

Staging tele-presence by embodying avatars: evidence from Let's Play Videos

Axel Schmidt & Konstanze Marx

Abstract

In so called *Let's Plays*, video gaming is presented and verbally commented by *Let's Players* on the internet for an audience. When only watched but not played, the most attractive features of video games, immersion and interactivity, get lost – at least for the internet audience. We assume that the accompanying reactions (transmitted via a so-called facecam) and verbal comments of *Let's Players* on their game for an audience contribute to an embodiment of their avatars which makes watching a video game more attractive. Following an ethnomethodological conversation analytical (EMCA) approach, our paper focusses on two practices of embodying avatars. A first practice is that *Let's Players* verbally formulate their actions in the game. By that, they make their experiences and the 'actions' of avatars more transparent. Secondly, they produce response cries (Goffman) in reaction to game events. By that, they enhance the liveliness of their avatars. Both practices contribute to a co-construction of a specific kind of (tele-)presence.

Keywords: Let's Play, computer game, conversation analyses, tele-presence, embodiment

Corresponding authors:

Prof. Dr. Axel Schmidt

[axel.schmidt@ids-](mailto:axel.schmidt@ids-mannheim.de)

mannheim.de

Institut für Deutsche Sprache

(IDS)

R 5, 6–13

68161 Mannheim (Germany)

Prof. Dr. Konstanze Marx

[konstanze.marx@uni-](mailto:konstanze.marx@uni-greifswald.de)

greifswald.de

Institut für deutsche

Philologie

Rubenowstraße 3, R. 2.04

17487 Greifswald (Germany)

1 Introduction

The research objects of our contribution are so-called *Let's Plays*, in which video games are played for an internet audience (cf. Hale 2013; Ackermann 2016a). To make watching a video game attractive for viewers, the gamer(s) usually produce verbal as well as embodied comments on their game. In addition, their face often appears in an extra video embedded within the feed of the video game (a so-called facecam) (cf. section 2, fig. 1). This allows spectators access to the verbal and embodied reactions of the players during the game.

Essentially, gamers take on the additional role of a moderator, not only playing but also mediating their game play activities for an audience. By doing this they try to make their game play 'watchable' (cf. Schmidt/Marx 2020). One crucial aspect that enhances the pleasure of watching a video game is to make moves in the game more transparent and understandable for viewers. For this purpose, gamers verbally formulate their game moves (e.g. *now I knock at a door*). In addition, they produce exclamations or following Goffman (1981b) response cries during or after their actions like *oh* if they are surprised or *ahhh* if they are shocked or frightened by game events. In this way they animate their avatars. We understand such practices of either formulating actions or animating avatars as an embodiment of avatars by gamers for viewers.

By doing so, gamers construct a specific kind of participation framework (cf. Goffman 1981a; Goodwin/Goodwin 2004) consisting of a) a human-machine-interaction (playing the game) which results b) in represented virtual activities (avatars 'doing' something within the game) which are c) presented for viewers. By animating their avatars and making their game moves more understandable by verbally formulating their actions, gamers stage (tele-)presence as they allow the spectators to participate in their immediate experience.

Our contribution focuses on the types of practices gamers employ to embody their avatars in video games for an internet audience. We investigate two main practices of *Let's Players*, *formulating actions* and *animating avatars via response cries*, which interact in the embodiment of avatars.

Our contribution follows an interactional perspective, or more precisely, a multimodally extended EMCA approach, which asks how participants create social reality. EMCA stands for Ethnomethodology and Conversation Analysis (cf. Heritage 1984a). By multimodal extension, we mean that the focus is not only on talk, but also on embodied actions, including the use of objects, media technology and space (cf. Deppermann 2013; Mondada 2008; Streeck/Goodwin/LeBaron 2011). Specifically, our analysis focuses

on the strategies or practices that players employ to make their *Let's Plays* attractive and engaging for potential viewers (and therefore “watchable”).

In section two and three we outline the specific participation framework of *Let's Plays*. After a brief introduction of our data base in section four, we analyze practices of formulating one's own actions (5.1) and of animating avatars via response cries (5.2) in section five.

2 *Let's Plays* and their participation framework

Our paper aims to contribute to a growing body of studies dealing with computer gaming from an interaction-theoretical perspective (cf. summarizing Reeves/Greifenhagen/Laurier 2016). Most studies emphasize that playing a computer game extends interaction to a virtual world and, by that, creates a different and – in the case of several players – a more complex *participation framework*¹ (cf. Keating/Sunakawa 2010; Laurier/Reeves 2014; Mondada 2012; Piirainen-Marsh 2012; Tekin/Reeves 2017) as well as different time layers (at least the time prescribed by the game play and time in terms of interaction) which have to be temporally coordinated (cf. Mondada 2013).

In the simplest case of gaming, which is playing alone, two levels of (inter)action arise: First, there is an interaction between the player and the computer, that is, a kind of human-machine-interaction, whose base is a *control action* by the player (e.g. moving the mouse). Secondly, this creates a *represented (inter)action* within the game (e.g. an avatar's knocking at a door or an avatar's fight with a non-play character). We call this *game action*. Control actions happen in the *real world*, game actions in the *game world*.

Computer games can be played alone or together. Playing together creates additional levels of interaction in the real world as well as in the game world. When several players control different avatars in a joint game, interactions between the players (either because they are co-present or because they are connected via technology) and/or between their avatars on screen may occur (cf.

1 The notion “participation framework” (Goffman 1981a: 137) describes the relations which participants accomplish in and through their communication. Participation frameworks are on the one hand constrained and enabled by situational and technological parameters (for instance, by whether a communication is face-to-face or technically mediated, whether it is written or oral etc.; cf. Meyrowitz 1990). On the other hand, participation frameworks are indexed by the ongoing activities of the participants, and therefore are in constant flux (cf. Goodwin/Goodwin 2004; Arminen/Spagnolli/Licoppe 2016).

Mondada 2012; Baldauf-Quilliatre/Colón de Carvajal 2019; Marx/Schmidt 2019). However, in the following we focus mainly on single players who present their gaming activities to an internet audience.

In *Let's Plays*, video games are not only played but presented to an internet audience. Following Dynel (2014), this creates a communication with videos (in this case gaming videos) on a sender-recipient-level (see also Schmidt/Marx 2019 for a more detailed discussion of the participation framework of *Let's Plays* as YouTube videos). In addition, *Let's Players* not only present videos but comment verbally and bodily on them via a facecam while playing the video game. Figure 1 shows the typical representation of *Let's Plays*.



Figure 1: Typical representation of a *Let's Play*

The verbal comments and bodily reactions of the players add an extra level of communication, a kind of para-social interaction (cf. Horton/Wohl 1956) or – as Ayaß (1993) has suggested renaming the term – social para-interaction. This has three effects: First, gamers produce talk and embodied conduct for an audience to which they have no direct access. Similar to mass media products such as television, direct address is frequently used to create an intimate interaction situation, even though there is no immediate feedback from the audience, and so the audience-centric talk is only based on assumptions about typical viewers (cf. Ayaß 2005; Hausendorf 2001). Secondly, as the *Let's Players* are constantly physically present in close-up via a facecam, viewers can continuously and

closely observe their behavior (cf. Meyrowitz (1985) for a similar argument in relation to television). Finally, viewers of *Let's Plays* are not able to participate, and importantly, they are not able to influence the game. All they can do is watch and listen to the comments of those who are actively involved in playing. Interactivity and immersion are considered to be computer games' most attractive features (cf. Freyermuth 2015). Although the term *Let's Play* promises a joint gaming experience, both of these features are lost for viewers of *Let's Plays*, as joining a *Let's Play* merely means to be in the role of a spectator (and, therefore, 'Watch Me Play' may actually be a more appropriate name, as suggested by Ackermann (2016a)). As our examples will show, this *inter-passivity* or *de-interactivization* (cf. Pfaller 2008; Ackermann 2016b) is faced by the *Let's Players* by a kind of *interactivity by proxy* (cf. Ligman 2011). This means that players convey their own interactive immersion by letting viewers participate in their immediate experiences. A prevalent technique to achieve this is by various practices of *embodying avatars*.

