Article:
Reed, Darren James (2024) Turning Heads and Making Conversation on Twitch. Discourse, Context, and Media. 100802. ISSN: 2211-6958

https://doi.org/10.1016/j.dcm.2024.100802

Discourse, Context & Media 60 (2024) 100802

Contents lists available at ScienceDirect

Discourse, Context & Media

journal homepage: www.elsevier.com/locate/dcm

Turning heads and making conversation on Twitch

Darren J. Reed

Law and Management Building, University of York, UK

ARTICLE INFO

Keywords:
Conversation analysis
Twitch
Multimodal technology
Embodied interaction

ABSTRACT

This paper undertakes an analysis of the streaming platform Twitch. Twitch is premised upon a single Streamer and a large audience. Interaction between a Streamer and an individual Chat member is constructed by the Streamer through a form of acknowledgement and ratification that leads to a momentary encounter. Strategic 'reading' and reading aloud not only function to link and sustain conversation, they also structure the embodied performance of the streamer, who must shift their visual attention away from the primary activity. These head movements are then a primary resource for interaction.Through a Conversation Analytic approach, the paper identifies a 'reading-reading aloud-responding' action or R(RA)R as an interactional resource for engaging audience members in ongoing conversation. The analysis is in line with the 'respecification' of 'mediated interaction' (Arminen et al., 2016) and moves towards a 'digital CA' (Giles et al., 2015). It contributes to an ongoing project that looks to 'digitise Sacks' (Housley et al., 2017) and inform methodological development of future analysis of technology-mediated interaction.

1. Introduction

As more of our social life occurs online there is a need to understand the ways that collective engagement is established and maintained. Asking how a 'community' might be established and maintained centres our attention on a sociological understanding of contemporary forms of technological communication. A pressing concern is the manner in which behaviours through such technologies engender interaction.

One communication technology that appears to engender collective engagement is Twitch, a platform oriented to the mass viewing and engagement with individual 'streamers'.[1] In August of 2022, Twitch had 813 billion minutes of content viewed to that point in the year, with an average of 100 thousand concurrent streamers and an average viewership 2.7 million (twitchtracker.com). A large audience, and any notion of community, is sustained through successful interaction between streamers and audience members.

Twitch engagement is achieved through interaction strategies deployed by individual streamers. One strategy is the reading aloud of text-chat messages. To achieve this, a streamer must turn their gaze away from the main computer monitor towards a second monitor on which chat messages are displayed. This simple behaviour turns out to be crucial in understanding 'interaction' as both achieved and observed. Being seen to interact is a key component of community dynamics. This paper is concerned therefore with embodied movements ('head turns') as afforded by the technological features within the spatial environment of the streamer.

An important point is that 'interaction' in Twitch is already recorded and archived and does not require further technology to accomplish this. It is therefore a naturalistic laboratory, or 'perspicuous setting' (Garfinkel, 2002: 182), for understanding these changes in relation to the competences and understandings of human actors and the affordances of technology, thereby revealing the changes brought about to the "interaction order" (Goffman, 1983) by technology (Housley et al., 2017) (see also Arminen et al., 2016).

This paper will use the term 'multimodal' to refer to features of the technology (video, audio, text) and 'embodied behaviour' to refer to the talk and movement of actors (speaking, gesturing, head turns, object manipulation). The separation allows for an account of behaviours afforded and constrained by the different technologies available to different actors (streamer/audience member/chat participant/analyst).

Through embodied Conversation Analysis (CA) this paper identifies one embodied interaction phenomenon – the 'reading-reading aloud-responding' strategy or R(RA)R. This is premised upon the 'reading-aloud and responding' (RAR) action sequence discovered by Licoppe and Morel (2018) in their analysis of the mobile streaming application Periscope. While there are similarities between Periscope and Twitch, and the RAR sequence is present in both, the technological and social contexts of Twitch provide for additional analytical insights. The

E-mail address: Darren.reed@york.ac.uk.

[1] The word 'streamer' is used to indicate the generic category or role. A capitalised version ('Streamer') is used for the specific streamer in the analysis.

https://doi.org/10.1016/j.dcm.2024.100802
Received 20 September 2023; Received in revised form 27 June 2024; Accepted 14 July 2024
Available online 27 July 2024

Fig. 1. The twitch interface.

interactions on Twitch are between a streamer and a large established distributed audience. An important aspect of the streamer's role, therefore, is to maintain the continued engagement of this audience. Twitch is also different to Periscope because it involves different technologies situated in different ways. This point will be returned to in the discussion section.

2. Background

There is growing interest in researching Twitch as a social phenomenon. This includes research into user motivation (Chen and Lin, 2018; Hilvert-Bruce et al., 2018; Xu et al., 2021; Zhao et al., 2018) and the uses and gratifications of the platform (Cabeza-Ramírez et al., 2020; Dux, 2018; Gros et al., 2017; Sjöblom et al., 2017). Play and the performative qualities of the medium has motivated a range of studies (Lessel et al., 2017; Paz and Montardo, 2018; Scully-Blaker et al., 2017) including gender performance (Dargonaki, 2018; Zhang and Hjorth, 2019) and the different roles involved (Wohn, 2019; Woodcock and Johnson, 2019). Research includes questions about the relationship between Twitch participation and the traditional television viewer (Spilker et al., 2020). Twitch study extends from studies of 'esports' and, before that, computer game-play (Taylor, 2018). This is understandable, as much of the activity on Twitch involves playing computer games in front of others. More recently other types of content has increased, with streams dedicated to everything from live coding to music creation.

The naming of Twitch underlines its fundamental embodied character. The term'twitch gameplay' references the rapid body movements of early computer games (Zamora, 1981: 24). Anderson (2017) identifies a 'corporeal turn' in computer games research and extends it to social arrangements beyond the screen to include the 'offline' culture of computer game audiences. This provides a valuable reminder of the 'situatedness' (Mondada, 2011; Pink and Mackley, 2013; Suchman, 1987) of Twitch participation and interaction in terms of technology and the material contexts of use of that technology. To incorporate this situatedness, the following insights are generated through 'embodied' conversation analysis.

Conversation Analysis was originally premised upon the analysis of talk (Sacks et al., 1974). As the approach developed, analysis emerged that combined talk with embodied behaviours such as gaze, gestures, and posture (Mortensen, 2012). The term 'multimodal conversation analysis' was suggested to accommodate such changes (Stivers and Sidnell, 2005). However other CA analysts question this nomenclature and suggest an alternative, 'embodied conversation analysis', arguing that the use of the term multimodal implies separate 'modes' of human communication, rather than 'social action,' which necessarily combines them (Deppermann, 2013; Heath and Luff, 2000; Streeck, 2013). This paper agrees with this alternative naming and avoids *a priori* separation of human behaviours into different 'modes'. Instead, separation is recognised in the different technology affordances available to different groups of actors and notions of mode utilised conceptually to address the 'translation' of text into talk-in-interaction.

