Originally published in: Thurlow, Crispin/Dürscheid, Christa/Diémoz, Federica (eds.): Visualizing digital discourse. Interactional, institutional and ideological perspectives.- Berlin [u.a.]: De Gruyter Mouton, 2020. Pp. 131-150

Axel Schmidt and Konstanze Marx

6 Making Let's plays watchable: An interactional approach to gaming visualizations

1 Introduction: The watchability of Let's Plays

Entering the keyword 'Let's Play' on YouTube results in over 170 million hits. This is a genre that has become very popular in a very short period of time. Here, gamers document their gaming in films, so called Let's Plays, and tens of thousands of other people watch via YouTube. What is interesting about this relatively new phenomenon is that video games are not only played, but the playing is presented to a (potential) mass audience where viewers do not play the game but watch others do. Normally, the interactivity of a video game is one of its most outstanding, appealing features. In Let's Play, however, this interactivity can only be experienced vicariously by the audience. For this reason, Let's Players must support their audience's vicarious experience by strategically commenting verbally on their play moves and by deploying specific visual presentation techniques. In this chapter, we are interested in how the core activity of playing a computer game is transformed into a "platform format" (Goffman 1983) by verbal, embodied and visual means. In short, we ask what practices players use to make Let's Play watchable?

A prevalent practice used by players on Let's Play is to continuously comment verbally on their game moves. If game activities are accountable in their own right (see section 3), one can ask, why they have to be explained or commented on verbally in addition to the visible action on screen. Our thesis is as follows. On the one hand, continuous commenting fulfills the main objective of Let's Plays which is to present active gameplay. Comments should therefore generate an added value, not simply making gameplay visually accessible. This itself indicates how the presentation of the game alone is not sufficient. Often – as we will show – a player's own game actions are explicated verbally which makes the game play more transparent as an action and thus more attractive for viewers. In other words, how players make Let's Plays watchable for viewers. It is our contention that the watchability of video gaming depends to a large extent on the degree of viewers' insights into the players's motivations for

https://doi.org/10.1515/9781501510113-007

Publikationsserver des Leibniz-Instituts für Deutsche Sprache
URN: http://nbn-resolving.de/urn:nbn:de:bsz:mh39-96595

action and the player's experience of the game.[1] To this end, we examine a case-study example of a Let's Play with respect to its player's linguistic and embodied practices and the ways in which there (self-experienced) interactivity of video games is made accessible to others. It is in this way that media formats (i.e. video games) not primarily designed for watching, are transformed into watchable presentations.

In the sections which follow we give a brief overview of Let's Plays and their typical characteristics (section 2) followed by an outline of our data and method (section 3). Then, in section 4, we examine some of the key strategies of players with reference to selected extracts from our Let's Play case-study. In short, it becomes clear that the central practice in Let's Plays – that is, formulating one's own acts – is closely related to the specific kind of multimodal nature and form of visual presentation of Let's Plays.

2 What are Let's Plays?

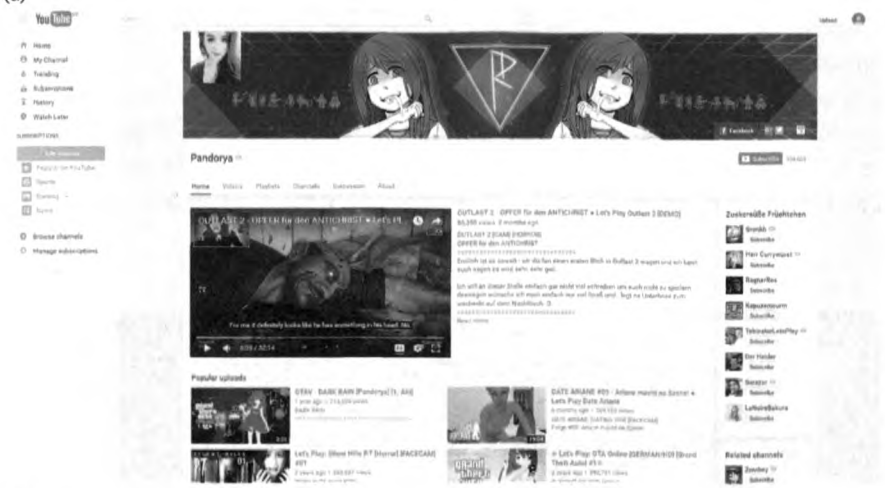
Let's Plays are defined simply as "playing videogames for the internet" (Hale 2013: 3). However, as we say, the game is not only played but also commented on verbally by players who usually appear in a facecam. In addition, both the playing process and any verbal comments are recorded and uploaded to video hosting websites like YouTube. The following stills show a Let's Play embedded in YouTube (Figure 6.1a) and in full screen mode (Figure 6.1b):

Let's Plays first appeared in 2006 and nowadays have very high "click rates". The German player *Gronkh* for instance has about 4.5 million subscribers, while the world's most popular player, *PewDiePie*, has over 60 million subscribers.