The first *Let's Plays* were created in the year 2006 on the platform *Something Awful*.² These videos are still very popular, ranking among the most popular videos on YouTube, and representing almost the entire content of the online live streaming platform *Twitch*.³ The German *Let's Player Gronkh*, for instance, has 4.8 million subscribers for his YouTube-channel. *PewDiePie*, a well-known global *Let's Player* from Sweden, can even record 60 million.

There are two different ways in which *Let's Plays* can be presented. One possibility is to record the gameplay as well as the video feed including the facecam and upload this video to websites like YouTube. Another possibility is to livestream the gameplay and the video feed on special platforms like *Twitch*. In this case, spectators usually have the possibility to participate directly via live chat tools, creating additional opportunities for interaction (cf. Schmidt/Marx/Neise 2020; Recktenwald 2017). In the following we focus on recorded *Let's Plays*. In contrast to live streams, no interaction via chat with an audience *during* gameplay is possible.

In the next section we introduce the notions of embodiment and tele-presence and argue that they play a crucial role when video games are presented to an audience.

² See <https://www.somethingawful.com/>.

³ Twitch is a live streaming video platform that specializes on live streams of *Let's Plays*.

3 Presented Gaming, Embodiment and (Tele-)Presence

As mentioned above, when computer gaming is presented for an audience, a specific kind of *participation framework* arises involving three levels: First, a human-machine interaction (HMI) between gamer(s) and game software; second, interactions between avatars within a represented virtual reality (VR); and finally, social para-interaction (SPI) between gamer(s) and viewers. The three levels are intertwined as the first level (HMI) generates the second level (VR) which in turn provides the content for the third level (SPI). It is this specific relationship that creates the affordances to embody avatars for viewers and, by that, to stage (tele-)presence. In the following we would like to take a closer look at these relationships.

Actions in computer games are usually mediated by movements of avatars. Players control most of the movements of their avatars via interfaces using technical devices such as a game controller or a mouse and a keyboard. The results of the gamers' control actions are displayed on screen as movements of their avatars. Like this, players can monitor the effect of their control actions immediately on screen. Thus, players and avatars are connected to each other by means of a cybernetic control loop in which inputs generate immediate outputs that are, in turn, the basis for further inputs as shown in Figure 2.

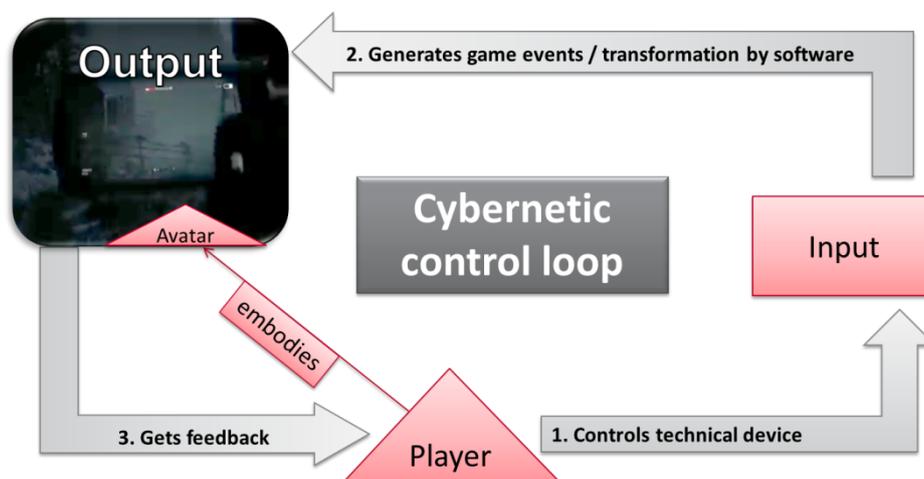


Figure 2: Gaming as cybernetic control loop

The close cycle between the player's control actions and the avatar's movements on screen forms the basis for an interactive and immersive game experience (cf. Klimmt 2006, 76 et seqq.). When an avatar, for instance, is knocking at a door in the game world, this 'action' is generated by a player's control action (e.g. by pressing a certain key on the keyboard). The avatar's action therefore has two

sides: First, it is triggered by a technical operation outside the game world; second, it is at the same time a representation of an action in the game world (in this case a 'door-knocking-action') that is visually displayed on the screen. Consequently, immersion is understood as a (partial) occupation of our senses by the events within virtual reality (Lombard/Ditton 2006).

This connection between controlling the game (HMI) and representation of the game (VR) is presented in *Let's Plays* in a certain way for viewers. Only the latter, the represented action, is visualized in *Let's Plays* and thus accessible for viewers. The real-world action of manipulating the controller remains invisible for spectators (cf. Schemer-Reinhard 2016). For spectators, the technical control action (e.g. pressing a key) and the representation of that action on screen (e.g. an avatar knocking at a door) therefore merge into one action. As a result, the experience (and pleasure) of playing a computer game is emulated for viewers as the representation acts as if there is only one relevant level of action, namely the events in the virtual world. As we will see in the data, this is additionally enforced by *Let's Players* as they often comment on such actions with expressions such as *I'm going to knock at that door*, thereby referring to the avatar's action as their own action. Using this kind of footing (cf. Goffman 1981a), the identities of players and avatars seem to overlap. By saying *I'm going to knock at that door*, players refer to both themselves and their avatars controlled by them as agents of the announced action. Such a construction of *player-avatar-hybrids* (cf. Baldauf-Quilliatre/Colón de Carvajal 2015, 2019) as agents in/of the game through the use of language enables and enhances illusion and immersion for viewers. It creates the impression that players are *directly active in the play world*.

These affordances enable and constrain a specific kind of participation framework which is, in turn, the basis of a specific kind of embodiment of avatars conducted by gamers for viewers. The notion of embodiment, roughly speaking, emphasizes the reflexivity of cognition and situated behavior, above all senso-motoric coordination (Clark 2001; Gibbs 1995; Rohrer 2007; Suchman 1987; Wachsmuth/Lenzen/Knoblich 2008). Most importantly, cognition is seen as rooted in the body so that perception is only possible within a functional cycle of sensing, kinesthetic and movement (Lakoff/Johnson 2011; Streeck 2008; Streeck/Goodwin/LeBaron 2011).

What occurs as a unit in real life is separated and reconnected via media technology in computer games. Acting in computer games is often mediated by *avatars* generating *two poles of agency*, the gamer and the avatar. In addition, gamers are involved in two situations simultaneously, the situation of playing a game in the real world and

the situation of being in a game in the virtual world. For most gamers, this creates a very enjoyable experience because they feel as if they really are part of the virtual world when they play, and so for them the real world can become temporarily replaced by the virtual world (cf. Klimmt 2006). Players are perceptively and psychologically immersed (Lombard/Ditton 2006). When gamers embody their avatars for viewers, they basically convey their own experience of acting in a virtual world. They reveal their thoughts and construct access to their (emotional) experiences during their actions. As outlined above, gamers and their avatars are closely connected. With respect to agency, they merge into player-avatar-hybrids. Taking this into account, the verbal and bodily construction of access to their 'inner states' (thoughts, feelings etc.) is transferred to their avatars. By that, avatars are equipped with human features (such as being sensitive, rational etc.).

Virtual reality in computer games, especially if presented in a point-of-view-perspective, “provide media users with an illusion that a mediated experience is not mediated” (Lombard/Ditton 2006: 1) which “creates for the user a strong sense of presence” (Lombard/Ditton 2006: 1). In view of recent developments in the field of media technology to convey interaction more and more realistically, traditional notions of presence are questioned (cf. Licoppe 2015; Spagnolli/Gamberini 2005). Traditionally, presence was tied to the spatial concept of situation, which means that two (or more) people are required to be in the same place at the same time (Goffman 1963; Gumbrecht 2012). With the help of technology, however, it is possible to create a realistic representation of interaction partners who then appear to be present or at least tele-present (Höflich 2005; Meyrowitz 1990; Shanyang 2005). In addition, virtual worlds promote immersive experiences that replace (or at least superimpose) the perception of a real world by that of the virtual world creating a strong sense of presence. When users consider something mediated to be present, they tend to “respond directly to what they see and hear in a mediated experience, as if what they see and hear was physically present in their viewing environment [...]” (Lombard/Ditton 1997: 10). This is exactly what *Let's Players* do to entertain their viewers: They treat the virtual world and react to it as if it were physically present. In the following we are interested in *Let's Players*' verbal and embodied practices of creating such a sense of presence.

