

Research Article

Dorottya Cserző*

Discourses and practices of attention in video chat

<https://doi.org/10.1515/mc-2020-0010>

Received April 8, 2020; accepted April 18, 2021; published online May 5, 2021

Abstract: This paper examines the use of video chat (VC) with a focus on expectations and construction of attention. It is based on micro analyses of recorded VC sessions (gathered between 2013 and 2015) and thematic analysis of 29 semi-structured interviews about VC practices (conducted in 2014 and 2015). Building on multimodal (inter)action analysis (Norris, S. (2004). *Analysing multimodal interaction: a methodological framework*. Routledge, Norris, S. (2016). *Concepts in multimodal discourse analysis with examples from video conferencing*. Yearbook of the Poznan Linguistic Meeting 2: 141–165) and key concepts from nexus analysis (Scollon, R. and Scollon, S.W. (2004). *Nexus Analysis: Discourse and the emerging internet*. Routledge), I examine how focused attention is constructed in VCs and how these practices are shaped by experiences with other forms of communication. I demonstrate that unlike other forms of distance communication, typical VC encounters require a full investment of attention. This can be formulated as an interactional maxim: *focus your attention on the VC interaction*. I discuss how other activities can be interwoven with a VC and examine the exceptional practice of lapsed VC encounters (previously open connections or always-on video). I argue that participants display an orientation towards the maxim when pursuing other courses of action, and that lapsed encounters operate under a different value system than typical focused VC encounters. Finally, I reason that VC is reserved for close relationships because of the required investment of attention.

Keywords: attention; digital communication; distance communication; video chat

1 Introduction

Communication platforms are increasingly converging, with previously specialised apps now incorporating instant messaging, live video, live audio, recorded video, recorded audio, and photo sending into the same platform. Smartphones and tablets come with multiple communication apps pre-installed, all of which can perform the same core functions. Initiating a video chat (VC) is easier than ever before. Despite these developments, VC practices changed very little in terms of who was contacted through this medium – at least until the outbreak of COVID-19. In this paper, I examine perceptions and practices of VC to understand its role in the context of the rapidly changing technological landscape.

Video chat, the most widespread platforms being Skype and FaceTime (Longhurst 2017), has been a popular form of distance communication for over a decade (Ames et al. 2010; Kirk et al. 2010). I use the term ‘video *chat*’ to index interactions that take place in domestic contexts between partners, family, and friends. My participants used a variety of platforms (Skype, FaceTime, Google Hangouts, Tinchat, Viber, Snapchat, WhatsApp, and Facebook Messenger) on a variety of devices (desktop PCs, laptops, smartphones, and tablets). Some of them brought in their experiences of videoconferencing in a professional context, but the focus of this paper is on dyadic interactions where the primary goal is personal relationship maintenance.

*Corresponding author: Dorottya Cserző, CUREMeDE, Cardiff University, School of Social Sciences, 12 Museum Place, Cardiff, CF10 3BG, UK, E-mail: cserzodc@cardiff.ac.uk. <https://orcid.org/0000-0003-2676-9427>

The paper examines discourses and practices of paying attention in VC, comparing them to those surrounding other forms of distance communication (instant messaging and phone calls) and face to face communication. The discourse around attention frames it as a crucial resource that we trade in our interactions: in what has been described as the *attention economy* we can choose where to *invest* our attention, we *pay* attention to others and hope to *get* their attention in return (Goldhaber 1997; Jones 2005). Attention is also finite and relative, which means that if we pay more attention to a certain person or activity, we are deemed to pay less attention to another. Although we cannot easily quantify attention, we have a sense of what it means to pay *enough* attention. I explore the practices of constructing attention in recorded VC sessions and interviews about VC habits. I consider the attention costs of participating in such an interaction and contrast it with the costs associated with other communication tools. I argue that outside of work, VC is reserved for our most intimate relationships because it typically requires the greatest investment of attention. The discussion is built on data and literature from the pre-COVID-19 era, and focuses on discourses and practices that belong to that time, recognising that VC practices have been starkly altered. I reflect on the recent changes in Section 4.4 and the conclusion (Section 5).