The analysis that follows is therefore premised upon a distinction between the technology-mediated behaviours of the streamer and the audience members. The streamer's behaviour is captured and presented through audio-visual means, while the audience members communicate through a text-chat system and notification messages. These different technology-based interactions sit in parallel with one another and typically result in separate communicative 'channels'. To accomplish audience interaction, streamers selectively read out the textual contribution of the audience members. This is more sophisticated than simple "mode switching" seen in video mediated interaction (Sindoni, 2021), and is instead a form of "modal translation" (Boria et al., 2019) in which the content of the text chat is re-performed as a "turn-at-talk" within a dialogic-monologue. Reading aloud, requires that the streamer physically reorients to a monitor on which the chat content appears. Head and body movements, therefore, become a primary means to signal these situated 'attentional shifts' and form the basis for the interaction. Such movements are observable by all and are an essential resource for sense-making.

3. Methodology

Like Sacks' use of telephone-based recordings (Sacks, 1995), the following analysis uses 'data to hand' produced and made available by technology, the Twitch web platform (https://www.twitch.com). As such, the analytic gaze is informed by 'naturalistic' data in the sense that the researcher is situated as one more audience member.

Over an extended six-month period instances of Twitch broadcasts (or 'streams') were observed, utilising an ethnographic method, and a sense of the behaviour and interactions gathered. Specific streams were chosen, and more detailed analysis carried out, premised upon the recording and transcribing of behaviours. The subsequent analytic framing was Conversation Analysis (CA), adapted to online interaction, which entails identification of interactional moments, gathering these

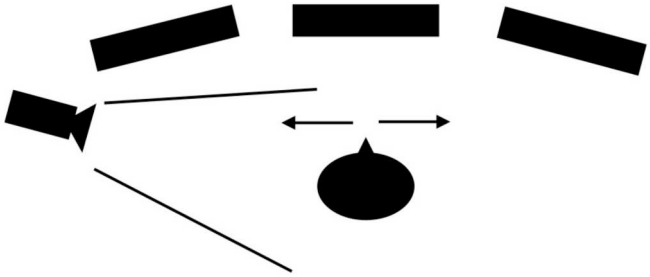

Fig. 2. Head and monitor positions.

into 'collections' of similar phenomena, and detailing the sequential production of the activities.

The data are composed of: 1. a visual representation of the game-play in video format; 2. a 'picture-in-picture' audio-visual representation of the Streamer; and 3. a visual record of the 'text-chat' activities of audience members. The multimedia data collected conveys the view (point) of all audience members (see Fig. 1) including those who view the stream at some later date.

The picture-in-picture representation of the Streamer, in centre-left of the image (Fig. 1), allows for an examination of head and body movements in relation to the camera viewpoint (Fig. 2).

The analysis rests on a distinction between three head positions (forward, right and left) that indicate the streamer's orientation to different monitor screens. These screens are not visible to the viewer in the collected data (although streamers do at times position the camera so that all monitors are visible). All that is visible is the head movement and position relative to the camera view. The analytic description of one of these monitors as containing the Chat[2] information and the other the composite broadcast stream is born of the ethnographic element of the study. That is over time it became clear that when looking in one direction the Streamer had access to Chat information and in the other the broadcast content. This reading of particular screen content is not supportable through the data directly. Yet it is necessary for the analysis. More importantly it is a piece of knowledge that all viewers, including the researcher, require and is the consequence of an 'everyday analytic'. As such it is a constitutive part of the sense-making practices of viewers and a competence or form of epistemic literacy. It was precisely the question 'why is he moving his head to the side?' that was asked by the analyst when first observing the streamed activities.

The analytic process includes the transcription of the behaviours observed. The resulting transcripts are based on the Jeffersonian system (Atkinson & Heritage, 1984; Heath et al. 2010), adapted to detail embodied interaction (See Appendix). Transcription is used to present relevant behaviours in a simple and readable fashion.

There are a number of novel aspects to studying streaming behaviour. Viewing a live or recorded broadcast does not require an account or specialist knowledge. Taking on the role of stream-viewer is effortless and the researcher has free rein to choose any stream and accompanying chat. The choice of data collected therefore rested on technological features. The platform filtering system that prioritises streams by viewer count was used to guide the research focus and a 'popular' streamer chosen. Arguably this experience and choice replicates that of the novice audience member.

While the ease of access side-steps several research issues, including questions of recruitment, research participation, and influencing behaviours through observation and camera placement, etc., it nevertheless prompts important ethical questions.

Much of the current work on Twitch, and other video streaming sites such as Youtube, treats the streams as publicly available data (Johnson and Woodcock, 2019). Yet, clearly, the participants do not broadcast their behaviours for the benefit of social researchers. Hence, social research should involve greater care.

For example, there are limitations in terms of assessing the competence and age of participants. There are also reputational issues; The Streamer is a public figure who makes a living from this activity. Arguably the streamers have ownership rights to their image and name (etc.). In addition, Twitch is a company with rights to the broadcast materials that are detailed in the terms and conditions (T&C) of the site.

The author undertook a review of these issues, that included the reading of the T&Cs of the Twitch site, the following of institutional guidelines on the collection of social media data, and the reading of relevant research ethics publications (British Sociological Association (BSA) (2017); Townsend and Wallace, 2016).

The collected data were anonymised, removing identifying content, such as Streamer name, and the visual materials manipulated so that the physical identity of the streamer was obscured. Chat member usernames and Chat post contents were, however, left intact because user names do not link to 'real-world' identities (unlike other forms of text based media such as Twitter, Twitch chat is not searchable) and messages are short and context free. Where the name of the streamer appears in chat content (through what is known as a'direct message' or DM) it has been obscured.

The Twitch terms and conditions permit the production and sharing of 'clips' (short audio-visual segments) by registered viewers. But they prohibit the 'monetisation' of such activities. They do not claim rights over non-registered viewer behaviour. Given the expectation of greater care on the part of the social researcher, the recorded data were stored securely on a password protected computer and only the necessary amounts of data were collected. Data collection did not include so-called 'scraping' methods seen in the collection of Twitter data through an API (Burnap et al., 2015). The strategy was to collect 'surface' materials only, limited by what an audience member can see, hear and read. To that end four streams were collected, totally 38 h of content.

4. Analysis

Licoppe and Morel (2018) identify a 'Read Aloud and Respond' (RAR) action in the use of the mobile streaming application Periscope,

"RAR responses are constructed as a two-component turn, a first one in which the streamer reads the message aloud, and a second one in which he or she responds to the message (though this second part may also occasion multiple expansions)" (Licoppe and Morel, 2018: 16).

The following analysis starts by detailing the most basic adaptation of the Licoppe phenomenon – and the primary finding of this paper – the embodied activity of reading, indicated by a head pan and gaze fixation. To incorporate this action, Licoppe and Morel's acronym 'read-aloud and respond' (RAR) has been extended to 'read, read-aloud, and respond' or R(R-A)R. The addition of the 'read' action at the beginning of the definition becomes the primary focus. Of course, the Licoppe definition incorporates a physical reading, however, it is in the *visible* aspects – that is visible to audience members – of this action in Twitch

[2] The term 'Chat' is used to indicate the group of people contributing text messages to the stream. It is also a term used by the Streamer to reference this collective identity.

behaviours that analytic purchase is found.

Read, (Read-Aloud) and Respond – R(RA)R.

(1)

[[Transcript 1.png]].

