap caused by his controller troubles. It is also worth noting that, whereas the users experiencing the device-related breakdowns faced challenges determining the cause, for the other users in the session this was even more difficult. Thus, we often saw other users present remarking on the obvious in-game symptoms of the technical issues should they appear.

Some of this lacklustre success in troubleshooting technical issues is likely related to the nature of the HMDs themselves. Consider a display issue with a traditional desktop computer set-up, with a monitor connected to a central unit via a cable: Should issues with the display arise, the user has immediate access to the cable connection and the monitor’s own power switch and other control buttons to begin their troubleshooting with in most use-cases. With a HMD like the Vive, however, a user wishing to inspect the hardware itself in case of trouble arising must go through the process of first removing the headset, which in itself is also a less familiar piece of technology to the average individual than the ever-ubiquitous desktop computer monitor. Similarly, other individuals present face limitations in helping a user affected by technical difficulties; those in the real world may not be able to see what the user is experiencing within VR if no secondary monitor is available, and other VR users can only see the possible in-game effects of the issues at hand (such as desynchronized controllers). This, in turn, may serve to emphasize instances of self-initiated attempts at resolving device-related problems over other-initiated ones.

Ultimately, while the users facing hardware-related breakdowns in the examined sequences were the ones who initiated repair attempts, they were never able to reach truly successful resolutions for the troubles they faced without the aid of others; they either waited for the issue to seemingly pass before continuing on with their prior course of action or carried on through the trouble (circumventing or ignoring it), or required help from other individuals present in the recording sessions. Notably, in hardware-related breakdowns, the affected users had to turn to individuals *outside* VR to ask for help in resolving the trouble, in contrast to excerpts 1-3 where the relevant interactions remained *within* VR. In the most severe cases, the device-related problems resulted in a major disruption to the interaction and the users' gameplay experience (excerpt 4) or forced them to quit their play session entirely (excerpt 8).

The next section will see the conclusion of this thesis. We will sum up the key points of our findings and briefly discuss future avenues of research.

## 6. Conclusion

At the beginning of this thesis, we set out to explore the impact of breakdown events on co-present interaction between users within virtual reality. To this end, a total of eight sequences – three featuring breakdowns relating to users' understandings of the game's internal rules, five relating to issues with the VR devices being utilised – were analysed in depth, with a focus on how the users responded to the troubles they faced and how they attempted to resolve them. These sequences were based on video materials recorded of 12 VR users interacting within the Rec Room virtual reality game, which the users played in six pairs for approximately one hour each in the presence of multiple observers. Footage was recorded from each user's individual HMD alongside a roof-mounted camera which offered a view of the recording environment, after which the feeds were combined into single synchronized video files. Conversation analysis was used as the basis for analysing this footage in conjunction with past research and theory regarding the phenomenon of breakdowns in human-computer interaction and repair in interaction.

Of the two broad types of breakdowns, game-related ones proved to be less disruptive to the overall user experience, with users being able to resolve the issues without requesting help from outside the game; instead, users collaborated when it became obvious that an issue had arisen, offering advice and suggesting potential solutions. Based on the sequences analysed, it did not appear that the game-related breakdowns experienced by the users had a notable negative effect on their level of engagement with the game, with their focus remaining within VR throughout the interactions. Device-related breakdowns proved more disruptive to the user experience than game-related ones, albeit with a great degree of variance; some breakdowns resulted in only a brief pause, while those of greater severity led to users being removed from the Rec Room game temporarily or forcing the entire recording session to come to a premature end. The problems faced were also focused on individual users and proved challenging to both diagnose and successfully resolve or circumvent. A lack of information, be it of the game's own internal rules or technical difficulties affecting the VR devices, was a recurring theme across all of the breakdowns examined. Repeating actions preceding the trouble and adopting a trial-and-error troubleshooting approach were common responses from affected users across both breakdown types.