This research project is unique within the field of VC studies in that it combines a micro analysis of recorded VC interactions with semi-structured interviews (the methods and data are discussed further in Section 3). Previous work has focused either on interviews (for example Longhurst 2017; Miller and Sinanan 2014) or on VC interaction at a micro level (Geenen 2017; Licoppe 2017; Norris 2019; Norris and Pirini 2017; Rosenbaun and Licoppe 2017), but not both. By drawing on nexus analysis (Norris and Jones 2005; Scollon and Scollon 2003, 2004) and multimodal (inter)action analysis (Norris 2004, 2016), I was able to systematically link observable chains of *lower-level actions* to discussions about VC practices.

VC studies have so far identified one maxim governing these interactions: *put the face of the current speaker on the screen* (Licoppe and Morel 2012). Like Grice's maxims of conversation (1989), this is not an unbreakable rule but rather a principle that participants use to generate and interpret meaning during interaction. If something other than the speaker's face is shown, there must be a reason for it; whether that is because a participant wanted to show her room or the VC device was moved for comfort. In this paper I identify a second underlying maxim which is dictated by the attention economy: *focus your attention on the VC interaction*. The second maxim does not prohibit participation in other activities (for example eating or tidying) as long as they can be interwoven with the VC smoothly. Nonetheless, I show that VC users hold each other accountable for keeping the VC at the foreground of their attention during the interaction.

2 Investing attention in distance communication

In this paper I take an interactional view of attention as it is laid out in multimodal (inter)action analysis (Norris 2004, 2016, 2019; Norris and Pirini 2017; Pirini 2016). In this approach the basic analytical unit is the *lower-level action*, which, chained together with other *lower-level actions*, constructs a *higher-level action*. A *lower-level action* is defined as the smallest meaningful unit, for example an utterance, gesture, or a shift in posture or gaze direction. Chains of *lower-level actions* are combined to form a recognisable *higher-level action* such as a conversation or dinner with friends. There are always multiple *higher-level actions* running in parallel, which are allocated different degrees of attention. By analysing the different communicative modes (such as speech, proxemics, gaze, gestures, posture) we can determine whether an action is in the foreground, midground, or background of a participant's attention.

This approach is well suited to examine the use of communication technologies, because their use is so closely intertwined with other higher-level activities. Ethnographic studies by Jones (2004, 2005, 2010) show that instant messaging (IMing) usually involves switching between multiple chat windows as well as other activities such as listening to music, browsing the internet, and sending photos. Jones (2004) suggests that the attraction of instant messaging is precisely that users are able to engage in multiple activities at the same time, *while displaying appropriate attention in multiple interactions*. In other words, IMing requires a relatively low investment of attention, which makes it easy to integrate with other activities.

Phone calls demand more attention than IMing. Some activities that commonly accompany IMing, such as listening to music or talking to other people, would be very disruptive during a phone call. There is also evidence that unlike in IM exchanges, people expect to maintain joint attention for the entire duration of the phone call (Baron 2008; Rettie 2009). However, since people cannot see each other while talking on the phone, it is possible to carry out some activities (such as walking, cleaning, or checking emails) without alerting the other person and/or disrupting the conversation (Kirk et al. 2010).

We carry in our *historical bodies* (Scollon and Scollon 2003, 2004) numerous experiences of sitting in front of our screens, of holding our phones, shifting our attention between different conversations, sources of entertainment, or tasks. However, during a VC it is almost impossible to conceal such shifts in attention (Ames et al. 2010; Kirk et al. 2010; Miller and Sinanan 2014, p. 154). There is a tension between the desire to engage in other activities during the VC, and the expectation of getting the full attention of the VC partner. Consequently, VC users must carefully negotiate their involvement in other activities with their VC partners or refrain from pursuing parallel activities altogether.

In the discussion (Section 4), I examine the extent to which participants aligned with these expectations in the interviews and the patterns that are observable in the video data. I argue that instances where expectations around attention are relaxed still display an orientation towards the proposed maxim (*focus your attention on the VC interaction*) in various ways such as body positioning and verbal accounts. Before these arguments can be laid out, I give a brief overview of the methods and data informing the study.