TRANSCRIPT: S111120 6:36 - 7:04

((right screen = chat, middle screen = gameplay))

```
01  Sh:   ▶▶▶▶|6.5|●1                                  ((to right; at right screen))
02 S:     how are you | ay::|
03  Sh:               |▲▲▲▲ |                                    ((to centre))
04  Sh:   | 0.6|~~~~~|                              ((at centre screen; small nod))
05  Sh:   ▶▶▶▶▶▶▶▶▶▶▶▶                                 ((to right, while nodding))
06 S:     I'm tired (.) | is what i am|
07  Sh:                |▲▲▲▲         |●2                          ((to centre))
```

- 1 Image One

- 2 Image Two

In extract (1) the Chat screen is to the right of the Streamer. The segment is a question answer sequence but with the Streamer being the only speaker. On line 01 the Streamer turns his head to the right and leaves it here for 6.5 s. He then utters 'how are you' while retaining this head position. A message appeared in the chat 9.6 s earlier which read "gorkem_one: how are you @[Streamer name]". The utterance at line 02 replicates the content of the message, minus the Chat member's username and a direct message convention (the addition of an '@' symbol to the streamer's name), which highlights the message in the chat screen. This utterance is then a "reading aloud" (RA) action.

Immediately upon producing the RA utterance, the Streamer utters "ay::" ("I") and turns his head to the centre computer screen. On line 04 there is a vocal pause of 0.6 s followed by a small head nod. The Streamer then pans his head to the right for 1.2 s while nodding, on line 05. He then produces a two-part response. The first "I'm tired (.)" is delivered while maintaining a gaze alignment with the Chat screen ('right screen'); the second part "is what I am" is accompanied by a head and gaze pan back to the centre screen. The two head positions at either end of the sequence can be seen in the images that accompany the transcript (Image 1 and Image 2) with their timing indicated in the transcript by a bullet points and numbers (●1 and ●2).

The head movement in line 01 is the basis for an embodied 'reading' action. This is not evidenced in the prima facie examination of the visual and audible data because the three computer screens are not visible in the video. Instead, it is an implied visual-cognitive action born of the next utterance and the comparison of this next utterance with the Chat content. Recognition of this as a physical and mental 'reading' is an everyday analytic outcome available to both the researcher and the onlooking participants.

Line 01 exhibits the basic 'reading' phenomenon, comprised of gaze panning – indicated by head movement – followed by gaze fixation, with head static. This is then followed by the read-aloud and response elements identified by Licoppe and Morel in lines 02 to 07.

Physical reading is premised upon embodied adaptation to the environment and (viewing) technologies as objects within that environment. Its capture is reliant upon a camera positioned within that environment and it is this camera that provides novel 'insight' for Twitch audience members and viewers. This affords interactional opportunities for audience members.

It is also the case that there is head and gaze movement within the sequence, such that the Streamer turns away from the chat screen and then returns to it in between the depicted head positions. These intermediate gaze patterns are common in the Streamer's interaction with the Chat screen. We see a similar set of (intervening) gaze patterns in the second example (2).

(2)

TRANSCRIPT: S111120 9:40-9:46

```
01 S:     good st|uff guys |
02  Sh:         |►►►►►►►  |1.8|                    ((to right; at right screen))
03 S:     hows your day g | going| huh |hokay |I guess
04  Sh:                   |▲▲▲▲  |                                ((to centre))
05  Sh:                              |►►►►►►|                      ((to right))
```

In this segment, the Streamer has completed a previous interaction with the Chat with 'good stuff guys' (line 01) as he does this he moves his gaze to the Chat screen. Leaving it there, while reading the screen (line 02) for 1.8 of a second. He starts to read-aloud a message (line 03) and pans his gaze away from the screen (line 04) as he finishes the reading-aloud on a restarted 'going'. Responding with 'huh hokay I guess' he pans back to the chat screen during 'hokay' (line 05). As will shown, this this final pan as well as the initial gaze pan to chat during a prior utterance become relevant when considering the 'chaining together' of R (RA)R sequences later in the paper.

The text message 'read, and read-aloud' by the streamer in instance (2) has appeared around nine seconds before in the Chat screen (see Fig. 3). At the point of reading, it is at the top of the Chat window (messages 'scroll' from the bottom). Amongst what can be seen to be a range of chat posts, as with the previous instance he chooses the one that is directed at him with a greeting and address term "Hi [name of Streamer] How is your day going?!".

Key in this activity is the choosing of certain messages to respond to amongst a collection of available messages. This choice effects the social engagement and exhibits preferential recipiency. As audience member there is a sense of choice on the part of the Streamer and perhaps an indication of which types of messages will likely be incorporated into such recipiency activities. This lends the participation a social dynamic in which text chat become implicated in a set of social interactional strategies for gaining attention.

4.1. Streamer response elicitation

To this point, the R(RA)R sequence has been seen in the selection of messages by the Streamer. There are other instances, however, when the Streamer actively elicits responses from the Chat members. For example, he asks a question and then waits for relevant responses, reads them aloud, and premises topic continuation on these responses. Analysis of these instances requires that we detail extended elements of the Chat messages. The multiple messages produced by audience members become *candidate* read-aloud targets.

In the following extract (3) the Streamer asks a question while facing forwards towards the gameplay screen: 'how do I meditate' (line 01). He qualifies this question with a comment about 'killing time' and then re-issues the question, reformulated as 'how do I do that' (line 02). At this point he is still looking forwards towards the main gaming screen. After a 2.2 s pause he utters 'actually don't know' and turns his head right towards the screen on which the chat messages are displayed.

(3)

TRANSCRIPT: S111120 1:22:30-1:22:59

```
((Head is forward toward gaming screen))
01 S:     how do I meditate (1.7) there's a way you can like kill time
02 S:     (0.9) how do I do that (2.2)
03 S:     |actually don't kn|ow (0.5)|
04 C:     |M1
05  Sh:                    |►►►►►►  |                        ((head to right))
06  Sh:   |9.7   |                                             ((head right))
07 C:     |M2-14|                                            ((chat messages))
08  Sh:   ▲▲▲▲|---►►►►---|---▲▲▲▲--►|              ((head movement and position))
09 C:         |M15-18     |M19-21     |M22-23                 ((chat messages))
10  Sh:   |1.7    |                                             ((head right))
11 C:     |M24-28|                                            ((chat messages))
12 S:     |hold e|m |
13  Sh:   |▲▲▲▲  |~~|                          ((head to forward; head forward))
14  Sh:   |►►►►  |                                            ((head to right))
15 C:     |M29-30|                                            ((chat messages))
16 S:     hold gee
17  Sh:   ▲|▲▲▲|---
18 G:      |~~~|                    ((action icons appears as circular overlay))
19 S:     oh
20 G:     |2.3|                              ((moves cursor over each icon until
                                          'Meditate' text appears then clicks))
```

Over a period of time audience members post a series of messages. The timing of which are indicated in the transcript below with 'M' followed by a number, with some arriving at the same time (e.g.'M2-14, line 07).

These Chat messages are split across two types of answer ('hold em' and 'hold gee') referring to two suggested keyboard actions that might put the gameplay character in 'meditate' mode. One or more of these suggestions are incorrect.