Responding to breakdowns did not always fall to the affected user alone. While the users who encountered the problems tended to try and resolve them by themselves initially, other individuals, both within VR and in the real world, often became involved with the troubleshooting process. With game-related breakdowns, the interaction remained within VR, with the users present communicating amongst themselves in order to try and resolve the issue. With device-related breakdowns, we saw the users experiencing the trouble interacted with both their fellow VR users and with individuals in the real world in an effort to troubleshoot the issues; the other VR users were incapable of offering much help in diagnosing the root of the problem but did attempt to offer advice in regards to dealing with symptoms caused by it, such as unresponsive controls, while the actual diagnosing of the issue often fell to the affected user themselves and the observers present in the recording sessions. In some cases, however, other users present failed to notice that their fellow users were facing trouble, typically courtesy of being distracted by some form of in-game interaction. This occurred with both game-related and device-related breakdowns. Finding the correct words to describe the trouble being faced to other users also proved challenging numerous times in the analysed sequences, with users often having to rely on vague descriptions.

While the analysed video material was of good quality and available in abundance, supplemental information gathered before and after the recording sessions could have offered further insight into the users' experience of breakdowns. Information regarding the study participants' general familiarity with video games outside of the context of VR and post-session interviews with the participants on key events which occurred during their recording sessions would have offered more insight into their own decision-making progress while facing the issues in-game. With only the recorded footage to go by, it was challenging to try and accurately evaluate such aspects during the sequences (such as the breakdowns' impact on the participants' sense of presence) which were examined. Similar research carried out in the future may benefit from considering these aspects during the data collection process. Furthermore, gathering data from a broader selection of software could make for an interesting comparison on how different design decisions made by developers affect the occurrence and resolution of breakdowns in VR.

As technology advances and the desire – and need – for people the interact with one another remotely in ever more intricate and immersive ways, the relevance of VR in entertainment, educational, and professional fields is likely grow over the course of the coming decade. While users will inevitably encounter breakdowns when using any technology, especially one as new as contemporary VR, minimising the issues they face and their disruption to whatever a user wishes to do is important to maintaining a good user experience, and in the case of VR, a strong sense of presence. Through studying the troubles users face in co-present virtual environments, designers can identify and remedy potential sources of breakdowns in hardware and software alike, in turn enabling the creation of VR experiences which are more seamless, user-friendly, and approachable.

## References

Abrash, M.What VR could, should, and almost certainly will be within two years. Paper presented at the *Steam Dev Days 2014,* Retrieved from https://steamcdn-a.akamaihd.net/apps/abrashblog/Abrash%20Dev%20Days%202014.pdf

Alasuutari, P., Alasuutari, P., Bickman, L., Brannen, J., Alastalo, M., Alexander, V. D., . . . Wong, V. C. (2008). *The SAGE handbook of social research methods*. Thousand Oaks (Calif.) ; London: SAGE Publications. Retrieved from https://oula.finna.fi/Record/oula.973964

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. doi:10.1080/13691180801937290

Arminen, I. (2013). Environmentally coupled repairs and remedies in the airline cockpit: Repair practices of talk and action in interaction. *Discourse Studies, 15*(1), 19-41. doi:10.1177/1461445612466463

Bailenson, J. N. (2006). A longitudinal study of task performance, head movements, subjective report, simulator sickness, and transformed social interaction in collaborative virtual environments. *PRESENCE: Teleoperators and Virtual Environments, 15*(6), 699-716. doi:10.1162/pres.15.6.699

Barr, P. (2007). *Video game values: Play as human computer interaction* Retrieved from https://pdfs.semanticscholar.org/d6f1/ca3e1b0b55c63c09d3db6d49157bcc838505.pdf

Barricelli, B. R. (2016). Semiotics of virtual reality as a communication process. *Behaviour & Information Technology, 35*(11), 879-896. doi:10.1080/0144929X.2016.1212092

Bates-Brkljac, N. (2012). *Virtual reality*. New York: Nova Science Publishers, Inc. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=540105&site=ehost-live

Benjamin, T., & Mazeland, H. (2012). *Conversation analysis and other-initiated repair* doi:10.1002/9781405198431.wbeal1310

Berger, M. (2016). Interaction and space in the virtual world of second life. *Journal of Pragmatics, 101*(C), 83-100. doi:10.1016/j.pragma.2016.05.009

Bergström, I. (2016). First-person perspective virtual body posture influences stress: A virtual reality body ownership study. *PloS One, 11*(2), e0148060.