4. Data and Method

To illustrate the core strategies and practices *Let's Players* employ for making *Let's Plays* watchable we draw on an example of a so-called 'blind play' (where the game has not been played before by the players) by the popular German player *Pan*⁴, who presents a current computer game from the adventure-action-genre. A few examples are taken also from multiplayer *Let's Plays*, in which several well-known German *Let's Players* participate in a joint adventure-horror game. The combination of blind play and the adventure/action/horror genre promises situations that are potentially unpredictable and/or surprising and therefore both require explanation by the player(s) and provoke spontaneous reactions. Selecting popular *Let's Players* offers us a chance to pinpoint more typical, well-established practices in this community of *Let's Players*.

Our analysis is based on a screening of ten different single and multiplayer *Let's Play* videos, all published on YouTube (equivalent approx. 20 hours) and an in-depth analysis of the two selected videos mentioned above (equivalent to approx. two hours).

In the following section five, we first illustrate briefly the two focused practices used by *Let's Players* to make their presented video gaming 'watchable' – formulating one's own actions and animating avatars via response cries – drawing on two simple and clear cases (5.1). In the following two sections, we present a more detailed analysis of both practices (5.2, 5.3).

5 Analysis

5.1 Two central practices and their function for making *Let's Plays* 'watchable'

In the following analysis we investigate two highly frequent practices used by *Let's Players* to embody their avatars for viewers. The first practice relates to *formulating one's own actions*, while the second one deals with *animating avatars via response cries*.

A very simple example for *formulating actions* is shown in the following transcript: The gamer introduced above, *Pan*, plays a demo version of the action-adventure game *Outlast 2*. The transcript shows her talk (original German with an English translation in bold) as well as important game events (GE) and game

⁴ *Pan* is the short form for the pseudonym *Pandorya*, a well-known German *Let's Player*, and at the same time her nickname as *Let's Player*.

sounds (GS). Special characters (such as * or ~) align non-verbal events with talk.⁵ Still images are represented by an extra line termed *Fig*, and their exact position in relation to talk is indicated by a hashtag (#).

Transcript (1): knocking (video 1)

```
01 P   ich klopfe hier einmal *~AN.#
      i'm gonna knock here now
      GE                               *avatar's
                                       fist/knocking
      GS                               ~knocking sound
      Fig                               #fig. 3
```

In this example, *Pan* is performing a *knocking-action* with her avatar which is represented in a point-of-view-perspective (cf. Figure 3).⁶

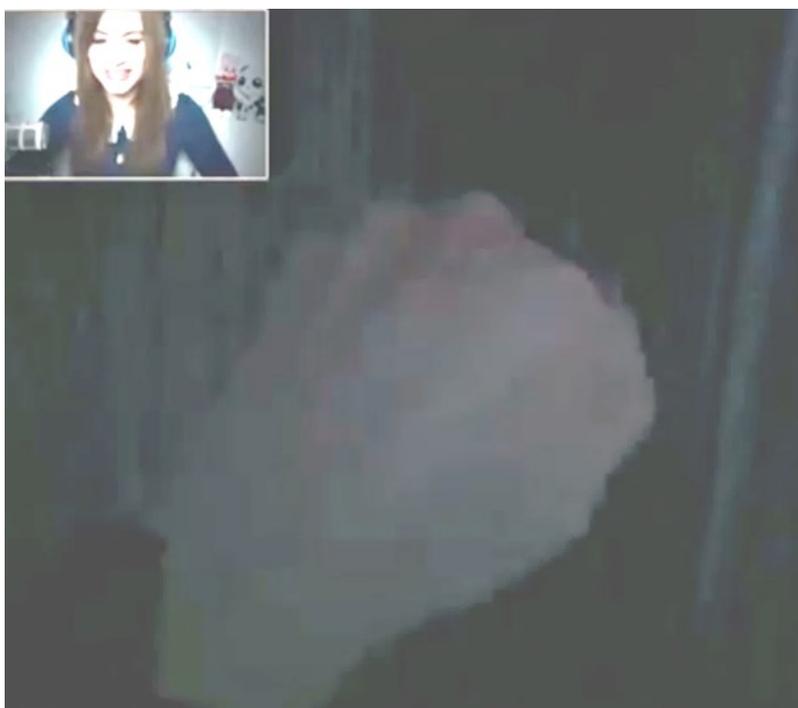


Figure 3: Knocking action in point-of-view-perspective

The action is not only carried out by manipulating the avatar but is also *verbalized* by saying “*Ich klopfe hier einmal an*”|“*I’m gonna*

⁵ Talk is transcribed according to GAT2 (cf. Selting et al. 2009), embodied conduct according to multimodal conventions following Mondada (2014). For a more comprehensive overview of the transcription convention used, see Appendix.

⁶ Video games in point-of-view-perspective enhance the impression, for players and viewers alike, to be directly active in the game (cf. Neitzel 2013).

knock here now” (line 1). With this verbalization, *Pan* is letting us know what she is doing. We call such practices *formulating actions*.

A simple example for *animating avatars via response cries* is the following case: *Pan*’s avatar is hiding in a locker in a deserted hallway. As she is about to open the door in order to get out again, a monster suddenly appears in front of the locker. Her strong physical reaction is accessible via the integrated facecam (FC).

Transcript (2): being scared (video 2)

01 Pan kannst du mal BITte *wieder;
 can you please again
 GE *starts to open the locker-
 -->*

02 Pan DANke schön das wär super*
 thanks that would be great
 GE -----*

03 **\$(0.5)**
 GE \$monster appears suddenly in front of the
 locker

04 Pan *eeeeehhhhhhh
 eeeeehhhhhhh
 FC *distorts the face, tears both hands upwards,
 jerks back from the screen
 Fig #fig. 4

05 ***(0.5)**
 FC *grabs her head with both hands, eyes and mouth
 torn open

06 Pan *huuuuuaaa
 huuuuuaaa
 FC *takes hands down, gets closer to the screen
 again

In this example, *Pan* shows strong *embodied reactions* to an unexpected game event, the sudden appearance of a monster. The extract begins with *Pan* talking to ‘the game’ by requesting for a possibility to get out of the locker (line 1), which she thanks for when she figures out how to reopen the door (line 2). At this point the monster appears, and after a half-second pause (line 3), *Pan* moves her upper part of the body quickly away from the screen and produces a fright sound accompanied by a corresponding facial expression (line 4; cf. Fig. 4).



Figure 4: Appearance of a monster and *Pan's* scared reaction

With this reaction, she displays a stance (she is shocked) towards an emergent result in the game (the appearance of a monster). At the same time, the game event is supplemented by an immediate and physical reaction of the gamer. Such reactions not only enrich game events with additional meaning (cf. also Recktenwald 2017), but give avatars a lively, audio-visually accessible experience. We call such practices *animating avatars*.

In the following we argue that there is an *inherent relation* between *formulating actions* as an attempt to make intentions accessible and *animating avatars* as displaying spontaneous reactions which add an emotional dimension to represented game events. That is, in *Let's Plays*, both forms of comments, formulating actions and animating avatars, can be understood as different but *complementary practices of embodying avatars*.

In the following we aim to support this hypothesis by first discussing practices of *formulating actions* and, secondly, practices of *animating avatars* in order to show their *capability to embody avatars*.

5.2 Formulations

Especially in *Let's Plays* which are dedicated to an audience, *Let's Players* are forced to make their gaming activities attractive for viewers. This is done by making them more *transparent* with the help of *verbal moderation* like for example when players tell viewers what they plan to do next or which problems they may

expect (cf. Marx/Schmidt 2019). Within the moderation a frequent practice is to *verbalize one's own actions*. A case in point is the formulation of the knocking-action as discussed above (“*I’m gonna knock here now*”). The way in which *Pan* formulates this action is designed to *ascribe intentions* to the visible actions of her avatar. Importantly, saying *I’m gonna knock* is not formulating *her* action of manipulating the game controller (like pressing a key to trigger the knocking action in the game); rather it formulates the result of her control action, which is an *audio-visual representation* of an avatar knocking at a door. By this, her *intention* to knock in the game by pressing a certain key is *transferred* by the game technology to a visible and *accountable action* of an avatar in the game world. Nevertheless, she is using the pronoun *I*, thereby conflating her own actions with those of the avatar. Therefore, formulating one’s own actions in this way equips avatars with plans, intentions and, in the long run, with rationality. We seemingly get to know why an avatar is acting in a certain way within the game.