The following table shows the relevant messages, detailing the temporal relationships between message appearance and talk through a notation key (M1, M2, etc.) inserted into the transcript. Note, the appearance of messages does not have a duration as such, hence timing points ("|") are used in the transcript to show when each message is first visible on the Chat screen (e.g. line 04).

While the first message-as-answer in line 04 (M1) contains text that the Streamer will eventually 'read-aloud' ('hold m' – line 12), there is no way of knowing if it is this message that he reads. There are in fact four other message with the same content (M3, M14, M16, M27) all of which are visible to the Streamer before he produces the utterance (Lines 7, 9, 11). The utterance on line 12 occurs after a period of moving the gaze to and from the Chat screen on line 08 and is produced as he pans his gaze to the centre screen (line 13). It might be assumed that at this point he enacts the key press without consequence because he then pans back to the Chat screen on line 14, during which two more messages appear (M29 and M30), one of which reads 'Hold G'. Again, there is no way of knowing if this is the actual message being read-aloud on line 16 because M9, M20, M23 all contain the same text. In addition to the potential for direct reading, there are also messages relevant to both suggestions that could have been summarised as 'hold m' or 'hold g' but are actually simply the letter 'm' or 'g' (M4, M5, M6, M7, M10, M11, M13, M15, M17, M18, M19, M21, M22, M24, M25, M28) or a similar phrase 'Press g' (M29). It could, of course, also be that the Streamer is summarising the mass of responses.

Episodes such as this, in which the Streamer surveys the Chat messages, are accompanied by extended physical readings produced through glancing to and from the Chat screen. In one sense, it doesn't matter which precise message is being read and then read-aloud. The gaze alignments with the chat screen indicates 'availability to read' and the chat participants continue to make suggestions. Each message containing the precise read-aloud text, and indeed those that merely give the information in a different format, could be seen, and understood by audience members, as the read-aloud message. Hence, each Chat participant that produced a relevant message could experience the interaction as relevant to them.

That the issued question does not have a specified referent or recipient and the ambiguity of the recipiency status of any given 'read' and 'read-aloud' message lends the interaction a sense of collective action and engagement. The question-utterance by the Streamer in combination with head movements to and from the Chat monitor precipitates thirty interaction-relevant Chat messages, produced over a seventeen second period. This concentration of messages indicates of a sense of expectation that they will become the target of the R(RA)R sequence.

In addition, the messages are sent at strategic moments in relation to the talk and head movements of the Streamer. The first message (M1) arrives at the end of the third turn construction unit (TCU) of the Streamer, 'how do I meditate (1.7) there's a way you can like kill time (0.9) how do I do that (2.2)' (line 04), and in overlap with a fourth, 'actually don't know'. This last message arrives before the Streamer turns to the Chat screen.

Over the next nine seconds, as the Streamer continues this gaze alignment with the Chat screen, thirteen more messages appear (lines 6–7). As the Streamer pans between the main monitor and the chat monitor for 4.5 s (lines 8 and 10) a further thirteen messages appear. The final two messages that give the alternative answer of 'hold g' are only sent once the first answer 'hold m' is uttered (and potentially actioned without the required outcome) on line 15.

It can be seen that the Chat participants are responding to the head movements of the Streamer. Aside from the very first message, each subsequent message sent is oriented to the potential to be read and read-aloud. This underlines the assertion that head and gaze position of the Streamer is an interactional resource for the Chat participants.

This, then, speaks to a central dynamic in Twitch interaction of 'potential attentional recipiency'. First in terms of the Streamer's availability to interact and second in terms of the positioning of Chat members as potential interaction partners. Looking at a person in face-to-face talk can instigate recipiency (the talk is for that person only) and form part of an 'access ritual' (Goffman, 1961 Kendon, 1990). In Twitch a similar dynamic is in play. The physical reading and reading-aloud of a Chat member's message draws them briefly into a 'focussed encounter'. The speaking of a message as a form of 'self-talk' (Goffman, 1983, p. 79) situates message writers as a conversational partner. While multiple candidate messages disperse that potential recipiency amongst message creators, each relevant message creator is situated as recipient and conversational partner.

4.2. *Extending conversations*

Now that we have the general sense of the R(RA)R sequence, deployed either with the Streamer drawing on existing messages or eliciting responses, the next question is how this sequence scales up to form ongoing talk.

One way this occurs is through turn extensions. Here the response to a read-aloud event is used to generate a longer stretch of talk. This is similar to the 'extensions' seen by Licoppe in Periscope, however here it directly relates to the embodied movements of the Streamer and the continuity of the collective behaviour and topic dynamics (either as a single topic or through what will be called 'topic juggling' later).

In extract (4), the R(RA)R sequence (lines 01 to 03) is followed by a verbal pause and then a further utterance that elaborates on the answer given. 'I'm archiving everything to a YouTube channel' is delivered as the Streamer pans to and then looks at the centre gaming screen. In the pause that follows, he turns back to the Chat screen and then utters 'as we speak' (line 06).

(4)

TRANSCRIPT: S111120 10:36-10:44

```
01  Sh:   ▸▸▸▸▸▸▸ |1.4|                          ((to right; at right screen))
02 S:     why can't I see your old streams I delete everything
03        (1.0)
04 S:     |i'm archiv|ing everything to a youtube channel
05  Sh:   |▲▲▲▲▲     |                                        ((to centre))
06 S:     (|------|) as we speak
07  Sh:    |▸▸▸▸▸▸|                                          ((to right))
```

This is a relatively small extension of the original interaction and acts as an account for what might be seen as a 'snap answer' (given that the second pair-part answer segment is delivered immediately upon the completion of the read-aloud question on line 02). The later head and gaze pan to the chat screen (line 07) serves the function of allowing for further reading, as will be seen in the next section.

In extract (5) the R(RA)R sequence (lines 01–05) is produced with the head positioned initially toward the Chat screen. This is followed by a head turn between the question and answer element (in overlap with the end of 'morales' and the beginning of the 'e::r' token). The topic of buying a playstation (the required console for the game) is then extended after a half second pause with three turn-like units. First 'chances are:: (0.5) I don't have that luxury' (lines 07 and 09). Then 'unless I buy from a scalper [a person who sells overpriced versions of the console] but' (line 11) and then after a pause of 0.5 s 'I'm not gonna do that' (line 13).

(5)

TRANSCRIPT: S111120 10:24-10:36

```
01  Sh:   ▸▸▸▸▸▸▸ |3.4|                          ((to right; at right screen))
02 S:     will you be playing spiderman
03 S:     miles moral|es E:|:r if I get my
04  Sh:             |▲▲▲▲▲|                                  ((to centre))
05 S:     hands on a playstation five abso| lutely |(0.5)
06  Sh:                                  |▸▸▸▸▸▸ |            ((to right))
07 S:     |chances ar:: |(0.5) ar:: I don't
08  Sh:   |▲▲▲▲▲▲▲▲     |                                  ((to centre))
09 S:     h|ave that luxury|
10  Sh:    |▸▸▸▸▸▸▸▸      |
11 S:     |unless i buy from a| scalper but
12  Sh:   |▲▲▲▲▲▲           |                              ((to centre))
13 S:     (|-----|) I'm not gonna do that
14  Sh:    |▸▸▸▸▸|                                         ((to right))
15        (|--------|)
16  Sh:    |▲▲▲▲▲▲▲▲|                                     ((to centre))
```

The utterance elements are formulated as a series of alternatives: First 'yes if I can get a playstation five'; then 'this is unlikely'; then the introduction of a solution, and then a decision statement.