3 Methods and data

The aim of the study was to explore how VC is used for maintaining personal relationships. I decided to focus on university students, because many of them have moved to a new city for their studies, giving them strong incentive to use VC with friends and family. Furthermore, university students are also likely to have access to VC devices. Focussing on young adults living independently made it possible to complement findings of previous VC studies, which highlighted the experiences of mothers and grandparents (Ames et al. 2010; Longhurst 2013). The videos feature dyadic interactions with occasional appearances from additional people and the interviews also focus on dyadic VC sessions. Data was collected between 2013 and 2015, at a time when front-facing cameras were already common among my participants, but mobile data packages were still relatively expensive. This provided a unique insight into fast-changing practices: my participants used VC on their phones, but only in places where they had access to wi-fi. Some reported using VC on the go, but these anecdotes were framed as exceptional. Attitudes have undoubtedly changed now that cheap unlimited mobile data has made VC devices truly portable; it is becoming more commonplace to see people engaging in VC in public spaces while for example walking or shopping.

The research project is built on a micro analysis of recorded VC sessions and thematic analysis of interview data. Insights from the two types of analysis were consolidated using key concepts in nexus analysis (Scollon and Scollon 2003, 2004) and the closely related mediated discourse analysis (Norris and Jones 2005). A total of 29 VC interactions featuring 44 participants were analysed using an approach based on multimodal (inter)action analysis (Norris 2004, 2016) to study speech, gaze, posture, gesture, and camera movement. All video recordings were imported into Transana, where I viewed and annotated them tracking shifts in topic, all visible activities (including for example drinking or eating), interruptions, and sequences where talk did not play the primary role (for example during video chat tours, analysed in Cserző 2020). By tracking the modes of talk, proximity, posture, gesture, head movement, gaze, and screen layout I was able to identify changes in the *modal density* of the interaction (Norris 2004, 2016). Such changes occur when one of the modes comes into focus or becomes less prominent. Sequences where there was a change in *modal density* were transcribed and analysed. Throughout Section 4, I report the patterns found across the data with illustrative examples from the videos.

In order to find out about participants' VC habits, preferences, and their perceptions of the medium, I also conducted interviews with 29 participants (including six participants who also provided video recordings). The audio recordings of the interviews add up to over 11 h of data. The average length of the interviews was 22 min

with the shortest one being 11 min and the longest lasting for over an hour. I asked participants about the frequency of their VC sessions, how long they have been using VC, and who they talk to on VC. We also discussed how and whether VC sessions are prearranged, the possibility of having a spontaneous VC session, and the spaces that are most suitable for conducting VC. The interviews were transcribed and analysed through inductive qualitative coding (Gibbs 2007; Mason 2002) using NVivo (version 10 and later 11), resulting in lists of descriptive codes. The final coding scheme resulted in 28 codes, the two most prevalent ones being *space* and *attention*. These high-level codes were closely related to a number of more specific codes: *space* was linked to *location*, *camera space*, *distance*, *background*, *screen*, *public space*, and *virtual tour*; while *attention* was linked to *digital mirror*, *comparisons with other communication technologies*, *developing habits*, *multi-party interactions*, *showings*, *privacy*, and *intrusions*.

In the next stage, the findings of the interview analysis were brought together with the micro-analysis of the videos using central concepts from nexus analysis (Norris and Jones 2005; Scollon and Scollon 2003, 2004). The video recordings allowed me to observe the *interaction order* (Goffman 1963; Scollon and Scollon 2003, 2004): the way in which participants take up and maintain relationships with each other during a social encounter, including the way they structure their attention during the VC. Further context was provided by discussions about relationships in the interviews, making apparent who is invited to participate in this type of *interaction order*.

The interactions were carried out using the *mediational means* of physical space, objects, body parts, language, and so on (Bernal 2008; Scollon 2001). The *mediational means* are recorded in the videos, but they also circulated through our discussions in the interviews. In my analysis I consider the *affordances* (Hutchby 2001) of the *mediational means*, that is, the kinds of actions are amplified or restricted by them. For example, I discuss the implications of using a PC-based VC platform (such as Skype) or a smartphone-based VC platform (such as WhatsApp) in terms of contact list management.

The comparisons with other modes of communication in this paper are driven by the comparisons participants made in the interviews. These conversations were a rich source of information about their *historical bodies* (Scollon and Scollon 2003, 2004), their previous experiences with other technologies and their VC habits. The interviews reveal that their VC practices are informed by experiences with instant messaging, phone calls, and face to face communication.