If we have a closer look at the knocking-example with respect to its temporal structure, we realize that the verbal part precedes the represented action on the screen, which immediately follows in slight overlap (Transcript 1 is presented here again as Transcript 3 for convenience):

Transcript (3): knocking (video 1)

```

0 p    ich klopfe hier einmal *~AN.#
1      i'm gonna knock here now
      GE                                *avatar's fist/knocking
      GS                                ~knocking sound
      Fig                                #fig. 3

```

This is the typical ordering as it occurs in most of our instances. A *verbal formulation* (line 1: *I’m gonna knock here now*) is followed by a *corresponding action* in the game. Thereby both parts are tied together in a *reflexive* way, in that the verbal part appears as a projection and the avatar’s action itself as its fulfilment. This is made possible by the fine-grained temporal coordination of verbal projection and embodied implementation. At the same time, game actions get a verbal label (here: *knocking*) whereby visually accessible conduct is unambiguously categorized (here as *knocking*).

Sometimes next actions are not only verbally projected, but specific expectations concerning the results of next actions may also be explicitly anticipated. This is done by embodied conduct indicating expectations. In the following example *Pan* is exploring a room with several doors. One of these doors is half open, which motivates her to take a closer look at the door. Her exploration is

accompanied by facial expressions and shifts in body posture which indicates anticipated trouble.

Transcript (4): anticipation (video 3)

01 Pan +(xxx) ^was ist HIER?
 (xxx) what is here
 GE +approaches a half-open door--->+
 FC ^tilts head to side--->^

02 **+(0.6) ^#(2.23)**
 GE +starts to open the door--->+
 FC ^pulls tilted head back, twists mouth-->^
 Fig #fig. 5

03 Pan YE::^%AH,
 yeah
 FC ^bends her head forward--->>
 GE %turns on Night Vision--->>

04 **(0.4)**

05 Pan +halLO?
 hello
 GE +moves towards/through the door

When *Pan* is approaching the half-open door, she produces the question *what is here?* (line 1) which serves as an *exploratory announcement* indicating her (following) action as ‘*exploring something*’.⁷ She tilts her head slightly to the side as if she is peeking carefully at what is behind the door. When she starts to open the door (line 2) she pulls her tilted head back and twists her mouth (cf. Fig. 5).

⁷ For a more extensive analysis of exploratory announcements in *Let's Plays* cf. Marx/Schmidt 2019.

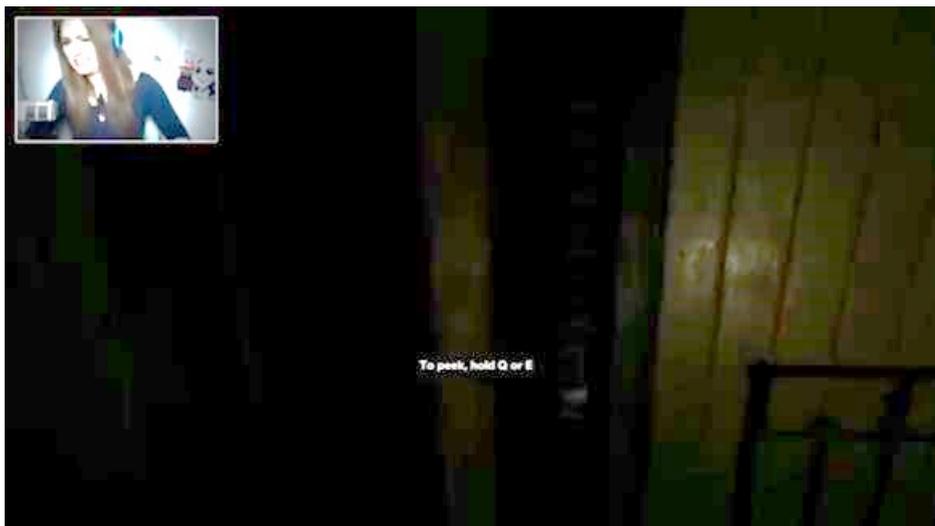


Figure 5: Facial expression anticipates trouble

By her embodied conduct, especially by her posture shift and her facial expression (cf. Fig. 5), *Pan* indicates that she *expects* to be scared by a sudden event (as it was the case above with the monster). But in this case, she makes her *expectation* of a specific possible result of her exploring action publicly available before being confronted with the actual outcome. She maintains her strained (facial) expression for several seconds (line 2), and only relaxes it when the potentially dangerous situation is resolved, accompanied by a drawn out *Yeah* indicating relief (line 3). At the same time she turns on the night vision to have a better view. Once the danger is over, she starts to explore what is behind the door (line 4).

Action formulations and embodied expectations (like in the case before) are not always used in an anticipating function preceding the next action. They can occur at *different temporal positions* within an action process. Besides their occurrence as projections in initial position (like in the two cases above) they can also appear *during* and *after* action processes. As we will see in the following examples, they can mark actions as relevant next actions, as ongoing actions or as completed actions, depending on their temporal positioning.

In the next example, *Pan's* avatar moves quickly through a cornfield. As she performs that action, she repeats the phrase *I'm looking for something*, thereby indicating an action as ongoing:

Transcript (5): I'm looking for something (video 4)

```
01 Pan    ^oKAY,
          oKAY
          GE    >>moves through gaming world--->>
          FC    ^leans slightly forward --->>
02 Pan    ich SUche was,
          i'm looking for something
```

03 Pan ich SUche was,
 i'm looking for something
04 Pan ich SUche was;
 i'm looking for something

In this example *Pan* repeats the formulation *I'm looking for something* three times. All three instances are very similar. They use the same words in the same order, and prosodically they are produced in a nearly identical manner. This redundancy, achieved by almost exact repetition, is used to signal that the *action is still underway* and not yet completed. As Stivers (2006) has shown, the main function of multiple sayings is displaying that the turn is addressing an in-progress course of action. Furthermore, the repetitions here indicate that there is no change with regard to the ongoing action process, and especially that it has not yet been successful.

However, by using the expression *looking for something*, a specific kind of outcome is anticipated, which is *finding something*. This means that the current action is not only categorized (as *looking for something*) and marked as ongoing by accompanying verbalizations while it is underway, but at the same time it is prospectively limited by implying a possible end point (*finding something*).

Formulations are not only used before and during actions processes, as in the examples above. They can also be used *after* the completion of an action process. However, usually they follow projections, so that they hardly ever occur alone. Thus, they are embedded in *sequences* of projecting a next action, conducting that action and finally evaluating it afterwards.

In the next example, two such sequences of action are concluded by evaluative formulations (lines 3 and 6). We join the action when *Pan* is exploring a room she has just entered:

Transcript (6): action sequence (video 5)

01 **(0.2)**
GE >>focus on a book on a table in a dark room ->*

02 Pan kann ich das hier LEsen?*

can i read this

03 Pan *nein kann +ich NICH.

no i can`t

GE *defocus book
 +focus cupboard/attempt to open--->+

04 **(0.29)**

05 Pan kann ich den SCHRANK-

can i the cupboard

06 Pan +°h ich kann den SCHRANK a-
 °h i can the cupboard o
 GE +zooms in on cupboard >>
 07 Pan °h (.) A:A:H.
 °h (.) aah

In this case, two action processes are projected (lines 1 and 5) and respectively completed by statements about the action's outcome (lines 3 and 6) which can be understood as evaluations as they assess the success of the projected actions retrospectively. The first one starts with an exploratory announcement in form of a question (line 2: *can I read this?*) followed by a negative evaluation of the action's outcome (line 3: *no, I can't*). The second one is launched in a similar way by an aborted question (line 5: *can I the cupboard...*) and followed by a positive evaluation (line 6: *I can the cupboard...*). Between projection (questions: *can I...*) and evaluation (*I can't/I can*) the projected actions are tested (*reading the titles of a book | opening a cupboard*). In both cases the final evaluation signals a *completion* of the action process and is bracketed by initial and terminal formulations. As the second action sequence (cupboard) shows, verbal formulations are adapted to the pace of visually conveyed action processes (and not vice versa). Once *Pan* has managed to open the cupboard, she aborts her question (line 5) and proceeds seamlessly to an evaluation (line 6) which she also aborts in favor for a change-of-state-token⁸ (a drawn-out *ah* in line 7) conveying that she has learned something. The fact that the action sequence *try to open the cupboard* is treated as completed at this point is also indicated by a camera action, the *zooming in* on the cupboard (line 6). By doing this, *Pan* switches her focus from opening to exploring the cupboard.