Bolden, G. B. (2018). Speaking 'out of turn': Epistemics in action in other-initiated repair. *Discourse Studies, 20*(1), 142-162. doi:10.1177/1461445617734346

Bosch-Sijtsema, P. M., & Haapamäki, J. (2014). *Perceived enablers of 3D virtual environments for virtual team learning and innovation* doi:https://doi-org.pc124152.oulu.fi:9443/10.1016/j.chb.2014.04.035

Bowman, D. A. (2007). Virtual reality: How much immersion is enough? *Computer, 40*(7), 36-43. doi:10.1109/MC.2007.257

Bowman, D., Kruijff, E., LaViola, J., & Poupyrev, I. (2004). *3D user interfaces: Theory and practice* (1st ed.) Addison-Wesley Professional. Retrieved from https://proquestcombo.safaribooksonline.com/9780133390599

Chen, W. (2014). Perceptual conflicts in a multi-stereoscopic immersive virtual environment: Case study on face-to-face interaction through an avatar. *PRESENCE: Teleoperators and Virtual Environments, 23*(4), 410-429. doi:10.1162/PRES_a_00209

Drew, P. (2005). Conversation analysis. In K. L. Fitch, & R. E. Sanders (Eds.), *Handbook of language and social interaction* (pp. 71-101). Mahwah, NJ: Lawrence Erlbaum Associates.

Fox, B. (2012). *Conversation analysis and self repair* doi:10.1002/9781405198431.wbeal0214

Fox, B., Benjamin, T., & Mazeland, H. (2012). *Conversation analysis and repair organization: Overview* doi:10.1002/9781405198431.wbeal1314

Fox, J., Arena, D., & Bailenson, J. N. (2009). *Virtual reality: A survival guide for the social scientist* Hogrefe Publishing. doi:10.1027/1864-1105.21.3.95

Frohlich, D., Drew, P., & Monk, A. (1994). Management of repair in human-computer interaction. *Human-Computer Interaction: Exploratory Sequential Data Analysis, 9*(3-4), 385-425. doi:10.1080/07370024.1994.9667211

Geszten, D., Hámornik, B. P., Komlodi, A., Hercegfi, K., Szabó, B., & Young, A. (2015). Qualitative analysis of user experience in a 3D virtual environment. *Proceedings of the Association for Information Science and Technology,* (1), 1-4. Retrieved from https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/pra2.2015.1450520100124

Greiffenhagen, C., & Watson, R. (2009). Visual repairables: Analysing the work of repair in human-computer interaction. doi:10.1177/1470357208099148

Hepburn, A., & Bolden, G. B. (2013). The conversation analytic approach to transcription. In Tanya Stivers, & Jack Sidnell (Eds.), *The handbook of conversation analysis* (pp. 57-76). GB: Wiley-Blackwell. Retrieved from https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=1000686

Iachini, T., Coello, Y., Frassinetti, F., & Ruggiero, G. (2014). Body space in social interactions: A comparison of reaching and comfort distance in immersive virtual reality. *PLoS ONE, 9*(11), 1-7. doi:10.1371/journal.pone.0111511

Iacovides, I., Cox, A. L., McAndrew, P., Aczel, J., & Scanlon, E. (2015). Game-play breakdowns and breakthroughs: Exploring the relationship between action, understanding, and involvement. *Human-Computer Interaction: Special Issue on HCI and Games, 30*(3-4), 202-231. doi:10.1080/07370024.2014.987347

Iacovides, I., Woods, W., & Scanlon, E.What can breakdowns and breakthroughs tell us about learning and involvement experienced during game-play? Paper presented at the *, 2011*- 275-281. Retrieved from https://oula.finna.fi/PrimoRecord/pci.scopus2-s2.0-84900553655

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13-31). Amsterdamn: John Benjamins Publishing Co. Retrieved from http://gateway.proquest.com/openurl?ctx_ver=Z39.88-2003&xri:pqil:res_ver=0.2&res_id=xri:ilcs-us&rft_id=xri:ilcs:rec:abell:R03478781

Kapralos, B., Collins, K., & Uribe-Quevedo, A. (2017). The senses and virtual environments. *The Senses and Society, 12*(1), 69-75. doi:10.1080/17458927.2017.1268822

Kitzinger, C. (2012). Repair. In J. Sidnell, & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 229-256). GB: Wiley-Blackwell.