The research project was granted ethical clearance by the relevant institution and followed established practices in the field of language and communication research (The British Association for Applied Linguistics 2016). All names in this paper are pseudonyms, and tracings are used in the place of screenshots for participants who did not wish for screenshots to be used in publications. These tracings preserve the analysed features while obscuring identifiable details. Participation was voluntary, and participants were free to withdraw their contributions at any time. Participants signed consent forms before the interviews and recording videos and discussed the publication and dissemination of research materials with the researcher.

Videos were recorded by the participants using freely available screen recording software (Debut Video Capture) on their laptops in 2013 and 2014. The videos only feature the use of Skype, but in the interviews participants also mentioned using FaceTime, Google Hangouts, Tinchat, Viber, Snapchat, WhatsApp, and Facebook Messenger for their VC sessions. In terms of devices, in the interviews, participants discussed using desktop PCs, laptops, smartphones, and tablets, but the video recordings only featured laptops, one PC, and one smartphone. The video data is presented primarily through verbal accounts with illustrative images in order to situate the analysed extracts within the context of the whole interaction in an economic manner.

Twenty interview participants were full time undergraduate students between the ages of 18 and 25. These interviewees had very similar living arrangements (occupying a single room in a shared house) and comparable VC habits, especially in terms of location. The other nine participants had a range of occupations (in full time employment, studying and working part time, on maternity leave, and one retired participant) and different living arrangements (for example living in a shared house, with a partner, or with young children). At the time of the interviews, all participants had been using VC for at least two years. The participants quoted in this paper are all undergraduate or postgraduate students between the ages of 21 and 31. Further details about the quoted participants can be found in the appendix.

4 Analysis and discussion

4.1 Video chat sessions as focused encounters

Interactions where people orient to maintaining a single joint focus of attention can be described as *focused encounters*. This term was coined by Goffman and originally referred to face-to-face communication which happens “when persons gather close together and openly cooperate to sustain a single focus of attention, typically by taking turns at talking” (Goffman 1963). However, it has been argued that phone calls are also treated as focused encounters (Baron 2008; Rettie 2009) in the sense that people expect to maintain joint attention for the duration of the call. Here, I apply this concept to VC interactions, demonstrating that VCs typically entail substantial investments of attention from all participants.

The analysis of the video recorded VC sessions supports the idea of VC sessions as focused encounters. This is created by maintaining *high modal density* (Norris 2004, 2016) throughout the majority of the interaction through conversation and adhering to the talking heads arrangement (Licoppe and Morel 2012). The talking heads arrangement (shown in Figure 1A below) leaves very little room for changes in posture and gaze. All a participant needs to do is turn their head (Figure 1B) or lean forward (Figure 1C) to change the configuration in a very prominent way. There was only one video where the talking heads arrangement was altered for longer stretches of time (for about half of the VC). This exceptional video is discussed in Section 4.2 below. In all other videos, participants did not deviate from the talking heads arrangement for longer than a minute at a time. Deviations were primarily prompted by interruptions via the phone or other people entering the space. The other main reason for departing from the talking heads arrangement was to leave the room to fetch something or use the bathroom. Finally, participants also altered their position when opening up files or websites;



A

Figure 1A: Talking heads

Ray: yo yo whassup man

As the video feed goes live the two men smile and greet each other. Their faces are clearly visible and they are looking directly at their screens.



B

Figure 1B: An interruption

Ray turns away from his screen, in the direction where his partner is (off camera). Paul continues to look at his screen.



C

Figure 1C: Paul briefly leaves

Paul: I'll just grab a water yeah

Paul turns his back to the camera as he exits through the door at the left of the screen. Ray is hunched close to the screen (he is reading) and his mouth is off camera.

however, the recordings show that these files were central to the ongoing interaction (for example jointly working on a Word document or sharing a photo). In contrast, talking to other people or leaving the room entails a temporary suspension of the VC interaction.

Overall, the videos were characterised by focused interaction with a few short breaks. For example, in the longest VC (which lasted for 1 h 18 min) the participants, Paul and Ray, shifted their focus of attention to something other than the VC in only five instances. For the remainder of the time, they were sat in front of their laptops and engaging in conversation with each other. Changes in their focus of attention are listed below:

2 s – Both video feeds go live, Paul and Ray exchange greetings (shown in Figure 1A).