Previous studies of video-gaming show that players and their spectators used to talk with each other, commenting on game play while playing the game (cf. Baldauf-Quilliatre & Colon de Carvajal 2019; Piirainen-Marsh 2012; Tekin and Reeves 2017). Furthermore, phases "in the game" and "out of the game" are ordinarily distinguishable and lead to different concepts of time ("game time" vs. "interaction time") and, in turn, different forms of involvement (cf. Mondada 2013). By contrast, Let's Plays adopt a different participation framework. Instead of a shared physical setting where every co-present person

1 In contrast to other well-known games like chess, many video games are self-generated and therefore particularly opaque to observers.

(a)



(b)

Figure 6.1: (a) Pan embedded in YouTube. (b) Pan in full-screen mode.

is able to participate, Let's Plays entail a unidirectional form of communication that excludes the viewer from direct feedback.[2]

2 We refer here to recorded Let's Plays uploaded to video portals like YouTube. Such Let's Plays have the status of films (they are "fixed products") and, as uploads, they are

Before we go into detail, we give an impressionistic overview of the main characteristics of Let's Plays with reference to selected moments in a Let's Play uploaded on YouTube on the 15th of June 2016 by a famous German Let's Player called *Pandorya* (nickname: *Pan*). In this case, Pan is playing a demo version of the survival horror game *Outlast 2*.[3] At the beginning, *Pan* refers to an audience by uttering "*a warm welcome to Outlast 2*" accompanied by a look into the facecam (see Figure 6.2), all reminiscent of forms of para-social interaction (cf. Horton & Wohl 1956, Ayaß 1993).

Figure 6.2: Pan's audience address.

Also right at the beginning, Pan emphasizes her individual experience and general attitude by categorizing herself (ironically) as a "*miserable fan-girl*". During the whole Let's Play she shows many affective reactions conveyed visually by means of the integrated facecam (e.g. scared, as in Figure 6.3).

permanently available to a dispersed, potentially mass audience. Due to the affordances of the medium, feedback from viewers can only be given in written form afterwards (cf. Dynel 2014, Frobenius 2014). Such commentary activities are not included in this article. We also do not consider the opportunity used by many players to distribute their Let's Plays live (e.g. on *Twitch*), where viewers can chat with them during the game (cf. Recktenwald 2017).

3 The actual Let's Play discussed here can be found (21.07.2019) at: https://www.youtube.com/watch?v=P9StHBpbnlA&t=186s.

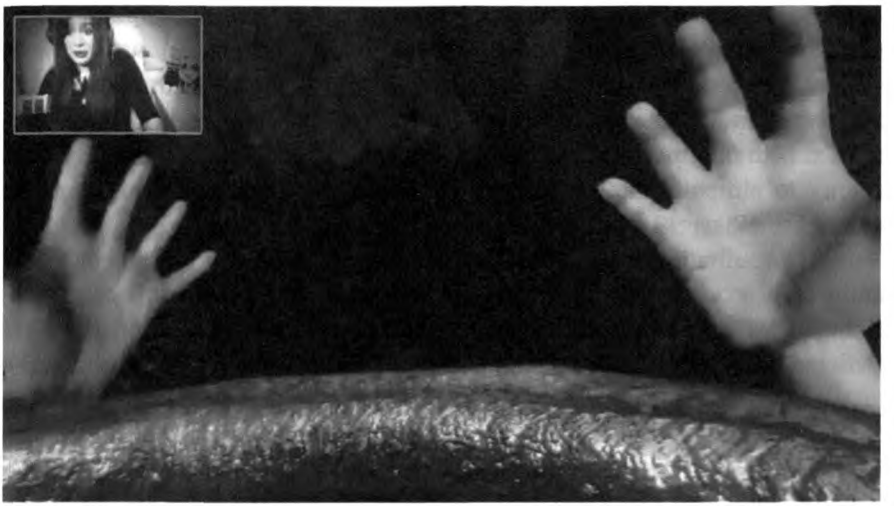


Figure 6.3: Frightening.