There is not time to develop the point further here, but this kind of 'self-talk' (Goffman, 1983) construction, in which the streamer takes alternative or alternating perspectives as though they were a conversation interchange is a very common feature of streamer talk. The important and relevant aspect of the behaviour is instead the head movements.

A second way that the R(RA)R is productive of an ongoing set of verbalisations is through the sequencing or chaining of a series of R(RA)R instances. There is an indication of this when the Streamer looks back towards the chat screen at the end, or during, his verbal responses in instances 2 and 4 and 5.

The R(RA)R sequence, when appended by a head turn during the 'response' element is infinitely extendable, in that one R(RA)R sequence can lead to another. On each occasion of 'reading', however, the Streamer is dependent on the Chat contents to either extend the current topic or introduce a new topic. Not identifying a read-aloud target would curtail the chaining of R(RA)R sequences together. The searching for relevant messages helps explain the repeated head turns to and from the chat screen.

4.3. *Conversation juggling*

While Chat messages may be relevant to the current conversation, there is also the possibility of jumping to and from topics. Here, messages relevant to earlier topics are chosen. This can be seen in the following longer set of instances. Given that this is a longer sequence, each R(RA)R sequence will be shown in turn and the line numbering will continue from one instance to the next.

Extract (6) follows on immediately from instance (5).

(6)

TRANSCRIPT: S111120 10:45-10:52

```
08 S:     can you block enemy attacks like in odyssey
09 S:     yeah with a |shield|
10  Sh:              |▲▲▲   |                              ((to centre))
11        (|----|----|)
12  Sh:    |►►►►|~~~~|                          ((to right; at right screen))
13 S:     |i don't think you | could block
14  Sh:   |▲▲▲▲▲▲▲          |                              ((to centre))
15 S:     (0.8) regularly I think
16 S:     you |c'd block with a sh|ield
17 S:         |►►►►►►►          |                               ((to right))
```

In line 07 of extract (5), the Streamer has moved his head to the right. In extract (6) (above) he reads aloud a message "can you block enemy attacks like in odyssey" (line 08). He then answers and extends this answer over lines 09–17.

Extract (7) (below) continues from the previous one and shows the last two lines in which the Streamer moves their head and gaze to the right. There is then a gap of fifteen seconds during which the streamer looks to the right in the direction of the chat screen.

(7)

TRANSCRIPT: S111120 10:51-11:18

```
16 S:     you | c'd block with a sh|ield
17 S:         |►►►►►►►          |                               ((to right))
18        (15.0)                                        ((at right screen))
19 S:     it's on pee ess four dude (0.3)
20 S:     you think I'm gonna play it on pee ess four (0.5)
21 S:     you're fu|cking c|razy | (0.4) you think
22  Sh:           |▲▲▲▲▲▲ |►►►►►|                    ((to centre: to right))
23 S:     i'm gonna | touch that game on pee ess | four|
24  Sh:            |▲▲▲▲▲▲                      |►►►► | ((to centre; to right))
25 S:     (1.1) ↑ha ↑huha good it's a good joke
```

The utterance 'it's on pee ess four dude' (line 19) does not make sense in terms of the preceding topic of gameplay (making a block with the shield). It is, instead, a reference to the earlier question about playing a certain game (Spiderman Morales) in extract (5). The Chat member is pointing out that the Streamer does not need to purchase a new console but can use an existing one (which the Streamer has in his possession).[3] The chat message does not make explicit the link to the earlier topic. The use of "it's" ("it is") is an indexical expression that requires contextual understanding.

As part of the utterance on line 24 the Streamer turns his head back to the right screen during "four" (line 23). Again, we see the preparation for the reading action nestled in the response element of the R(RA)R sequence. There follows another lengthy gap of 11.4 s (line 26) and the utterance "wait so there's no parry" (Instance 8, lines 26–27).

[3] At the time a new console (the PS5) had just been released with higher performance. The PS4 was the existing console but newer games would not run as well on it.

(8)

TRANSCRIPT: S111120 11:13-11:45

```
23 S:    i'm gonna | touch that game on pee ess| four|
24  Sh:           |▲▲▲▲▲▲                     |▸▸▸▸ |  ((to centre; to right))
25 S:    (1.1) ↑ha ↑huha good it's a good joke
26       |11.4|                                          ((at right screen))
27 S:    wait so there's no parry
28 S:    you can |parry | js |can't blo:ck|
29  Sh:          |▲▲▲▲▲ |    |▸▸▸▸▸      |            ((to centre; to right))
30 S:    (0.3) I can't use like a
31 S:    (|------|---|---) li|ka::
32  Sh:   |▲▲▲▲▲▲|   |▸▸▸▸   |                        ((to centre; to right))
33 S:    (0.9) | speer| or (0.4)
34  Sh:        |▲▲▲▲▲ |                                ((to centre))
35 S:    i think you can block with a speer atchly (0.4)      ((croaky voice))
36 S:    i can't use | like a sw|ord
37  Sh:             |▸▸▸▸      |                       ((to right))
38 S:    an |sit there| and hold bl|o:ck
39  Sh:    |▲▲▲▲     |▸▸▸▸       |                    ((to centre; to right))
40 S:    (--|----) i could |parry |
41  Sh:    |▲▲▲▲          |▸▸▸▸  |                    ((to centre; to right))
42 S:    i just can't hold a block with a sword
```

Reference to a "parry", a game move in which the blow of an opposing character is blocked, is a shift back to the earlier topic – that of blocking an enemy attack seen in extract (6).

Finally in extract (9), the overlapping R(RA)R is again seen followed by a reading action of 6.8 s (line 46) and the reading aloud of a chat message (lines 47 to 49) that references an even earlier topic of archiving the previous recordings on Youtube (extract 4).

(9)

TRANSCRIPT: S111120 11:41-11:54

```
43 S:    (--|----) i could |parry |
44  Sh:    |▲▲▲▲          |▸▸▸▸  |                    ((to centre; to right))
45 S:    i just can't hold a block with a sword
46       |6.8|                                          ((at right screen))
47 S:    is that the new thing streamers are doing uploading
48 S:    vods into youtube |YeS: |
49  Sh:                    |▲▲▲▲ |                     ((to centre))
```

This series of segments shows the skilful way the Streamer 'juggles' topics through selective reading aloud actions. The R(R)R sequence is repeated, and the preparatory head turn is embedded in the response and extension element of each instance. This juggling is akin to the 'threaded' nature of topic generation in various forms of text-chat (not only Twitch). Here, messages are discontinuous and individual posts reference messages from earlier points in the activity (Meredith, 2019).