2 min – Ray is in the middle of a narrative when suddenly he says ‘Hold on 1 s’, turns to the side, and calls to his partner ‘Yeah, Tessa, come in’. He holds this position silently for 4 s (shown in Figure 1B), but Tessa is not audible on the recording and she does not appear on screen. Ray turns back and picks up the narrative where he left off.

14 min – Paul leaves the room to get water (shown in Figure 1C), returning 20 s later. While he is away Ray continues to read something on his screen, presumably the paper they are discussing before Paul leaves and after he returns.

47 min – Tessa walks on screen and has a 30 s exchange with Paul. Initially she walks through in the background, disappearing quickly. However, Paul greets her and asks how she is doing. She comes back in view of the camera, they exchange a bit of small talk, and she announces that she is going to bed. Paul wishes her goodnight and promises not to keep Ray for too long and Tessa leaves the room.

55 min – Paul receives a chat message asking if he is available, he writes back to say he is still talking to Ray.

1 h 8 min – As Paul is talking about his plans, Ray stops him to go to the bathroom and tells Paul not to go. Ray is gone for 50 s, during this time Paul also goes off camera for 5 s but he returns before Ray. When Ray comes back Paul finishes telling him about his plans.

1 h 18 min – Paul hangs up and stops the recording.

Relevant screenshots are presented below with the corresponding talk (if any) and a brief image description.

Figures 1B and C show examples where participants signal shifts in attention through their posture. Another important source of information was the layout of the recording participant’s screen. The standard arrangement was to display the VC window on the full screen, and deviations from this layout were noted during the video analysis. There were only six instances where participants opened windows that were unrelated to the VC across the entire data set. One example comes from Paul, who responds to an incoming chat message during his VC with Ray. The coordination of these side activities with the ongoing VC is a site for future research; the point here is that the scarcity of these examples provides further indication that the VC was in the foreground of the participants’ attention (Norris 2004, 2016) for the majority of the recording.

The interview analysis provided further indication that my participants view VCs as focused encounters. For example, Mark stated emphatically that compared to texting and IMing, VC and phone interactions demand a greater investment of attention:

Mark: I wouldn’t like to talk to someone [on VC] and not be giving them not my hundred percent of my full attention. If someone wanted to do that to me I’d be quite annoyed. On text-based mediums you kind of expect people to be talking to other people at the same time but with phone or Skype (...) you should give the other person your full attention if possible.

Most of my participants (16 out of 29 interviewees, including Mark) expressed similar opinions, explicitly stating that they prefer to focus on the VC and/or expect the same from their partners. Two further participants mentioned that although they would happily engage in other activities during the VC, their conversational partners ask them to focus on the interaction. These accounts indicate that there can be disagreement on whether other activities are appropriate during a VC, which I explore further in Section 4.2. Furthermore, they also demonstrate that VC users are held accountable for dividing their attention. Three further participants mentioned that they might ask the other person to move somewhere else if they feel that they are

“too distracted” by their surroundings. In the most extreme case, April even rescheduled the VC because she wanted to discuss personal topics with her best friend, who was distracted by other people in the room and incoming phone messages. These comments indicate that it is not enough to foreground the VC interaction: a focused VC encounter requires users to choose locations and times where distractions can be minimised.

In line with previous research, my analysis of the videos and interviews indicates that in VC sessions there is an expectation to keep foregrounding the VC interaction. This expectation is stricter than in the case of face to face meetings: Norris (2004) gives numerous examples of the complexity of face to face meetings (for example eating in a cafeteria with a friend; having a chat with a colleague at work; or even balancing meeting a friend, writing a shopping list, speaking on the phone, playing a game, watching the children, and interacting with the researcher) and Scollon (2001) has examined the interweaving of chains of actions between friends meeting in a coffee shop. In contrast, during a VC even turning away from the screen is a marked and accountable action, as shown above. However, it is clear that in practice the expectation to focus on the VC is not always met. In the following sections, I examine the conditions under which parallel activities may be carried out. Firstly, participants may engage on parallel *higher-level actions* during a VC (4.2). In some exceptional cases, participants can pursue independent courses of action, backgrounding the VC for long stretches of time (4.3).