Overall, her verbal moderation of the game play uses a spontaneous language and takes place partly as an ad hoc reaction to events emerging in the game; there is obviously no post-editing. Finally, and interestingly, Pan problematizes the moderation itself: about 13 minutes after the start of the game, she says, "*hey, how do you present it now?*"[4] Evidently, the need to make one's own playing attractive to others is also a concern for players themselves (and not only for analysts like us). A key reason for this is that Let's Plays are recorded and thus fixed products. As such, targeted viewers can no longer play the game, but only watch how others play it. The game itself unavoidably loses its interactivity therefore. For players, thus, the challenge is to make watching the game attractive – in other words, to make it watchable.

3 Data and method

In order to demonstrate core strategies and practices of making Let's Plays watchable we draw on the same case already introduced. This is an example of a so-called *blind play* (i.e. the game has not been played before by this player)

4 In the German original Pan uses the expression *moderieren* which explicitly signals broadcasting contexts.

by the popular German player Pan who presents a current computer game from the *adventure-action*-genre. The combination of *blind play* and the *adventure/action* genre promises situations that are potentially unpredictable, surprising and in the need of explanation since players are typically confronted with obscure settings. Selecting a popular Let's Play player offers us a chance to pin-point more typical, well-established practices in this community of Let's Player.

Our contribution follows an interactional perspective, or more precisely a multimodal extended EMCA approach, which asks *how* participants create social reality. EMCA stands for Ethnomethodology and Conversation Analysis (cf. Heritage 1984a). By multimodal expansion, we mean attention is paid not only to talk, but also to embodied actions, including the use of objects, of media technology and of space (cf. Deppermnann 2013, Mondada 2008, Streeck/Goodwin/LeBaron 2011). Specifically, our analytic focus is on the strategies or practices by which players design their Let's Plays to be attractive, engaging and therefore watchable. EMCA research on mediated interaction in general (cf. Arminen, Licoppe, & Spagnolli 2016, Schmidt & Marx 2017) and video games in particular (cf. Reeves, Greiffenhagen, and Laurier 2016 for an overview) has emphasized that the mediation of a given interaction is not determined by technical circumstances alone. Rather, mediation is taken to be an interactional achievement. Within media settings, and following Arminen, Licoppe & Spagnolli (2016), participants' use of technology also reflects their notion of a reasonable *accountable* handling of the technology. In short, the way practical problems are solved in mediated settings is related to the manner in which interaction is shaped by the specific affordances of the technology. For example, video players in co-present multiplayer settings only turn their bodies halfway around in order to react to other players in their backs thereby at the same time being able to turn back quickly to their own running video game. Through the half turn of the body, what Schegloff (1998) coined "body torque", players simultaneously display their temporary double involvement, which makes their action accountable for others and at the same time reflects the specifics of the media setting.

According to Reeves, Greiffenhagen, and Laurier (2016) there are thus *two levels of accountability* in video game settings: First "real world actions" like the above mentioned body orientations within the game setting, but also steering the game and possible verbal conduct (concerning the game play or anything else). Secondly and additionally in comparison to face-to-face-settings, there are (inter-)actions in the virtual world conducted by player-driven avatars. The latter is illustrated by Reeves, Greiffenhagen, and Laurier (2016) using the case of a cooperative multiplayer game in which one of the avatars kneels

down. This visible action is read by the (competent) co-players as a display that the player's avatar is about to set a trap. They, in turn, display their understanding by embodied actions of their avatars (in this case they form a semicircle behind and around the trapper). In this way, the game itself produces socially organized and comprehensible interactions based on the control actions of the players.

In order to reconstruct how participants produce Let's Plays, we use multimodally extended GAT2-transcripts that show the use of different modal resources in their temporal interplay.[5] In the transcripts (see extract 6.1 and Figure 6.4) speech and pauses appear bold and provide the temporal framework for aligning physical activities, game events (after the abbreviation GE), game sounds (after the abbreviation GS) and status displays (after the abbreviation SD). The alignment is done with special characters (%, *, +, etc.), one symbol indicates the beginning (%), a double symbol (%%) the end and the double-headed arrow (—>>) a continuation beyond the transcript of an activity/event.

Figure 6.4: Relevant events for transcription.

5 See Selting et al. (2011) for GAT 2-conventions; Mondada (2014) for conventions of multimodal transcription; and Recktenwald (2017) for the transcription of LP- and Twitch-formats. A selection of the GAT 2-conventions we use is provided below.