Levin, L., Cromdal, J., Broth, M., Gazin, A., Haddington, P., McIlvenny, P., . . . Rauniomaa, M. (2017). Unpacking corrections in mobile instruction: Error-occasioned learning opportunities in driving, cycling and aviation training. *Linguistics and Education, 38*, 11-23. Retrieved from https://doi.org/10.1016/j.linged.2016.10.002

Lin, J. J., Duh, H. B. L., Parker, D. E., Abi-Rached, H., & Furness, T. A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. *Proceedings IEEE Virtual Reality 2002,* , 164-171. doi:10.1109/VR.2002.996519

Marini, D., Folgieri, R., Gadia, D., & Rizzi, A. (2012). Virtual reality as a communication process. *Virtual Reality, 16*(3), 233-241. doi:10.1007/s10055-011-0200-3

Marsh, T. (2001). Evaluation for the design of experience in virtual environments: Modeling breakdown of interaction and illusion. *Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society, 4*(2), 225. doi:10.1089/109493101300117910

Mondada, L. (2013). The conversation analytic approach to data collection. In Tanya Stivers, & Jack Sidnell (Eds.), *The handbook of conversation analysis* (pp. 32-56). GB: Wiley-Blackwell. Retrieved from https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=1000686

Mondada, L. (2016). Conventions for multimodal transcription. Retrieved from https://franzoesistik.philhist.unibas.ch/fileadmin/user_upload/franzoesistik/mondada_multimodal_conventions.pdf

Mortensen, K. (2012). *Conversation analysis and multimodality* doi:10.1002/9781405198431.wbeal0212

Muhanna, A. (2015). Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions. *Journal of King Saud University - Computer and Information Sciences, 27*, 344-361.

Murray, C. D., & Gordon, M. S. (2001). Changes in bodily awareness induced by immersive virtual reality. *CyberPsychology & Behavior, 4*(3), 365-371. doi:10.1089/109493101300210268

Nevile, M. (2007). Action in time: Ensuring timeliness for collaborative work in the airline cockpit. *Language in Society, 36*(2), 233-257. doi:10.1017/S0047404507070121

Nicolajsen, H. W. (2019). IT-enabled idea competitions for organizational innovation: An inquiry into breakdowns in adaptation. *Creativity and Innovation Management, 28*(4), 436-449. doi:10.1111/caim.12330

Norris, S. (2004). *Analyzing multimodal interaction : A methodological framework*. London: Routledge. Retrieved from https://oula.finna.fi/Record/nelli28.1000000000247774

Pan, X., & Hamilton, Antonia F. de C. (2018). *Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape* John Wiley & Sons, Ltd. doi:10.1111/bjop.12290

Pelletier, C., & Oliver, M. (2006). Learning to play in digital games. *Learning,* Retrieved from https://oula.finna.fi/PrimoRecord/pci.ucl10003112

Raudaskoski, P. (1990). Chapter 7 - repair work in human-computer interaction: A conversation analytic perspective. () doi:10.1016/B978-0-08-050264-9.50012-9 Retrieved from https://oula.finna.fi/PrimoRecord/pci.elsevier_sdoi_10_1016_B978_0_08_050264_9_50012_9

Ryan, W., & Sigel, M. A.Evaluating interactive entertainment using breakdown: Understanding embodied learning in video games. Paper presented at the *DiGRA International Conference,* 1-9.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696-735. doi:10.2307/412243

Schegloff, E. A. (1992). Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology, 97*(5), 1295-1345. doi:10.1086/229903

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language, 53*(2), 361. doi:10.2307/413107

Schulze, S., Pence, T., Irvine, N., & Guinn, C. (2019). (2019). The effects of embodiment in virtual reality on implicit gender bias. Paper presented at the *Virtual, Augmented and Mixed Reality. Multimodal,* 361-374.