4.2 Interweaving higher-level actions

The recorded VC sessions captured one example where a participant was engaged in another *higher-level action* (Norris 2004, 2016) throughout the entire VC. In this video Sian calls Tracy to interview her for a piece of coursework, and she is surprised to find that Tracy is in the kitchen. Tracy reveals that she is cooking, which she continues to do while answering the interview questions. In a detailed analysis of this interaction (Cserző 2016) I argue that the first maxim of VC (Licoppe and Morel 2012) is relaxed in order to accommodate the requirements of cooking: Tracy remains “on camera”, although her face is not always visible and she stands relatively far away from her laptop. In this interaction, Tracy switches back and forth between a position indicating primary involvement with the VC (Figure 2A) and one which allows her to attend to the cooking (Figure 2B).

When Tracy is standing at the hob, her face is not always visible as her head leaves the range of the camera (for example in Figure 2B). However, she later adjusts the camera so that her face remains in the frame even when she’s standing at the hob (Figure 3).

In this case, cooking and participating in the VC can be intertwined relatively easily because the two activities draw on different modes, similarly to the case of talk at the hair dressers’ (Stefani and Horlacher 2017). Cooking and cutting hair require manual manipulation, leaving the participant free to engage in conversation at the same time. Throughout the interaction, Tracy shifts her focus of attention between cooking and the VC interview. However, even when focussing on the cooking, she provides appropriate and timely responses and remains at least partially within view of the camera. This indicates that the VC is in the

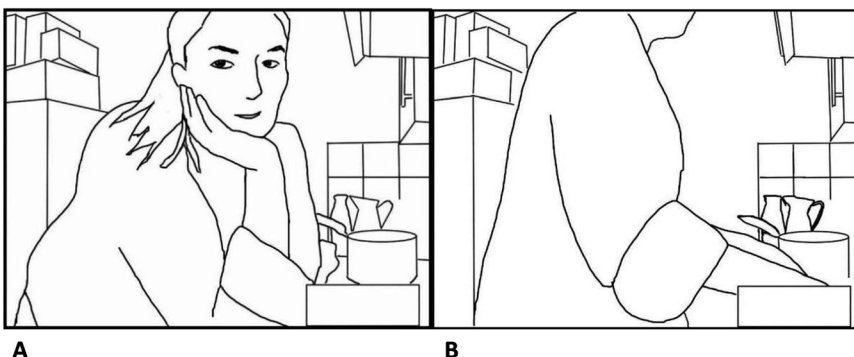


Figure 2: A. Tracy leaning on the counter. B. Tracy standing at the hob.



Figure 3: Tracy standing at the hob after adjusting the screen.

midground if her attention when cooking is foregrounded (Norris 2004, 2016). Tracy shifts her position eight times and she splits her time between the position at the hob and the position at the laptop almost equally, spending 6 min and 20 s in total at the hob and 6 min and 9 s in total close to the laptop. This is in sharp contrast to the instances where participants engage in other activities on their screens, quickly read and reply to incoming messages or briefly open up web pages unrelated to the conversation without altering their posture.

In Section 4.1, I summarised responses from participants who approach VC sessions as focused encounters. For these participants, a “good” VC session is one where both parties dedicate their attention to the interaction. However, eight participants indicated that the requirement to focus on the VC can be overly demanding and restrictive. For them, the VC can be more enjoyable if they are free to pursue other activities during the conversation. For example, Gemma reported that she likes to have the freedom to move about when she is talking to her mother or sister on VC. She contrasts these with “an intense conversation” where she is not allowed to leave the desk.

Gemma: If I want to sit down and have like an intense conversation then I can. But it's also nice to know that it can just be kind of relaxed like I can still facetime and have the conversation face-to-face, but I don't have to be completely tied to sitting on my chair at my desk. I can move around in my room, I can you know do little like bits and bobs while I'm talking, which is what I would do at home. So I suppose it is good to not have that intensity, to have like a bit of option in what I wanna do as well as talking.

Gemma's account and the other similar accounts contain many caveats. Firstly, like in the cooking example, the intertwined activities must be compatible with maintaining a conversation. This means that they cannot be too noisy, and they should not take the participants too far from the VC device. Secondly, engaging in parallel activities was mentioned in the context of close relationships (family members or partners). Finally, such VC interactions were presented as a practice that developed over time: interviewees described a “relaxing” of their initial approach. With an accumulation of VC experience in the *historical body*, it appears that more leeway can

be given in the *interaction order* in well-established relationships and participants may gain more freedom to deviate from the talking heads arrangement. In the most extreme cases, this can lead to a new type of *interaction order*, which is the lapsed VC encounter.