The analysis has revealed the central importance of the embodied head movement: how it helps develop the RAR sequence identified by Licoppe; how it is used to select messages to form ongoing talk in either a responsive or elicited manner; and how it is used as a resource for Chat participants to engage the Streamer. The R(RA)R sequence is also the basis of topic extension. Finally, an embedded physical reading as part of the 'response' element enables the chaining together of R(RA)R actions through 'threaded' topic development and topic juggling.

5. Discussion

Identification and analysis of the Read-(Read-aloud)-Respond sequence in Twitch interaction is premised upon the availability of

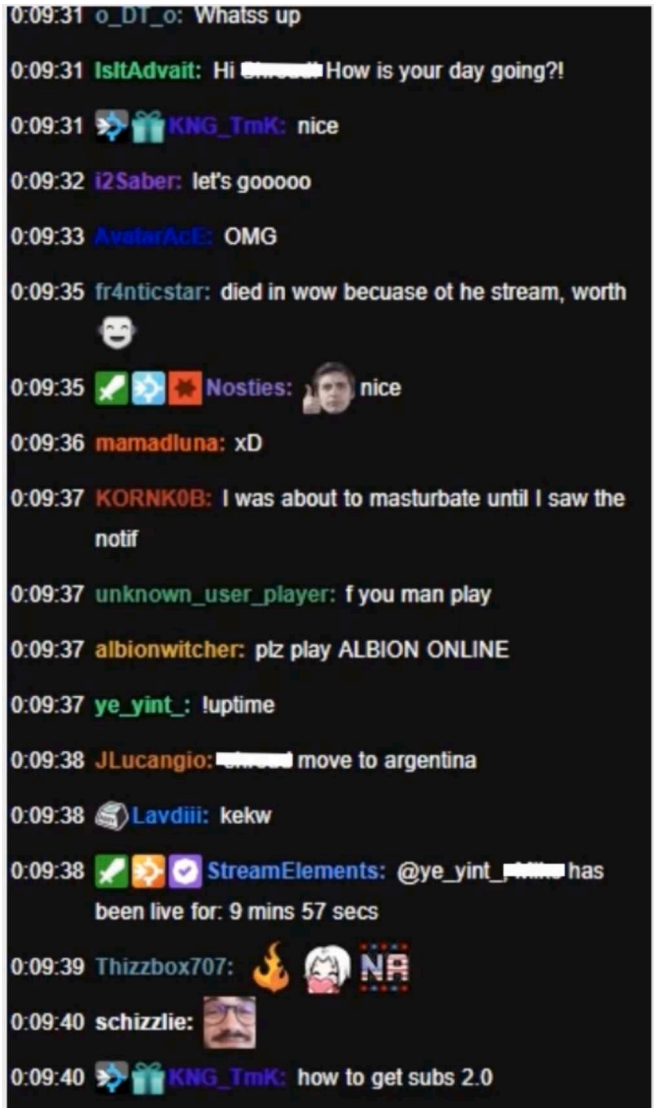

Fig. 3. Chat Screen contents.

technology afforded information. This information is a product of the configuration of physical technology and resultant audio-visual and textual contents in Twitch. This multimedia configuration includes specific technologies (video camera, computer monitor, mouse, keyboard), their affording of (multimodal) recording and archiving, and the spatial arrangements and orientations of component elements.

The behaviours of the Streamer, the Chat members, and crucially the analyst, are constrained and afforded by this multimodal technology configuration. Within this configuration, the streamer and audience member have different 'rights to act' – or more precisely different 'means' and 'affordances' to act.

The Chat member can only communicate through asynchronous text messages. They 'post' these messages to a system that while temporally ordered (based on submission time) lends little control on position and sequence within the message window. Underlining the 'multimodal/embodied' distinction, while 'visual' these messages are not necessarily 'visible'; that is while posted to the message board they are not necessarily seen by the Streamer. Like the cheers and claps of a theatre audience, these individual activities are potentially subsumed within the mass of other messages.

The Streamer has more control. They can use voice to communicate. Yet they still need to work within the technology affordances. For them, the gameplay audio-visual materials are in one place and the audience chat another. Transferring visual attention from one to the other requires the novel embodied behaviour at the centre of this paper – the head turn.

The Streamer utilises the ongoing commentary Chat for the purposes of topic and conversation-like development. To do this, the Streamer moves his attention to a second monitor, reads out selected Chat contributions and responds to them as though they were part of a spoken interaction. As with other forms of multimedia interaction (Reed, 2017), the Streamer applies the norms of verbal conversation to the multimedia components, transforming textual contributions into a verbal 'turn' and 'response'; He transforms textual commentary into a 'first pair part' of an adjacency pair, which then requires a 'second pair part', a response, answer, or the like (Sacks et al., 1974).

For Manovich digital technologies not only allow for the separation and combination of content – sound, image, and text – but also allows for the transference of the content's 'fundamental techniques, working methods, and ways of representation and expression' (Manovich, 2013: 110). This is seen when texts become 3D animations, sounds become visualised, and live action footage combines with computer generated materials. Reed (2017) extended Manovich's formulation to include the 'fundamental techniques' of talk-in-interaction'- i.e. turn-taking, adjacency pairs, and the like – such that interaction through the music sharing platform Soundcloud combined the norms of text-chat interaction with those of face-to-face interaction.

There is a similar dynamic with Twitch, which incorporates the norms of text-chat with the norms of talk. For Schönfeldt and Golato (2003) text chat involves the achievement of 'virtual adjacency',

"adjacency in chats is thus an achievement of the participants' reading of, and selection from, a quickly changing stream of messages addressed to them; it is thus a "virtual adjacency.'' This virtual adjacency is not merely a construct on the part of the analysts but a true reading on the part of the chat room participants (p. 251).

For Manovich, such processes are a form of 'digital materialism' (2002: 10) or 'deep remixability' (2013: 25). They not only co-opt practices and techniques from other formats, they also have the potential to become new fundamental techniques of emerging media.

The R(RA)R sequence, and other similar techniques, has the potential to be a 'fundamental technique' of Twitch interaction. Just as turn-taking is a fundamental technique of face-to-face conversation, so interaction on Twitch relies entirely on the ability to cross modal boundaries to connect Streamer with Chat participant. That the R(RA)R sequence rests on the fundamental techniques of talk further evidences the claim in CA that everyday talk is a primary resource for social interaction.

While only indicated in this paper, the activities in the Chat are multiple and complex. While often 'about' the stream contents they never-the-less have a separate character and trajectory. It is only in selective moments that those 'immediately present' are included in the stream. Interaction is entirely dependent on the Streamer's ability to ratify another person through the R(RA)R sequence. The Chat is then a 'subordinated' encounter in relation to that centred upon the Streamer.

There are a number of other situations in which the R(RA)R sequence is operative, which cannot be detailed here. Briefly, they include activities in which general Chat commentary is relevant to the ongoing gameplay and commented upon by the Streamer (typically formulated through reference to the group identifier 'Chat' – as in 'the Chat is really mad today'). Other instances include the paraphrasing of Chat messages to fit the tenor of the ongoing gameplay action. Finally, there is a category of messages that carry a normative expectation that they be addressed: 'Donations' and 'Subscriptions' are forms of interaction available to viewers that involve monetary contribution to the Streamer. Here, normative expectations of a response are reinforced and the typical dynamics of Streamer ratification are briefly suspended.