Sharples, M. (1993). A study of breakdowns and repairs in a computer-mediated communication system. *Interacting with Computers, 5*(1), 61-77. doi:10.1016/0953-5438(93)90025-O

Sharples, M. (2009). Methods for evaluating mobile learning. *Researching Mobile Learning: Frameworks, Tools, and Research Designs,* , 17-39. Retrieved from https://www.researchgate.net/publication/252110088_Methods_for_Evaluating_Mobile_Learning

Sidnell, J. (2013). Basic conversation analytical methods. In Tanya Stivers, & Jack Sidnell (Eds.), *The handbook of conversation analysis* (pp. 77-100). GB: Wiley-Blackwell. Retrieved from https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=1000686

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1535), 3549-3557. doi:10.1098/rstb.2009.0138

Slater, M. (2018). *Immersion and the illusion of presence in virtual reality* doi:10.1111/bjop.12305

Slater, M., Spanlang, B., Sanchez-Vives, M., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS ONE, 5*(5), 1-9. doi:10.1371/journal.pone.0010564

Spagnolli, A. (2002). Situated breakdown analysis for the evaluation of a virtual environment. *PsychNology Journal, 1*(1), 5-17. Retrieved from https://oula.finna.fi/PrimoRecord/pci.scopus2-s2.0-84883430137

Tham, J. (2018). *Understanding virtual reality: Presence, embodiment, and professional practice* doi:10.1109/TPC.2018.2804238

Thomas, L. M., & Glowacki, D. R. (2018). Seeing and feeling in VR: Bodily perception in the gaps between layered realities. *International Journal of Performance Arts and Digital Media, 14*(2), 145-168. doi:10.1080/14794713.2018.1499387

Vlachos, A. (2015). Advanced VR rendering. Retrieved from https://steamcdn-a.akamaihd.net/apps/valve/2015/Alex_Vlachos_Advanced_VR_Rendering_GDC2015.pdf

Wagner, J. (2012). *Conversation analysis and transcription and data* doi:10.1002/9781405198431.wbeal0215

**Appendices**

Symbols for transcription (Jefferson, 2004)

| | |
|---|---|
| [word] | Brackets: Overlapping talk. |
| = | Equal signs: No break or gap. |
| (1.2) | Numbers in parentheses: Elapsed time by tenths of seconds |
| (.) | A dot in parentheses: A brief interval of approx. one tenth of a second |
| word | Underscoring: Some form of stress, either in pitch or amplitude. Longer underscore indicates greater stress |
| :: | Colons: Prolongation of the immediately prior sound. Longer colon row indicates longer prolongation. |
| :_ | Colon-underscore combinations: Intonation contours. The underscore 'punches up' the sound it occurs beneath. |
| ↑↓ | Arrows: Indicate a shift into especially high or low pitch. |
| .,?? | Punctuation markers: Used to indicate 'the usual' intonation. |
| WORD | Upper case: Indicates especially loud sounds in contrast to the surrounding talk. |
| ◦word◦ | Degree signs: Sounds bracketed by degree signs are softer than the surrounding talk. |
| * | Asterisk: Percussive non-speech sounds. |
| - | Dash: Indicates a cut-off. |
| .hhh | Dot-prefixed row of 'h's: Indicates an in-breath. A row without a dot prefix indicates an outbreath. |
| whhord | A row of 'h's within a word: Indicates breathiness |
| ( ) | Empty parentheses: Unclear talk |
| (word) | Words/speaker designations within parentheses: Especially dubious words/speakers |
| (( )) | Double parentheses: Transcriber's descriptions |

Conventions for multimodal transcription (Mondada, 2001), version 3.0.6

| | |
|---|---|
| * * | Gestures and descriptions of embodied actions are delimited between |
| + + | two identical symbols (one symbol per participant) |
| Δ Δ | and are synchronized with corresponding stretches of talk. |
| | |
| *---> | The action described continues across subsequent lines |
| ---->* | until the same symbol is reached. |
| | |
| >> | The action described begins before the excerpt's beginning. |
| --->> | The action described continues after the excerpt's end. |
| ..... | Action's preparation. |
| ---- | Action's apex is reached and maintained. |
| ,,,,, | Action's retraction. |
| ric | Participant doing the embodied action is identified when (s)he is not the speaker. |
| | |
| fig | The exact moment at which a screen shot has been taken |
| # | is indicated with a specific symbol showing its position within the turn at talk. |