4.3 Lapsed video chat encounters

In the interviews, six of my participants described using VC in a way that breaks all the rules of focused encounters: they withdrew from the joint attentional frame for relatively long periods of time to go about their day (for example to watch TV, do housework, or study); however, they maintained the connection, which allowed them to remain aware of each other's presence and summon each other back. Borrowing another of Goffman's (1963) terms, I refer to these interactions as *lapsed encounters*. Similar practices have been reported under the label *open connection* (Kirk et al. 2010; Neustaedter et al. 2015), *always-on video* (Miller and Sinanan 2014; Rosenbaun et al. 2016), or *virtual co-presence* (de Fornel and Libbrecht 1996; Develotte et al. 2010; O'Hara et al. 2006).

Goffman uses the term *lapsed encounter* to refer to people who are considered to "be together" and have the right to rapidly summon each other's attention for example while walking silently, dozing on the beach, or staring at the fire (1963, p. 102–103). This is a different type of *interaction order* from focused encounters because participants are free to engage in solitary activities while still maintaining some form of companionship. The term *lapsed encounter* can be applied to VC because by leaving the software running, VC users make it possible to restart communicating abruptly. In the silences between these exchanges they can remain aware of each other's activities despite the physical distance between them. The term *lapsed encounter* also highlights the boundedness of these interactions: VC users must engage in focused interaction before and after the lapsed encounter. It is currently impossible to start a VC without focused interaction, and, as one participant pointed out, ending the VC without saying goodbye would be considered rude.

Studies of VC in long distance romantic relationships have argued that lapsed VC encounters can create a form of virtual co-habitation (Kirk et al. 2010; Longhurst 2017, p. 111; Miller and Sinanan 2014, p. 57; Neustaedter et al. 2015). Similar sentiments were expressed by two of my participants (Camille and Bryn) talking about using VC with their partners, and interestingly also by April talking about VC sessions with her sister.

April: When I skype with my sister it's not for the purpose of having a conversation that much. Like when she calls me we talk about stuff that's going on (...) But there comes a point while I'm talking to her where there's just no need to talk. (...) we're just comfortable enough to stay connected even though we're not paying attention anymore. So it's like an unspoken agreement between the both of us that it's okay now to shift your attention to something else. But I think that is very special and limited to certain relationships and also to the fact that when I skype with my sister it's probably late in the evening, she's not doing anything else, she's just watching TV. And I'm not doing anything else, I'm just sitting there doing something and that's just it. It feels very natural because that's like the moments when we sit together and we talk when I'm at hers. And then we watch TV and we talk about something for 2 min and then we do something else again.

April, like Camille and Bryn, talks about having *lapsed encounters* via VC with one specific person only. There is a "natural transition" from focused encounters into *lapsed encounters*, which evokes the experience of cohabitating or physical visits. April describes a 'comfortable silence', one where there is no need to talk. In this context, silence is viewed in a positive light, because it is proof that there is no pressure to talk within the relationship (Jaworski 2000). Thus, paying less attention contributes to a feeling of intimacy, which is at odds with the *interaction order* of focused encounters.

Silences have divergent meanings during focused and *lapsed encounters*. In focused VC encounters, long silences are problematic because they mark an absence of response. However, in lapsed VC encounters participants give each other permission to relegate the VC to the background of their attention. In physical presence and via VC, long silences are acceptable (under the right circumstances) because presence is expressed by other means. In phone calls, where the only way to express presence is verbally, silence is

unacceptable (Rettie 2007, p. 42). Therefore, by transmitting live video, the *mediational means* of VC has made it possible to participate in a synchronous but loosely structured interaction.

Lapsed encounters also occur in other contexts, where instead of creating a sense of co-habitation the goal is to remain connected while working, for example during study sessions (Kirk et al. 2010; Miller and Sinanan 2014, p. 57), or at the work-place (Fish et al. 1993; Heath and Luff 1992). Three of my participants also reported using VC in this manner, keeping a VC window open while studying or writing fiction, activities which are incompatible with maintaining focused interaction. In such VC sessions, participants shift back and forth between working individually and engaging each other in conversation for several hours.