Interestingly, by formulating their actions before, during and after their conduction, *Let's Players* not only reveal their plans, but sometimes they project an expected result on the basis of which the actions can be evaluated afterwards. This becomes particularly apparent in the case of failures. In the following example *Pan* explores a building from the outside.

Transcript (7): failure (video 6)

01 (1.13)
 GE >>focus on a window--->+
 02 Pan ich kann hier REINGucken;*
 i can look inside
 03 Pan +°h oke wahrscheinlich kann ich auch da
 REINGehn;

⁸ Cf. Heritage 1984b.

°h okay probably i can also enter there
 +panning movement, starts moving--->+
 04 Pan das werd ich ja wo_ma direkt MACHEN,
i will directly do that now
 05 **+(1.89)**
 GE +moves tw. entrance, tries to open the door-->+
 06 Pan +HÄÄh;
Huh
 07 **(0.39)**
 08 Pan halLO?
Hello
 09 **(0.46)**
 10 Pan KANN ich nich;
i can't
 11 Pan (.) OH;
(.) oh

In this example, visual evidence (line 2: *I can look inside*) leads to an inference (line 3: *I can also enter*) which serves not only as a projection for a specific kind of next action (*entering the house*) but at the same time as a projection of an expected (successful) outcome (*being able to enter the house*). After announcing the intention to do so (line 4: *I will directly do that now*), the following conducted action of opening the door of the house fails (line 5) and is commented on by a response cry-like surprise sound (*hääh/huh* in line 6), followed by a summon (*hello* in line 8). Both reactions convey her disbelief to have failed. Finally, the action sequence is evaluated with *I can't* (line 10) and a change-of-state-token (a freestanding *oh* in line 11) retrospectively contextualizing the action's outcome as surprising and the information sequence as complete (cf. Heritage 1984b).

By *projecting* a possible result of her action (*being able to enter the house*), *Pan* makes the *conditions of a failure explicit* (in this case not being able to enter) before actually conducting the action. By this, visually conveyed action processes in the game get both a *projected course* by marking start and end points and a normative structure by *projecting expected outcomes*. Both contribute to making *Let's Plays* more *transparent*.

Making action processes transparent in order to involve viewers is particularly important when it comes to *less obvious actions*. The next example is an extended version of transcript 5 where *Pan* is moving quickly through a cornfield:

Transcript (8): I'm looking for something (extended) (video 7)

01 Pan ^oKAY,
oKAY
 GE >>moves through gaming world--->>
 FC ^leans slightly forward --->>

02 Pan ich SUche was,
 i'm looking for something

03 Pan ich SUche was,
 i'm looking for something

04 Pan ich SUche was;
 i'm looking for something

05 **(1.08) & (2.0)**
 GE &Night Vision

06 Pan *DAS suche ich.
 that's what i'm looking for

 GE *water container visible, moves towards it

07 ***(0.92)**
 GE *climbs into water container

In this case, if only the images were available, the viewer would see a flow of movements that is hardly recognizable as a certain activity. Only the verbal descriptions of this flow of images as *looking for something* enables the viewer to comprehend what is happening on the screen. This example highlights how otherwise apparently *random movements* on the screen are framed by the *verbal formulations* as *rational and accountable* activities, in this case as a process of *searching*.

Furthermore, the flow of visual representations is *packaged* into comprehensible *action units*, in this example starting with a search lasting a while (marked by the repetitions in lines 2-5) and ending successfully, in this case by finding a specific place to hide (a water container), which is also verbally announced (line 6: *that's what I'm looking for*). Note that the found 'object' is not explicitly named but referred to with the demonstrative pronoun *that*. By using *that* (instead of a referential noun like *water container*), the visually conveyed game world and the talk about it are reflexively tied together as we have to scan the images in order to detect the reference of *that*. In turn, only by the verbal formulations, parts of the images get the status of a searched/found object. Moreover, by indexically referring to the game world, *Pan* makes her perception relevant. To understand the meaning of *that*, we have to see the game world with her eyes or – as Hausendorf (2003) has put it – we are invited to perceive her perception.

5.3 Animations via response cries

Game events are not only verbally formulated, but also commented on via *exclamations* or (in Goffman's words) via *response cries* which convey a *bodily involvement* and a *player's stance* towards actual game events. Having access to *reactions* of the player that appear *spontaneous* (transmitted via facecam) makes the game more transparent and attractive for viewers. A case in point is the *shock*

reaction of the player *Pan* described above, when suddenly a monster appeared (cf. transcript 2).

Reactions of this kind are very frequent. In our selected 30-minutes *Let's Play* from the player *Pan*, response cries occur every 30 seconds on average; altogether we found more than *105 instances* (for a compilation see video 8). They occur in lexicalized forms (such as *shit, oh my god* etc.) and non-lexicalized forms (e.g. shock cries as in the monster example, or pain cries, cf. below). They are related to various events, such as sudden game events (cf. transcript 2), anticipated game events caused by own actions (cf. transcript 6), the game control, status displays or simply the development of the story within the game. Depending on how demanding the game is at a certain point, forms, functions and density of response cries may vary. Especially when the game gets particularly thrilling, the use of response cries increases and tends to replace verbal comments overall. Furthermore, response cries play a crucial role in structuring and constructing action processes and thereby assign meaning to the images conveyed by the computer game (cf. section 5.2 above). Sometimes they animate avatars in terms of reacting or speaking on their behalf (see below the comments on *pain cries* and *interaction with non-play-characters*).

What all cases have in common is that the use of response cries connects game events with players' displayed emotional stances and thus roots them in physical experience. Often the results of players' announced actions are qualified affectively afterwards by response cries (cf. transcript 6). Goffman (1981b) has argued that response cries are designed to convey an internal state without doing it in an explicit communicative way (cf. also Baldauf 2002, Heritage 1984b). Saying *Ooops* in a public place, e.g. if you are stumbling over a step, signals to all those present within earshot that the stumbling person realized that it was an accident. In this way, it is framed as an exception and, by that, individuals show that they are normal members of society who are aware of their misconduct. At the same time, uttering *Ooops* does not oblige anybody to engage in an interaction or conversation. Response cries can be registered without any comment. One reason for that is that they are read, partly because of their non-lexicalized sound structure, as direct reactions of the body, which are then more of an indicator or an indexical sign than a full-fledged symbol or sign such as lexical units like words. Response cries are therefore seen as more rooted in autonomous reactions of the body than verbal assertions, for instance, which may refer to the same or similar circumstances. Their notoriously ambivalent status between autonomous bodily reactions and deliberately produced communicative acts constitutes their specific functional potential.

In this extract, C tries to fix a generator while talking about something else (line 1). He fails and gets an electric shock (line 1). In reaction to this, C produces a pain cry and a swearword (line 3: *ouch shit*).

(c) ahh

```

01B      glaube wir %~*könn da hier ^#[(xxx)]
          believe we can here xxx
02C      [!AH! ]
          ah
GE      >>C runs in crouching position
GE      %A hits C with machete from behind
GS      ~stroking sound
GE      *C falls over
FCC     ^moves back, eyes
          wide open
Fig.     #Fig.6
03A      du HUM[peilst nich mehr,]
          you're not limping any more
04D      [NEI:N NEI:N; ]
          no no

```

In this example, the killer (A) hits C with a machete from behind (line 2). C shrugs back from the screen with his eyes wide open and cries out loud (line 2), drowning out the end of B's prior utterance (line 1). Partly overlapping with A's ironic comment (line 3), C comments on his situation by saying *no* twice in a modulated voice (line 4).

In this collection of examples, represented physical states and pains of avatars caused by virtual events are embodied by the players producing both vocal sounds such as pain cries, whining and moaning, and embodied conduct such as bodily position changes and facial expressions accessible via the integrated facecam. Figure 6 shows C's reaction when hit by the killer from behind.