Unlike other streaming activities, such as Periscope, Twitch is characterised by continual observation of not only the view of the stream, but also the Streamer themselves. Their embodied action,

caught by the camera, is continually in view and continually watched by audience members. Indeed, it could be said that the primary attraction of Twitch is that watching others becomes normalised and commodified. As such, engagement and interaction are achieved through forms of sense-making peculiar to the medium. Close observation becomes a members' method for understanding how and when to create 'successful' contributions (as well as other activities). Watching when the Streamer is oriented to such contributions, and when she or he is not, is a vital component of the sense-making practices of participants and suite of actions as appropriate methods.

6. Conclusion

This paper has undertaken an initial analysis of the streaming platform Twitch. It has identified the R(RA)R ('Read (Read Aloud) Respond') action sequence and noted its similarity to sequences found in other platforms. Unlike those platforms, Twitch is premised upon a single Streamer and a large audience. Interaction between a Streamer and an individual Chat member is constructed by the Streamer through a form of acknowledgement and ratification that leads to a momentary encounter. Strategic 'reading' and reading aloud not only function to link and sustain conversation, they also structure the embodied performance of the streamer, who must shift their visual attention away from the primary activity. These head movements are then a primary resource for interaction.

The small noticing of a head movement was used to unpack and examine the data as a product of a set of technologies, which afford selective insight and analysis. In turn, the technology arrangements and configuration of Twitch was used to reflect on Conversation Analysis as practice. This is in line with the 'respecification' of 'mediated interaction' (Arminen et al., 2016) and moves towards a 'digital CA' (Giles et al., 2015). It contributes to an ongoing project that looks to 'digitise Sacks' (Housley et al., 2017) and inform methodological development of future analysis of technology-mediated interaction.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data are publicly available materials from the Twitch web site

Appendix

TOTALS BY MONTH

Month	Avg. concur. viewers	Avg. concur. channels	Time watched	Active streamers
March 2021	2,857,845	114,339	721M hrs	n/a
February 2021	2,945,135	122,623	1,979M hrs	9,517,967
January 2021	2,919,533	119,271	2,171M hrs	9,894,745
December 2020	2,522,869	106,845	1,876M hrs	9,241,666
November 2020	2,487,091	103,917	1,790M hrs	8,546,985
October 2020	2,385,206	95,074	1,774M hrs	7,844,426
September 2020	2,202,200	91,861	1,586M hrs	7,461,252
August 2020	2,178,080	96,309	1,620M hrs	7,709,990

twitchtracker.com - 11/03/21

Transcription Notation

The transcription notation and system used is adapted from Jefferson (Atkinson & Heritage, 1984) and Heath et al. (2010). It replaces numerical pause lengths where appropriate with a graphical representation so as to allow for clear alignment of the pause with a simultaneous or delayed onset action. Actions are presented graphically, showing duration in tenths of a second, and then a descriptive gloss given in double parentheses justified right. Identifiers showing embodied action are appended with a relevant character to differentiate them from talk.

A: participant identifier related to talk
A(h): participant identifier related to embodied action (in this case the movement of the head)
| timing point, relating embodied action to aligned point in talk line
[vocal overlap
|~~~| embodied action, aligned with vocal utterance, in tenths of a second, described in double parentheses right justified
|6.5| embodied action, in tenths of a second, described in double parentheses justified right
(———) graphical vocal pause, length indicated in tenths of a second
(0.5) numerical vocal pause
• image capture point aligned with talk
|~~~~|~~~~~| ((action1; action2)) graphical representation and description of *sequential* actions
|~~~~~~~~~~| ((action1, action2))
graphical representation and description of *simultaneous* actions
gaze pan in tenths of a second
►□►□►□gaze pan to the speaker's right in tenths of a second
◄□◄□◄□gaze pan to the speaker's left in tenths of a second
▲▲▲▲gaze to speaker's forward in tenths of a second

References

Anderson, S.L., 2017. Watching People Is Not a Game: Interactive Online Corporeality, Twitch. tv and Videogame Streams. Game Studies. Int. J. Comput. Game Res. 17 (1).

Arminen, I., Licoppe, C., Spagnolli, A., 2016. Respecifying mediated interaction. Res. Lang. Soc. Interact. 49 (4), 290–309. https://doi.org/10.1080/08351813.2016.1234614.

Atkinson, J.M., Heritage, J., (1984). Structures of Social Action. Studies in Conversation Analysis. Cambridge University Press.

Boria M and Tomalin M (2019) Introduction. In: Boria M, CARRERES Á, Noriega-Sánchez M, et al. (eds) *Translation and Multimodality: Beyond Words*. London: Routledge, pp. 1–23. DOI: 10.4324/9780429341557-1.

British Sociological Association (BSA) (2017) *Ethics Guidelines and Collated Resources for Digital Research. Statement of Ethical Practice Annexe*.

Burnap, P. et al. (2015). Detecting tension in online communities with computational Twitter analysis. Technol. Forecasting Social Change, 95, 96–108. [Online]. Available at: https://doi.org/10.1016/j.techfore.2013.04.013.

Cabeza-Ramírez, L.J., Sánchez-Cañizares, S.M., Fuentes-García, F.J., 2020. Motivations for the use of video game streaming platforms: the moderating effect of sex, age and self-perception of level as a player. Int. J. Environ. Res. Public Health 17 (19), 7019. https://doi.org/10.3390/ijerph17197019.

Chen, C.-C., Lin, Y.-C., 2018. What drives live-stream usage intention? The perspectives of flow, entertainment, social interaction, and endorsement. Telemat. Inf. 35 (1), 293 303. https://doi.org/10.1016/j.tele.2017.12.003.

Dargonaki, S., 2018. Performing gender on Twitch.tv: Gendered playbour through Butlerian theory. *International Journal of Media & Cultural*. Politics 14 (1), 103–110. https://doi.org/10.1386/macp.14.1.103_7.

Deppermann, A., 2013. Multimodal interaction from a conversation analytic perspective. J. Pragm. 46 (1), 1–7. https://doi.org/10.1016/j.pragma.2012.11.014.

Dux J (2018) Social Live-Streaming : Twitch.TV and Uses and Gratification Theory Social Network Analysis. In: *Social Computing and Social Media. Human Behavior*, Cham, 17 February 2018, pp. 47–61. Springer International Publishing. DOI: 10.5121/csit.2018.80305.

Garfinkel H (2002) *Ethnomethodology's Program: Working Out Durkeim's Aphorism* (ed. Ann W. Rawls). MD, USA: Rowman & Littlefield.

Giles, D., Stommel, W., Paulus, T., et al., 2015. Microanalysis of online data: the methodological development of 'digital CA'. Discourse Context Media 7, 45–51. https://doi.org/10.1016/j.dcm.2014.12.002.

Goffman, E., 1961. Encounters: Two Studies in the Sociology of Interaction. Bobbs-Merrill, Ind, US.

Goffman, E., 1983. The interaction order: american sociological association, 1982 Presidential Address. Am. Sociol. Rev. 48 (1), 1–17. https://doi.org/10.2307/2095141.