Extract 6.1: A transcript example

```
1   PAN   und äh wir können die *%+KAmer~a%% be~~nutzen?
          and äh we can use the camera
                                  %nods
    GE                          *camera use--->>
    SD                             +camera symbol--->>
    GS                                      ~bleeping
```

4 Making Let's Plays watchable

In this section, we present three key practices, two talk-based and a visual one, used by players in the overall attempt of making Let's Play engaging and, therefore, watchable for viewers; we do so with reference to the same case-study example of Pan playing the demo version of the survival horror game *Outlast 2*.

4.1 Commenting on self-generated actions in Let's Plays

We start with a simple example to show how moderation works in principle. In the following extract, *Pan* uses the integrated camera function at the very beginning of her game play:

Extract 6.2: "We can use the camera"

```
1   PAN     und äh wir können die *KAmera benutzen?#
            and uh we can use the camera
    GE/DS                          *camera use/camera symbol--->>
2   PAN     wir KENN_das ja in outlast–
            we already know that from outlast
3   PAN     wir *können wieder** RANzoomen,
            we can zoom in again
    GE         *zoom in/zoom out
4   PAN     und ich kann *natürlich auch die NIGHTvision anmachen?
            and of course I can also put on the night vision
    GE                         *night vision -->>
```

For now, we focus on just the visual impressions in order to highlight the particular use of verbal comments. First of all, potentially relevant objects of perception are attended to; for example, a landscape, a fence, a wind pump and a house in the distance as well as conditions of perception, here especially the

relative darkness, are noticeable. In addition, the image frame moves, it gets bigger and smaller, changes color etc. (see Figure 6.5a-c).

(a)

Figure 6.5a: Objects, conditions of perception.

(b)

Figure 6.5b: Zoom.

(c)

Figure 6.5c: Night Vision.

Depending on the player's background knowledge, game features can be recognized (for instance point-of-view-perspective, in-game-camera etc.). Altogether, however, specific action context is not easily inferred, except perhaps that someone might be looking for something. Even then, viewers would not easily discern what the exact focus or purpose is. Accordingly, what viewers get to see remains inaccessible and therefore less attractive.

This changes completely when we look at the same extract in conjunction with its verbal comments. Viewers learn *what* Pan is doing (that she uses the camera and its features); more than this, however, they also learn what her *purpose* it is: she uses the camera to *demonstrate* her functions and is not looking for anything as one might have otherwise imagined. This framing of the sequence as a demonstration is not obvious from the images alone – regardless of how much expert knowledge one might have about the game or about video games in general. In addition, the "demonstration frame" (cf. Goffman 1986, 66) shifts attention away from the visual content to the way it is technically captured and conveyed by using and demonstrating functions of the game-integrated camera. As such, the focus is not on the events in front of the camera, but about the camera action itself which is made (also by means of accompanying words)

accountable in a specific way.[6] The sequence is thereby segmented into three units concerning technical control over the game: camera, zoom, and night vision.

Furthermore, as the above case illustrates, another moderator effect is that the player identifies certain visual events as *self-generated results of their own actions*. Figure 6.6 demonstrates this relation. Since the deployment of the simulated camera is done during Pan's utterance "*and, uh, we can use the camera*" the visually conveyed game event becomes understandable as *her* action.[7]

und äh wir können die Kamera# benutzen
and, uh, we can use the camera

Figure 6.6: Self-generated events.

This raises two fundamental questions. First, which activity forms the basis for Let's Plays? Second, how is this activity presented visually?

First of all, the core activity in Let's Plays (i.e. computer gaming) is a form of human-machine communication usually described as a *cybernetic control loop*, as illustrated in the following Figure (6.7).