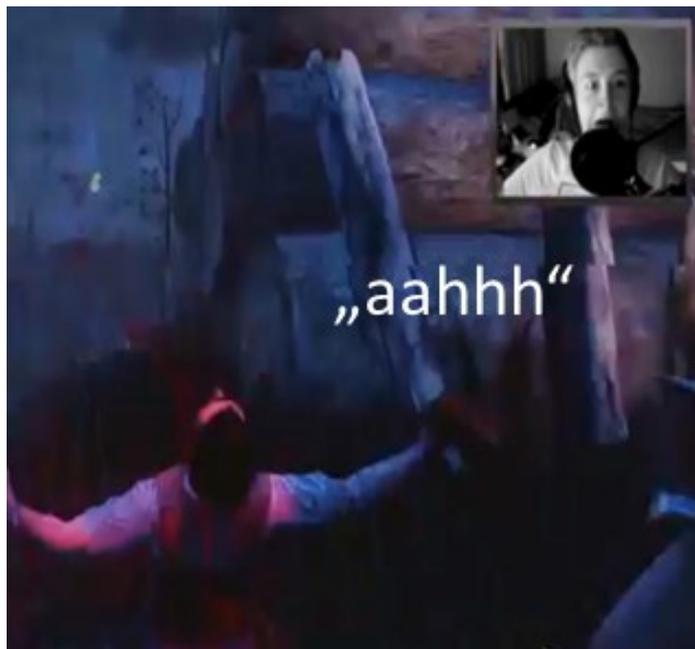


Figure 6: C's reaction when hit by the killer

The physical expressiveness indicates not only their involvement in the game but adds more liveliness to the overall experience of watching the gameplay.¹⁰ Pain cries in particular appear to be directly connected to physical experiences (cf. Ehlich 1985; Mannheim 1980).

The animating effect of response cries gets especially obvious when gamers speak as avatars with non-play characters (NPC)¹¹ as in the next collection of examples:

Transcripts (11): collection of interaction with NPC (video 10)

(a) leave me in peace

```
01          ** (0.81)
GS          >>threatening music>>
GE          *NPC reaches for her and tries to hold her>>
GE          +tries to escape>>
```

¹⁰ In television studies, 'liveliness' basically means – following Tolson (2006) – the impression of spontaneous (inter)actions despite their staged character, which leads to “(...) performances which are like acting in some ways but are not ‘actorly’” (11). Comparable concepts are Fairclough’s “conversationalization” (1995, 8), Goffman’s (1981a) “fresh talk”, Ochs’ (1979) “planned unplanned discourse” and Clark’s “extemporaneity” (1996, 9). In our case, it is not only and not primarily the (apparently) unscripted talk what enables the impression of liveliness but also and in particular the (apparently) spontaneous bodily reactions to virtual events in the game.

¹¹ Non-play characters (NPC) are characters in the game that are not controlled by a human player, but by the software of the game.

02 Pan !FUCK!,
Fuck

03 ***(0.22)**
 GE *fast camera movements

04 Pan <<lachend> WOAHA SCHEIÙe >
<<laughing> wow shit >

05 **~(1.78)**
 GS ~hittig sound

06 Pan !NEIN!,
No

07 **(0.29)**

08 Pan LASS mich;
leave me in peace;

09 **(0.42)**

10 Pan LASS ~mich;
leave me in peace

GS ~hitting soundsound, puffing sound

11 **(0.92)**

12 Pan lass mich RAUS,
let me out

13 **(0.55)**

14 Pan LASS mich einfach raus;
just let me out

15 Pan *lass mich einfach RAUS,
just let me out
 *escapes und flees>>

In this extract, *Pan* is suddenly attacked by an NPC (line 1) which tries to catch her. During the whole extract she tries to escape and get rid of him. First, she comments on her situation via response cries (*fuck* in line 2, *wow shit* in line 4, *no* in line 6), later she addresses the NPC directly telling him to stop (lines 8–15). As she slowly manages to free herself, her voice gets calmer and softer (lines 14 and 15).

(b) no no no no no

01 **(1.0)* (0.2)~ (1.0)**
 GE *NPC approaches, reaches for her
 GS ~hitting sound

02 Pan *NEI:N nein nein nein nein ~nein nein
 nein nein nein nein nein nein nein nein
 nein nein,
no no no
no no no

GE *escapes and flees>>
 ~hitting sound

03 Pan lass mich in RUhe,
leave me in peace

04 Pan lass mich in RUhe.
leave me in peace

As in the extract before, *Pan* is attacked and pursued by an NPC (line 1). This time she is able to escape a scuffle and flees directly. Her escape is accompanied first by a series of response cries (multiple *noes* in line 2), later by direct calls to the NPC to leave her alone (lines 3 and 4). Furthermore, her multiple sayings display that her turn is addressing an ongoing 'action' (Stivers 2006), in this case of an NPC.

As Aarsand/Aronsson (2009) have argued, animating avatars is a common way of engaging in joint gaming. In the above examples, *Pan* uses several response cry-like expressions (like outcries, negation particles etc.) when she is attacked by a non-play character. She not only reacts to game events but stages them as an interaction with a real counterpart. By using imperatives, for instance, (like get away from me, leave me in peace etc.), and a stronger voice, she treats the non-play character as a person-like figure that can change his behavior through interaction. This results in a more animated nature of her gameplay and the presented actions of her avatar.

6. Conclusion

Using the example of *Let's Plays*, we have shown the close connection between affordances, participation framework and practices of *Let's Players* to make their gaming 'watchable'. *Let's Plays* are video games in which players get involved in a virtual world by using software. Involvement basically means that the control actions of players are translated into represented movements of avatars on a screen which are interpreted as actions in a virtual world. As players and game/avatars are closely interconnected through a cybernetic control loop, actions of players get immediate feedback. This is the precondition of being immersed in a virtual world. Immersion means that our sense of presence has shifted from the real world to a virtual world. One indicator for immersion in a virtual world is that players sense and react to virtual rather than real world events. The pleasure and attractiveness of video games is largely the result of this possibility to be active and immersed in a virtual world.

That said, video games are not, at least not in the first place, made to be presented to an audience. The sense of being present in a virtual world that video games enable is first off only a gratification for active players, not for those who watch how others play. To make watching video gaming attractive, *Let's Players* try to convey this sense of presence to the audience.

Two aspects, as we have seen, are crucial for achieving this: First off, the screen representation of *Let's Plays* involves the game play

itself and the player appearing in a facecam. By simultaneously representing the game and those who play the game in a split-screen mode, the most important feature of video games is highlighted: its close connection between control actions and actions in a play world in form of a cybernetic control loop. Like this, for viewers, game actions and reactions of the player(s) are directly related. Based on this tele-presence, the viewers are able to vicariously experience the players' sense of presence continuously and immediately.¹²

Secondly, as we have shown, *Let's Players* have to invest a lot of work to actively achieve this kind of illusion. They are continuously oriented towards their viewers providing them with comments on their actions in the running game. Thereby they stage tele-presence with regard to viewers. For instance, saying *I'm now going to knock at that door* while performing this action in the game world is constructed to make the move in the game tangible for viewers. Its construction systematically takes viewers into account thereby staging their own experience with the goal to create a lively representation of their sense of presence (which is then, given the mediated character of *Let's Plays*, a kind of tele-presence).

As we have shown, *Let's Players* basically draw on two kinds of practices to convey their sense of presence when playing the game. On the one hand, they formulate their game actions in order to give action processes a recognizable and rational structure, thus making them more comprehensible. On the other hand, they produce response cries in reaction to game events. Producing response cries during the game not only indicates high involvement and apparently grants viewers access to players' emotions, but also contributes to an animation of avatars. Both formulating actions and animating avatars via response cries are part of *embodying* avatars. Formulating actions equips avatars with cognition (e.g. intentions), animating with sensibility (e.g. pain sensation). Both practices interact to make the visual events on screen understandable as actions. By bringing intentions (or more general: inner processes of consciousness) and external behavior together through spoken discourse (mainly by formulating actions) and embodied conduct (mainly by animating avatars via response cries) gamers facilitate the perception of avatars as full-fledged persons/characters. This is an important precondition

¹² Basically, this rests on mechanisms of vicarious role-taking, which is always relevant when we observe the conduct of others (cf. Ellis/Streeter/Engelbrecht 1983). In our case, the prerequisites for this are the continuously and immediately conveyed images of both the game actions and the reactions of players in the facecam, which make 'playing a video game' a present-like – or a tele-present – experience (cf. Clark (1996, 9 et seq.) for immediacy as a feature of face-to-face-interactions).

to *understanding* and thus to *enjoying* what is happening on the screen.