Gros D, Wanner B, Hackenholt A, et al. (2017) World of Streaming. Motivation and Gratification on Twitch. In: Meiselwitz G (ed.) *Social Computing and Social Media. Human Behavior*. Cham: Springer International Publishing, pp. 44–57. DOI: 10.1007/978-3-319-58559-8_5.

Heath, C., Hindmarsh, J., Luff, P., 2010. Video in qualitative research: Analysing social interaction in everyday life. Sage Publications Ltd.

Heath, C., Luff, P., 2000. Technology in Action. Cambridge University Press, Cambridge, UK.

Hilvert-Bruce, Z., Neill, J.T., Sjöblom, M., et al., 2018. Social motivations of live-streaming viewer engagement on Twitch. Comput. Hum. Beha. 84, 58–67. https://doi.org/10.1016/j.chb.2018.02.013.

Housley, W., Webb, H., Edwards, A., et al., 2017. Digitizing Sacks? Approaching social media as data. Qualit. Res. 17 (6), 627–644. https://doi.org/10.1177/1468794117715063.

Johnson, M.R., Woodcock, J., 2019. 'It's like the gold rush': the lives and careers of professional video game streamers on Twitch.tv. Inf. Commun. Soc. 22 (3), 336–351. https://doi.org/10.1080/1369118X.2017.1386229.

Lessel P, Mauderer M, Wolff C, et al. (2017) Let's Play My Way: Investigating Audience Influence in User-Generated Gaming Live-Streams. In: *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, Hilversum The Netherlands, 14 June 2017, pp. 51–63. TVX '17. Association for Computing Machinery. DOI: 10.1145/3077548.3077556.

Licoppe, C., Morel, J., 2018. Visuality, text and talk, and the systematic organization of interaction in Periscope live video streams. Discourse Studies 20 (5). https://doi.org/10.1177/1461445618760606, 146144561876060 29.

Manovich, L., 2002. The Language of New Media. The MIT press.

Manovich L (2013) *Software Takes Command: Extending the Language of New Media*. International texts in critical media aesthetics. New York ; London: Bloomsbury.

Meredith, J., 2019. Conversation Analysis and Online Interaction. Research on Language and Social Interaction, 52 (3), pp. 241–256. [Online]. Available at: https://doi.org/10.1080/08351813.2019.1631040.

Mondada, L., 2011. Understanding as an embodied, situated and sequential achievement in interaction. J. Pragmat. 43 (2), 542 552. https://doi.org/10.1016/j.pragma.2010.08.019.

Mortensen K (2012) Conversation Analysis and Multimodality. In: Chapelle CA (ed.) *The Encyclopedia of Applied Linguistics*. Oxford, UK: Blackwell Publishing Ltd, p. wbeal0212. DOI: 10.1002/9781405198431.wbeal0212.

Paz, S., Montardo, S.P., 2018. Performance play: consumo digital como performance no streaming de games da plataforma Twitch. Front. Estudos Midiát. 20 (2), 190–203. https://doi.org/10.4013/fem.2018.202.05.

Pink, S., Mackley, K.L., 2013. Saturated and situated: expanding the meaning of media in the routines of everyday life. Media Cult. Soc. 35 (6), 677 691. https://doi.org/10.1177/0163443713491298.

Reed, D.J., 2017. Performance and interaction on Soundcloud: Social remix and the fundamental techniques of conversation. J. Pragmatics, 115, 82–98. [Online]. Available at: https://doi.org/10.1016/j.pragma.2017.01.012.

Sacks, H., Schegloff, E., Jefferson, G., 1974. A simplest systematics for the organisation of turn-taking for conversation. Language 50, 696–735.

Sacks H (1995) *Lectures on Conversation. Volume I & II* (ed. G Jefferson). Oxford, UK: Blackwell Publishers Inc.

Schönfeldt, J., Golato, A., 2003. Repair in chats: a conversation analytic approach. Res. Lang. Soc. Interact. 36 (3), 241–284. https://doi.org/10.1207/S15327973RLSI3603_02.

Scully-Blaker R, Begy J, Consalvo M, et al. (2017) Playing along and Playing for on Twitch: Livestreaming from Tandem Play to Performance. In: 1 January 2017. DOI: 10.24251/HICSS.2017.246.

Sindoni, M.G., 2021. Mode-switching in video-mediated interaction: Integrating linguistic phenomena into multimodal transcription tasks. Linguist. Educ. 62, 100738 https://doi.org/10.1016/j.linged.2019.05.004.

Sjöblom, M., Törhönen, M., Hamari, J., et al., 2017. Content structure is king: an empirical study on gratifications, game genres and content type on Twitch. Comput. Hum. Behav. 73, 161–171. https://doi.org/10.1016/j.chb.2017.03.036.

Spilker, H.S., Ask, K., Hansen, M., 2020. The new practices and infrastructures of participation: how the popularity of Twitch.tv challenges old and new ideas about television viewing. Inf. Commun. Soc. 23 (4), 605–620. https://doi.org/10.1080/1369118X.2018.1529193.

Stivers, T., Sidnell, J., 2005. Introduction: multimodal interaction. Semiotica 2005 (156), 1–20. https://doi.org/10.1515/semi.2005.2005.156.1.

Streeck, J., 2013. Interaction and the living body. J. Pragmat. 46 (1), 69–90. https://doi.org/10.1016/j.pragma.2012.10.010.

Suchman, L., 1987. Plans and Situated Actions. The Problem of Human Machine Communication. Cambridge University Press, Cambridge University Press.

Taylor, T.L., 2018. Watch Me Play. Princeton University Press, Princeton University Press.

Townsend L and Wallace C (2016) *Social Media Research: A Guide to Ethics*. 5 July.

Wohn DY (2019) Volunteer Moderators in Twitch Micro Communities: How They Get Involved, the Roles They Play, and the Emotional Labor They Experience. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow Scotland Uk, 2 May 2019, pp. 160–173. ACM. DOI: 10.1145/3290605.3300390.

Woodcock, J., Johnson, M.R., 2019. Live streamers on twitch.tv as social media influencers: chances and challenges for strategic communication. Int. J. Strategic Commun. 13 (4), 321–335. https://doi.org/10.1080/1553118x.2019.1630412.

Xu X-Y, Luo X(Robert), Wu K, et al. (2021) Exploring viewer participation in online video game streaming: A mixed-methods approach. *International Journal of Information Management* 58: 102297. DOI: 10.1016/j.ijinfomgt.2020.102297.

Zamora, R., 1981. Olympic Decathlon: The Software of Champions. InfoWorld.

Zhang, G., Hjorth, L., 2019. Live-streaming, games and politics of gender performance: the case of *Niizhubo* in China. Convergence Int. J. Res. New Media Technol. 25 (5–6), 807–825. https://doi.org/10.1177/1354856517738160.

Zhao Q, Chen C-D, Cheng H-W, et al. (2018) Determinants of live streamers' continuance broadcasting intentions on Twitch: A self-determination theory perspective. *Telematics and Informatics* 35(J. Pers. Soc. Psychol. 66 5 1994): 406–420. DOI: 10.1016/j.tele.2017.12.018.