6 See also Laurier & Reeves (2014) who discuss the accountability of the camera use in different video games.

7 The hash (#) in our transcripts defines the exact location where a still image was drawn.

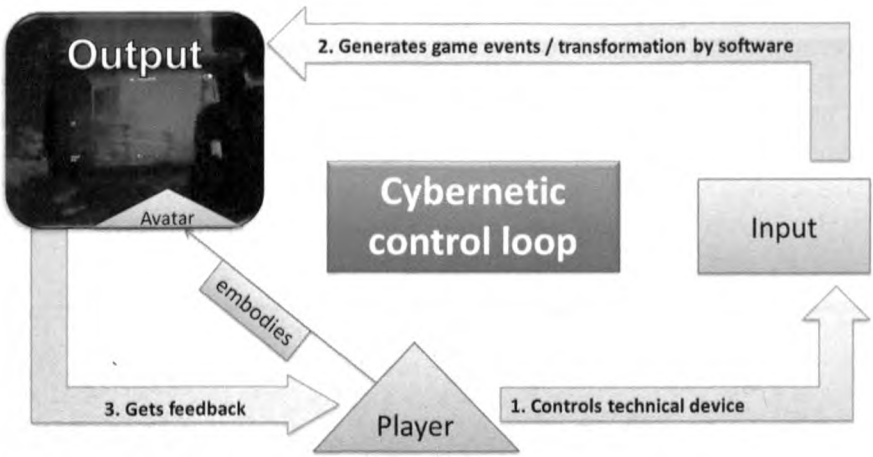


Figure 6.7: The cybernetic control loop of gaming.

As Figure 6.7 shows, playing a computer game entails generating game events by controlling a technical device which transforms software-based inputs into on-screen outputs whereby the latter serves as a feedback for the player. Since inputs are controlled and adjusted on the basis of outputs, gaming follows the model of a cybernetic control loop. Together with the audio-visual point-of-view-aesthetic, the technical ability to effect or cause events is fundamental for embodying one's avatar in a virtual world.

4.2 Visualization of self-generated actions in Let's Plays

The *visual presentation* of activities in Let's Plays rearranges the physical-spatial configuration of the real game situation which is usually body-/face-to-screen and hands-on-controller. First of all, the activity of controlling the game is not represented, so hands–on-controller does not usually appear in Let's Plays. Instead, via facecams, players' bodies are reduced to the face which appears smaller at the edge of the screen; by contrast, game play fills the entire screen and is thus marked as the main focus. As a result, player and gameplay appear simultaneously in a split-screen-optics spatially next to each other. Mondada (2009) has described the use of split-screen as a re-arrangement of interactional space. In our case, the use of the facecam together with the split screen makes players appear more like recipients than controllers of the game. Nevertheless, the control loop clearly remains crucial since the person who appears in the facecam generates, perceives and reacts to what is happening in the game (see Figure 6.8).

Figure 6.8: Visual representation of self-generated events.

It is precisely this self-generation of the gameplay that makes video games appear/feel attractive, interactive and immersive (cf. Calleja 2011). However, because this core aspect is missing when the gameplay is presented in Let's Plays, an additional orientation is needed for giving viewers an insight into the cycle of self-control. One solution to this problem is very frequent verbal references by players to their own actions. That means, players are not only playing but also simultaneously *formulating* their game actions, as in the example above where Pan, while using the camera, also formulates her actions verbally. Following Schegloff's (1972) "formulating place" and Goodwin & Goodwin's (1996) "formulating planes", we call this strategy "formulating gameplay". A particular problematic action in terms of comprehensibility for viewers is *seeing* or *visual perception*. Although seeing is an indispensable activity of playing a video game, seeing is difficult to convey for viewers on the basis of the images alone. Visual impressions alone never reveal whether seeing is a relevant activity at any moment – as opposed to some other activity like walking, fighting, opening a door, etc.. If it is relevant, nor is it clear what the focus is, how it is to be interpreted and how it leads to or is connected to subsequent actions.[8] Players therefore face the fundamental challenge of making their perception perceptible to viewers. This brings us to the third (and last) of our three practices for making Let's Plays watchable.

8 Goodwin (1994, 1995, 1996, 2000), in particular, has attended to "seeing as an activity".

4.3 Formulating perception in Let's Plays

In a further example we take a closer look at the ways processes of visual perception are formulated. In the following extract Pan explores the entrance area of a wooden house:

Extract 6.3: "Delicious"

1	PAN	***was ham wir HIER–**
		what do we have here
	GE	*hardly perceptiple game object
2	PAN	**=Ä&ähä:;**
	GE/DSSD	&camera use/camera symbol
3		**%(0.6)%%(0.53)*(1.00)&**
	pan	%looks down to the right
	GE	*camera focus on game object
	GE/SD	night vision/symbols
4	PAN	**°h LECker,**
		°h yummi
5	PAN	**&%ich HAbe es ver*MISST;**
		I have really missed it
	GE/SD	&end of camera use/camera symbol disappears--->>
	pan	%looks to facecam
	GE	*focus on door
6	PAN	**ich LIE+be ++outlast.**
		I love outlast
		+looks to facecam

Here, we are most interested in three aspects of this short extract: *visual impressions*, the role of *deixis*, and the verbally expressed *perspective* of Pan's visual perception.