Our contribution focused on practices of *Let's Players* which enhance an 'illusive' (Rapp 1973) experience of viewers. Such practices are designed to create the impression that players are directly active in the play world and that their avatars are capable of acting. However, *Let's Players* are not only deploying practices which are designed to create illusion (for viewers). There are instances in which they talk with their avatars, with the game itself (as in transcript 2) or in which they meta-communicate the fact of playing a video game (e.g. by discussing the production of the video game, its narration or the game control). Instead of concealing the process of mediation (as practices of embodying avatars do), meta-communicative practices disclose the impression of being present in a virtual world. Such practices do not create illusion, rather they destroy them. They are not illusive, they are 'inlusive', that is they are creating distance instead of immersion (cf. Rapp 1973). Further research is needed to explore this intricate relationship between opaqueness and transparency or illusion and inclusion in re-mediations such as *Let's Plays*.¹³

7. Appendix

Conventions for the notation of physical activities (cf. Mondada 2014)

Nonlinguistic events and activities

- appear after the abbreviations GE, GS, FC, and Fig
- in lines following pauses or conversation activities (without own number)
- are aligned with conversation/pauses with the help of special characters (like *, ~, + etc.) indicating the beginning and (if relevant) the end of events

¹³ Following Bolter/Grusin (2000), *Let's Plays* are understandable as a form of re-mediation as they mediate an already existing medium, the video game (see also Ackermann 2016b). Re-mediations can be opaque, then they promote the illusion of the re-mediated medium, or they can be transparent, then they expose the illusion of the re-mediated medium. In our contribution, we focused on practices which enhance an opaque re-mediation of video games in *Let's Plays*.

Further conventions for the notation of physical movement

--->	movement continues
--->\$	movement continues after the line until reaching
\$	the same sign
>>	continues after transcript ends
>>	starts before transcript

8. Bibliography

- Aarsand, André/Aronsson, Karin (2009): Response cries and other gaming moves – Building intersubjectivity in gaming. In: *Journal of Pragmatics* 41, 1557–1575.
- Ackermann, Judith (2016a) (ed.): *Phänomen Let's Play-Video: Entstehung, Ästhetik, Aneignung und Faszination aufgezeichneten Computerhandelns*. Wiesbaden: Springer VS.
- Ackermann, Judith (2016b): Einleitung. Entstehung und wissenschaftliche Relevanz eines Remediatisierungsphänomens. In: Ackermann, Judith (ed.): *Phänomen Let's Play: Entstehung, Ästhetik, Aneignung und Faszination aufgezeichneten Computerhandelns*. Wiesbaden: Springer VS, 1–15.
- Arminen, Ilkka/Licoppe, Christian/Spagnolli, Anna (2016): Respecifying Mediated Interaction. In: *Research on Language and Social Interaction* 49 (4), 290–309.
- Ayaß, Ruth (1993): Auf der Suche nach dem verlorenen Zuschauer. In: Holly, Werner/Püschel, Ulrich (eds.): *Medienrezeption als Aneignung*. Opladen: Westdeutscher Verlag, 27–41.
- Ayaß, Ruth (2005): Interaktion ohne Gegenüber? In: Jäckel, Michael/Mai, Michael (eds.): *Online-Vergesellschaftung? Mediensoziologische Perspektiven auf neue Kommunikationstechnologien*. Wiesbaden: VS, 33–49.
- Baldauf, Heike (2002): *Knappes Sprechen*. Tübingen: Niemeyer.
- Baldauf-Quilliatre, Heike (2014): Formen knapper Bewertungen beim Fußballspielen an der Playstation. In: Schwarze, Cordula/Konzett, Carmen (eds.): *Interaktionsforschung: Gesprächsanalytische Fallstudien und Forschungspraxis*. Berlin, 107–130.
- Baldauf-Quilliatre, Heike/Colón de Carvajal, Isabel (2015): Is the avatar considered as a participant by the players? A conversational analysis of multi-player videogames interactions. In: *PsychNology Journal* 13 (2–3), 127–147.
- Baldauf-Quilliatre, Heike/Colón de Carvajal, Isabel (2019): Interaktionen bei Videospiele-Sessions: Interagieren in einem hybriden Raum. In: Marx, Konstanze/Schmidt, Axel (eds.): *Interaktion und*

- Medien – Interaktionsanalytische Zugänge zu medienvermittelter Kommunikation*. Heidelberg: Universitätsverlag, 219–254.
- Bolter, Jay David/Grusin, Richard (2000): *Remediation: understanding new media*. Cambridge: MIT Press.
- Clark, Andy (2001): *Being there: putting brain, body, and world together again*. Cambridge: MIT Press.
- Clark, Herbert H. (1996): *Using language*. Cambridge: Cambridge Univ. Press.
- Deppermann, Arnulf (2013): Multimodal interaction from a conversation analytic perspective. In: *Journal of Pragmatics* 46, 1–7.
- Dynel, Marta (2014). Participation framework underlying YouTube interaction. In: *Journal of Pragmatics* 73, 37–52.
- Ehlich, Konrad (1985): The Language of Pain. In: *Theoretical Medicine* 6, 177–187.
- Ellis, Godfrey J./Streeter, Sandra K./Engelbrecht, Joann D. (1983): Television characters as Significant Others and the Process of Vicarious Role Taking. In: *Journal of Family Issues* 4 (2), 367–384.
- Fairclough, Norman (1995): *Media discourse*. London: Arnold.
- Freyermuth, Gundolf S. (2015): *Games, game design, game studies: eine Einführung*. Bielefeld: transcript.
- Gibbs, Raymond W. (2007): *Embodiment and cognitive science*. Cambridge: Cambridge Univ. Press.
- Goffman, Erving (1963): *Behavior in public places: notes on the social organization of gatherings*. London: Free Press.
- Goffman, Erving (1981a): Footing. In: Goffman, Erving: *Forms of Talk*. Philadelphia: University of Pennsylvania Press, 124–159.
- Goffman, Erving (1981b): Response Cries. In: Goffman, Erving: *Forms of Talk*. Philadelphia: University of Pennsylvania Press, 78–122.
- Goodwin, Charles/Goodwin, Marjorie Harness (2004): Participation. In: Duranti, Alessandro (eds.): *A companion to linguistic anthropology*. Vol. 1. Malden: Blackwell, 222–244.
- Gumbrecht, Hans Ulrich (2012): *Präsenz*. Berlin: Suhrkamp.
- Hale, Thomas (2013): *From Jackasses to Superstars: A Case for the Study of 'Let's Play'*. Online resource: https://www.academia.edu/5260639/From_-_Jackasses_to_Superstars_A_Case_for_the_Study_of_Let_s_Play_September_2013_
- Hausendorf, Heiko (2001): Warum wir im Fernsehen so häufig begrüßt und angeredet werden. In: Sutter, Tilmann/Charlton, Michael (eds.): *Massenkommunikation, Interaktion und soziales Handeln*. Wiesbaden: Westdeutscher Verlag, 185–213.
- Heritage, John (1984a): *Garfinkel and Ethnomethodology*. Cambridge: Polity Press.

- Heritage, John (1984b): A change-of-state token and aspects of its sequential placement. In: Atkinson, Maxwell/Heritage, John (eds.): *Structures of social action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press, 299–345.
- Höflich, Joachim R. (2005): Medien und interpersonale Kommunikation. In: Jäckel, Michael (eds.): *Mediensoziologie: Grundfragen und Forschungsfelder*. Wiesbaden: VS, 69–90.
- Hofstetter, Emily (2020): Nonlexical “Moans”: Response Cries in Board Game Interactions. In: *Research on Language and Social Interaction* 53 (1), 42–65.
- Horton, Donald/Wohl, Richard R. (1956): Mass Communication and para-social interaction: Observations on intimacy at a distance. In: *Psychiatry* 19, 215–229.
- Keating, Elizabeth/Sunakawa, Chiho (2010): Participation cues: Coordinating activity and collaboration in complex online gaming worlds. In: *Language in Society* 39, 331–356.
- Klimmt, Christoph (2006): *Computerspielen als Handlung: Dimensionen und Determinanten des Erlebens interaktiver Unterhaltungsangebote*. Köln: von Halem.
- Lakoff, George/Johnson, Mark (2011): *Metaphors we live by*. Chicago: Univ. of Chicago Press.
- Licoppe, Christian (2015): Contested norms of Presence. In: Hahn, Cornelia/Stempfhuber, Martin (eds.): *Präsenzen 2.0: Körperinszenierung in Medienkulturen*. Wiesbaden: Springer VS, 97–112.
- Ligman, Kris (2011): Let’s Play: Interactivity by Proxy in a Web 2.0 Culture. In: *PopMatters*. URL: <https://www.popmatters.com/lets-play-interactivity-by-proxy-in-a-web-2-0-culture-part-1-2496052960.html>, 06.02.2017.
- Lombard, Matthew/Ditton, Theresa (1997): At the Heart of It All: The Concept of Presence. In: *Journal of Computer Mediated Communication Computer-Mediated Communication* 3 (2). DOI: 10.1111/j.1083-6101.1997.tb00072.x.
- Mannheim, Karl (1980): *Strukturen des Denkens*. Darmstadt: Luchterhand.
- Marx, Konstanze/Schmidt, Axel (2019): Making Let’s Plays watchable – Praktiken des stellvertretenden Erlebbar-Machens von Interaktivität in vorgeführten Videospiele. In: Marx, Konstanze/Schmidt, Axel (eds.): *Interaktion und Medien – Interaktionsanalytische Zugänge zu medienvermittelter Kommunikation*. Heidelberg: Winter, 319–351.
- Meyrowitz, Joshua (1985): *No Sense of Place: The Impact of Electronic Media on Social Behavior*. Oxford: Oxford University Press.
- Meyrowitz, Joshua (1990): Redefining the Situation: Extending Dramaturgy into a theory of social change and media effects. In:

- Riggins, Stephen H. (eds.): *Beyond Goffman: studies on communication, institution, and social interaction*. Berlin: de Gruyter, 65–97.
- Mondada, Lorenza (2008): Using Video for a Sequential and Multimodal Analysis of Social Interaction: Videotaping Institutional Telephone Calls [88 paragraphs]. In: *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 9 (3), Art. 39.
- Mondada, Lorenza (2012): Coordinating action and talk-in-interaction in and out of video games. In: Ayaß, Ruth/Gerhardt, Cornelia (eds.): *The appropriation of media in everyday life*. Philadelphia: Benjamins, 231–270.
- Mondada, Lorenza (2013): Coordinating mobile action in real time: The timely organization of directives in video games. In: Haddington, Pentti/Mondada, Lorenza/Nevile, Maurice (eds.): *Interaction and mobility: language and the body in motion*. Berlin: de Gruyter, 300–341.
- Mondada, Lorenza (2014): *Conventions for multimodal transcription*. URL: https://franzoesistik.philhist.unibas.ch/fileadmin/user_upload/franzoesistik/mondada_multimodal_conventions.pdf.
- Neitzel, Britta (2013): Point of View and Point of Action – Eine Perspektive auf die Perspektive von Computerspielen. In: *RMFK. Repositorium Medienkulturforschung* 4, URL: <http://repositorium.medienkulturforschung.de/?p=360>.
- Ochs, Elinor (1979): Planned and unplanned discourse. In Givón, Talmy (ed.): *Syntax and semantics*. Vol 12: *Discourse and syntax*. New York: Acad. Press, 51–80.
- Pfaller, Robert (2008): *Ästhetik der Interpassivität*. Hamburg: Philo Fine Arts.
- Piirainen-Marsh, Arja (2012): Organizing participation in video gaming activities. In: Ayaß, Ruth/Gerhardt, Cornelia (eds.): *The appropriation of media in everyday life*. Philadelphia: Benjamins, 197–230.
- Rapp, Uri (1973): *Handeln und Zuschauen: Untersuchungen über den theatersoziologischen Aspekt in der menschlichen Interaktion*. Darmstadt: Luchterhand.
- Recktenwald, Daniel (2017): Toward a transcription and analysis of live streaming on Twitch. In: *Journal of Pragmatics* 115, 68–81.
- Reeves, Stuart/Greiffenhagen, Christian/Laurier, Eric (2016): Video gaming as practical accomplishment: Ethnomethodology, conversation analysis, and play. In: *Topics in Cognitive Science* 9 (2), 308–342.

- Rohrer, Tim (2007): The Body in Space: Dimensions of embodiment. In: Zlatev, Jordan/Ziemke, Tom/Frank, Roz/Dirven, René (eds.): *Body, Language and Mind*. Vol. 2. Berlin: de Gruyter, 339–378.
- Schemer-Reinhard, Timo (2016): Let's play without Controller. Zu den Effekten des Verschwindens der Steuerung im remediatisierten Game. In: Ackermann, Judith (eds.): *Phänomen Let's play-Video: Entstehung, Ästhetik, Aneignung und Faszination aufzeichneten Computerhandelns*. Wiesbaden: Springer VS, 55–83.
- Schmidt, Axel/Marx, Konstanze (2019): Multimodality as challenge: YouTube Data in Linguistic Corpora. In: Wildfeuer, Janina/Pflaeging, Jana/Bateman, John A./Seizov, Ognyan/Tseng, Chiao-I (eds.): *Multimodality. Disciplinary thoughts and the challenge of diversity*. Berlin: De Gruyter, 115–143.
- Schmidt, Axel/Marx, Konstanze (2020): Making Let's Plays watchable: An interactional approach to multimodality. In: Crispin Thurlow/Christa Dürscheid/Diémoz, Federica (eds.): *Visualizing (in) the New Media*. London: John Benjamins, 131–150.
- Schmidt, Axel/Marx, Konstanze/Neise, Isabell (2020): Produktion – Produkt – Rezeption? Medienketten in audiovisuellen Webformaten am Beispiel von Let's Plays. In: Marx, Konstanze/Lobin, Henning/Schmidt, Axel (eds.): *Deutsch in Sozialen Medien: interaktiv, multimodal, vielfältig*. Berlin: de Gruyter, 265–289.
- Selting, Magret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10, 353–402.
- Spagnolli, Anna/Gamberini, Luciano (2005): A Place for Presence. Understanding the Human Involvement in Mediated Interactive Environments. In: *PsychNology Journal* 3 (1), 6–15.
- Stivers, Tanya (2004): “No no no” and Other Types of Multiple Sayings in Social Interaction. In: *Human Communication Research* 30 (2), 260–293.
- Streeck, Jürgen (2008): *Gesturecraft. The manu-facture of meaning*. Amsterdam: Benjamins.
- Streeck, Jürgen/Goodwin, Charles/LeBaron, Curtis D. (2011): Embodied interaction in the material world: an introduction. In: Streeck, Jürgen/Goodwin, Charles/LeBaron, Curtis D. (eds.): *Embodied interaction - language and body in the material world*. New York: Cambridge Univ. Press, 1–26.
- Suchman, Lucy A. (1987): *Plans and situated actions: the problem of human-machine communication*. Cambridge: Cambridge Univ. Press.
- Tekin, Burak S./Reeves, Stuart (2017): Ways of spectating: unravelling spectator participation in Kinect play. In: *Proceedings of*

the 2017 CHI Conference on Human Factors in Computing Systems, 1558–1570.

Tolson, Andrew (2006): *Media talk: spoken discourse on TV and radio*. Edinburgh: Edinburgh Univ. Press.

Utterback, Camille (2004): Unusual positions – Embodied interaction with symbolic spaces. In: Wardrip-Fruin, Noah/Harrigan, Pat/Crumpton, Michael (eds.): *First Person. New Media as Story, Performance, and Game*. Cambridge: MIT Press, 218–235.

Wachsmuth, Ipke/Lenzen, Manuela/Knoblich, Günther (eds.) (2008): *Embodied communication in humans and machines*. Oxford: Oxford University Press.

Zhao, Shanyang (2005). The Digital Self: Through the Looking Glass of Telecopresent Others. In: *Symbolic Interaction* 28 (3), 387–405.

Video Games

Red Barrels (2017): *Outlast 2*.

Starbreeze Studios (2016): *Dead by Daylight*.