Referring only to visual impressions, the viewer has two clues for reconstructing Pan's perception. First, the facecam conveys the direction of her *gaze*. For example, in Figure 6.9 she is seen looking straight down (6.9a), diagonally downwards (6.9b,) and straight into the facecam (6.9c).

It is not possible to know exactly what Pan is looking at; only her looking into the facecam is easily understood as an audience address.[9] Secondly, Pan's point-of-view-perspective conveys what she is visually focusing on in the game.

9 To discover the real objects and intentions of Pan's gaze we would need access to the physical setting. From interviews, however, we know that a gaze straight downwards is directed at events on screen whereas a gaze diagonally downwards is directed to status displays. In Figure 6.9a, therefore, the gaze straight downwards is understandable as game gaze (cf. Aarseth 2004, Atkins 2006) and thus as the default mode for someone involved in the game.

Figure 6.9a: Gaze direction diagonally downwards.

Figure 6.9b: Gaze direction straight downwards.

Figure 6.9c: Gaze direction in facecam.

Hence, the player sees what the avatar sees and viewers see what the player-avatar-hybrid sees (in this case, a kind of bucket).[10] This is enhanced by the viewfinder of the camera indicating what is currently focused by a grey square inserted in the display (see Figure 6.10).

10 See also Baldauf-Quilliatre & Colon de Carvajal (2015).

Figure 6.10: We see what is seen.

The images thereby simulate the "perception of perception" – in other words, giving a sense of what another person is noticing or looking at.

As we say, images alone are apparently never sufficient. To further mark seeing *as* seeing players rely on the verbal resource of *deixis*. In the example above, Pan makes her perception a relevant action in line 1 by uttering "*what do we have here*". These exploratory announcements are very frequent in Let's Plays and they indicate that visual perception is currently relevant. In particular, this is achieved by the local deictic "*here*" which serves as a cue for specifying the context produced by *here* (cf. Hausendorf 2003). In face-to-face interaction this is usually done by means of pointing gestures (cf. Goodwin 2007); however, these are clearly not available in this mediated instance. Instead, and as we have argued above, viewers rely on the point-of-view-perspective for discerning the player's vision. Thus, the point-of-view-vision acts as a pointing instrument which indicates what Goodwin (1994) calls a "domain of scrutiny".

Exploratory announcements like "*what do I/we have here*" are usually fulfilled subsequently by formulating visual impressions. Those verbalizations again specify the perspective of perception. What is striking about the current case is that even though an announcement of exploration makes the identification of a scrutinized object pertinent, there is no such investigation. That is, no referent is verbally specified such as "a bucket with guts". This leads to a sort of mystery which itself creates tension and contributes to making Let's Plays

engaging and watchable. Instead, what Pan reveals verbally is reduced to assessments. First, in line 2, we have a response cry (Goffman, 1981) – *ähhhh* – for conveying disgust and which reflects her immediate experience. This is followed in line 4 by a positive evaluation term (*delicious*), accompanied by a smile evoking the sensual qualities of the object in question. *Delicious* is also presented in direct contrast to disgust. Although the choice of words and prosody suggest irony, subsequent events show how Pan is concerned also with the aesthetics of the game. In this regard, note how she no longer refers to the concrete object of scrutiny ("the bucket") with the pronoun *it* in "*I missed it*" (line 5), but refers rather generically to a type concept – namely the look or design of the game *Outlast*. The visual impression she stresses thereby becomes a trademark for the game itself. This is clearly shown in the euphoric evaluation of the game in line 6 of the extract ("*I love Outlast*"). Only through the accompanying words do viewers have access to Pan's personal perspective on the events unfolding visually, which are indicated as a typical feature of a game-specific aesthetic.

5 Summary

We started with the claim that Let's Plays transform an interactive video game into a fixed, broadcast-like product. This process can be understood as "de-interactivisation" (cf. Ackermann 2016). The playing of the game itself, however, remains interactive. As we have argued, the basic activity structure of Let's Plays is a "cybernetic control loop" which cannot be readily made accessible by drawing only on visual presentations. Thus, the main problem for players consists of how to make this "black box" accessible for viewers. Based on two typical cases we have shown how the players in Let's Plays deal with this problem. A pervasive practice throughout Let's Plays is to formulate one's own actions.[11] Thereby players make the gameplay comprehensible and at the same time watchable. This applies in particular – as we have shown – to formulations of visual perception, which make the self-organizing control loop of playing a video game transparent by creating a connection between prospective relevancies (like the above mentioned exploratory announcements) and retrospective evaluations of actions (like response cries or explicit verbalizations such as

11 There are other settings in which participants formulate their actions; for example, when think-aloud-methods are used in writing research (cf. Marx & Schmidt 2019 for a discussion of this aspect).

delicious). This relation cannot be conveyed solely through the visuals. Only the verbal comments are able to transform a stream of visual events into comprehensible actions. Let's Plays are thus an interesting case of re-mediated visualizations (cf. Bolter and Grusin 2000), which needs additional verbalization to be attractive for viewers. In addition, our example analysis has shown how a fully multimodal practice (i.e. Let's Plays) is visualized in specific ways and, thereby, how communicative processes are embedded. This way, visualization not only fulfils representational purposes (i.e. showing something), but is also used to simulate core processes of multimodal interaction such as using the in-game camera to indicate the direction of a player's gaze.

GAT 2 Transcription conventions (selection)

°h / h°	in- / outbreaths
(.)	micro pause
and_uh	cliticizations within units
uh, uhm	hesitation markers
=	fast, immediate continuation with a new turn or segment (latching)
:	lengthening
SYLlable	focus accent

Final pitch movements of intonation phrases

?	rising to high
,	rising to mid
–	level
;	falling to mid
.	falling to low

References

Aarseth, Espen. 2004. Genre trouble: narrativism and the art of simulation. In Noah Wardrip-Fruin & Pat Harrigan (ed.), *First person. New media as story, performance, and game*, 45–57. Cambridge: MIT Press.

Ackermann, Judith (ed.). 2016. *Phänomen Let's Play-Video: Entstehung, Ästhetik, Aneignung und Faszination aufgezeichneten Computerhandelns. Neue Perspektiven der Medienästhetik*. Wiesbaden: Springer VS.

Arminen, Ilkka, Christian Licoppe & Anna Spagnolli. 2016. Respecifying mediated interaction. *Research on Language and Social Interaction* 49 (4), 290–309.

Atkins, Barry. 2006. What are we really looking at? The future-orientation of video game play. *Games Cult* 1 (2), 127–140.

Ayaß, Ruth. 1993. Auf der Suche nach dem verlorenen Zuschauer. In Werner Holly & Ulrich Püschel (eds.), *Medienrezeption als Aneignung*, 27–41. Opladen: Westdeutscher Verlag.

Baldauf-Quilliatre, Heike & Isabel Colón de Carvajal. 2015. Is the avatar considered as a participant by the players? A conversational analysis of multi-player videogames interactions. *PsychNology Journal* 13 (2–3), 127–147.

Baldauf-Quilliatre, Heike & Isabel Colón de Carvajal. 2019. Interaktionen bei Videospiel-Sessions: Interagieren in einem hybriden Raum. In Konstanze Marx & Axel Schmidt (eds.), *Interaktion und Medien*, 219–254. Heidelberg: Winter.

Bolter, Jay David & Richard Grusin. 2000. *Remediation: understanding new media*. Cambridge: MIT Press.

Calleja, Gordon. 2011. *In-game: from immersion to incorporation*. Cambridge: MIT Press.

Deppermann, Arnulf. 2013. Multimodal interaction from a conversation analytic perspective. *Journal of Pragmatics* 46, 1–7.

Dynel, Marta. 2014. Participation framework underlying YouTube interaction. *Journal of Pragmatics* 73, 37–52.

Frobenius, Maximiliane. 2014. Audience design in monologues: How vloggers involve their viewers. *Journal of Pragmatics* 72, 59–72.

Goffman, Ervin. 1981. Response cries. In Ervin Goffman (ed.), *Forms of talk*, 78–122. Philadelphia: University of Pennsylvania Press.

Goffman, Erving. 1983. The interaction order. *American Sociolocical Review* 48 (1), 1–17.

Goffman, Erving. 1986. *Frame analysis*. Boston: Northeastern Univ. Pr.

Goodwin, Charles. 1994. Professional vision. *American Anthropologist* 96 (3), 606–633.

Goodwin, Charles. 1995. Seeing in depth. *Social Studies of Science* 25, 237–274.

Goodwin, Charles. 1996. Transparent vision. In Elinor Ochs, Emanuel Schegloff & Sandra A. Thompson (eds.), *Interaction and grammar*, 370–404. Cambridge: Cambridge University Press.

Goodwin, Charles. 2000. Practices of seeing visual analysis: An ethnomethodological approach. In Theo van Leeuwen & Carey Jewitt (eds.), *Handbook of visual analysis*, 157–182. London: Sage.

Goodwin, Charles. 2007. Enviromentally coupled gestures. In: Susan D. Duncan, Justine Cassell & Elena T. Levy (eds.), *Gesture and the dynamic dimension of language: essays in honor of David McNeill*, 195–212. Amsterdam: John Benjamins.

Goodwin, Charles & Marjorie Harness Goodwin. 1996. Seeing as situated activity: Formulating planes. In Yrjo Engeström & David Middleton (ed.), *Cognition and communication at work*, 61–95. Cambridge: Cambridge University Press.

Hale, Thomas. 2013. From jackasses to superstars: A case for the study of 'Let's Play'. Available online at: https://www.academia.edu/5260639/From_Jackasses_to_Superstars_A_Case_for_the_Study_of_Let_s_Play_September_2013_ (21 February, 2019).

Hausendorf, Heiko. 2003. Deixis and speech situation revisited: the mechanism of perceived perception. In Friedrich Lenz (ed.), *Deictic conceptualisation of space, time and person*, 249–269. Amsterdam, Philadelphia: John Benjamins.

Heritage, John. 1984. *Garfinkel and ethnomethodology*. Cambridge: Polity Press.

Horton, Donald & Richard R. Wohl. 1956. Mass Communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry* 19, 215–229.

Laurier, Eric & Stuart Reeves. 2014. Cameras in video games: Comparing play in Counter-Strike and Doctor Who Adventures. In Mathias Broth, Eric Laurier & Lorenza Mondada (ed.), *Studies of video practices: Video at work*, 181–207. Hoboken: Taylor and Francis.

Marx, Konstanze & Axel Schmidt. 2019. Let's Play (together) oder schau mal, wie ich spiele – (Interaktive) Praktiken der Attraktionssteigerung auf YouTube. In Konstanze Marx & Axel Schmidt (eds.), *Interaktion und Medien*, 319–352 Heidelberg: Winter.

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. 2009. Video recording practices and the reflexive constitution of the interactional order: Some systematic uses of the split-screen technique. *Human Studies* 32 (1), 67–99.

Mondada, Lorenza. 2012. Coordinating action and talk-in-interaction in and out of video games. In Ruth Ayaß & Cornelia Gerhardt (eds.), *The appropriation of media in everyday life*. pp. 231–270. Philadelphia: John Benjamins.

Mondada, Lorenza. 2013. Coordinating mobile action in real time: The timely organization of directives in video games. In Pentti Haddington, Lorenza Mondada & Maurice Nevile (eds.), *Interaction and mobility: language and the body in motion*, 300–341. Berlin; New York: de Gruyter.

Mondada, Lorenza. 2014. Conventions for multimodal transcription. Available: https://franzoesistik.philhist.unibas.ch/fileadmin/user_upload/franzoesistik/mondada_multimodal_conventions.pdf (26 March, 2018)

Piirainen-Marsh, Arja. 2012. Organizing participation in video gaming activities. In Ruth Ayaß & Cornelia Gerhardt (eds.), *The appropriation of media in everyday life*, 197–230. Philadelphia: John Benjamins.

Recktenwald, Daniel. 2017. Toward a transcription and analysis of live streaming on Twitch. *Journal of Pragmatics* 115, 68–81.

Reeves, Stuart, Christian Greiffenhagen & Eric Laurier. 2016. Video gaming as practical accomplishment: Ethnomethodology, conversation analysis, and play. *Topics in Cognitive Science* 9 (2), 308–342.

Schegloff, Emanuel A. 1972. Notes on a conversational practice: Formulating place. In David Sudnow (ed.), *Studies in social interaction*, 95–119. New York: Free Press.

Schmidt, Axel & Konstanze Marx. 2017. Interaktion und Medien. *Sprachreport* 33, 4, 22–33.

Selting, Margret, Peter Auer, Dagmar Barth-Weingarten, Jörg Bergmann, Pia Bergmann, Karin Birkner, Elizabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzlufft, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock & Susanne Uhmann. 2011. A system for transcribing talk-in-interaction: GAT 2. [trans. by Elizabeth Couper-Kuhlen & Dagmar Barth-Weingarten.] *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* (www.gespraechsforschung-ozs.de) 12, 1–51.

Streeck, Jürgen, Charles Goodwin & Curtis D. LeBaron. (eds.). 2011. *Embodied interaction. Language and body in the material world*. New York [u.a.]: Cambridge University Press.

Tekin, Burak S. & Stuart Reeves. 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.