The recorded VCs did not include such *lapsed encounters*. Although the joint attentional frame was visibly suspended on occasion, these suspensions were brief compared to the length of the VC, and participants displayed an orientation to resuming the interaction as quickly as possible. When Ray turned his face away to talk to his partner (Figure 1B), his shoulders were still facing the laptop, displaying a continued orientation to the backgrounded VC, and he resumed his conversation with Paul within seconds. When Ray left to use the bathroom, he returned within a minute and the conversation continued. This is in contrast with the accounts of *lapsed encounters* which depict longer VC sessions *characterised by* interactional lulls rather than focused interaction *punctuated by* brief suspensions.

Only a small group of my participants (six out of 29) reported habitually engaging in *lapsed encounters*, and their accounts indicate that such VC sessions occur only with a few specific people. This is consistent with the findings of previous studies, but contradicts the prediction (Miller and Sinanan 2014, p. 57) that *lapsed encounters* (in their terms “always-on webcam”) would eventually become the dominant way of using VC. This is not the case, and I suggest that there are both social and technological reasons for this. Socially, there are few relationships in which *lapsed encounters* are meaningful and desirable. In addition, experiments with always-on VC systems have highlighted that such usage requires a dedicated VC device, which is at odds with the multifunctional purpose of the devices used for VC today (Neustaedter et al. 2015). Cheaper purpose-built VC-only devices may bring about a change in attitudes towards *lapsed encounters*, but at the moment this type of *interaction order* is still the exception and not the rule. Domestic VC is not truly “always-on” but rather an interaction in which participants can move from focused interaction to a *lapsed encounter*. The *mediational means* make it possible to participate in *lapsed encounters*, but they are designed to facilitate focused interactions.

4.4 Who is contacted via video chat?

The mere act of starting or accepting a VC session makes a statement about the relationship between the participants, because not everyone is contacted through this medium. Before COVID-19, VC was a ‘niche’ communication tool in the sense that it was reserved for special occasions, and/or close relationships (O’Hara et al. 2006). The association of VC with intimate relationships is also implicit in other studies, which focus on the use of VC in long distance relationships (e.g., Longhurst 2017; Rintel 2013) or between parents, children, and grandparents (Kirk et al. 2010; Longhurst 2013). In response to COVID-19, there has been a huge increase in the use of video-conferencing in a work context (Nellis and Menn 2020). Domestic VC use has also increased, although the extent of this change is still being investigated and it remains to be seen how practices evolve if restrictions on travel and social contact are lifted.

Similarly to the reports in the studies cited above, my participants described regular VC sessions with specific people (partners, family, friends). These VC partners were people with close relational ties living in distant locations. VC was a means to maintain relational closeness during periods where physical meetings were not possible, in their case due to distance. The main difference today is that physical meetings may not be permissible, no matter how close we may be located. Therefore, there is now a strong incentive to use VC with people we would previously meet in person.

At the time of data collection and initial analysis, I observed a large gap between the number of people that could potentially be contacted via VC and the number of people that were regularly contacted in practice. In

contrast to the early days of Skype when contacts were added one-by-one, smartphone apps (such as WhatsApp, Viber, or FaceTime) automatically connect to the various contact lists stored on the phone. The *affordances* (Hutchby 2001) of the platforms encouraged connectivity, but the potential reach of VC was not fully exploited by users. It took a global pandemic for VC to become a truly ubiquitous tool. However, increased VC use has also spurred public discussions of ‘zoom fatigue’ (Jiang 2020): the exhaustion felt after participating in video calls. While the year 2020 has changed who we talk to via VC, it has also shed light on the additional cognitive effort required for participating in a VC compared to face-to-face interactions.

5 Conclusion

This paper has examined attention distribution in dyadic VCs between young adults in the mid-2010s through a micro-analysis of recorded VCs combined with insights from qualitative interviews. The benefit of the combined approach is that I was able to draw systematic links between discourses and practices. Attention was identified as a central concern in the interviews, reflecting wider discourses about attention as a valuable and limited resource. The analysis of the recorded VC sessions revealed the interactional work that is required to maintain a focused VC interaction in terms of remaining in view of the camera and refraining from engaging in other activities. Furthermore, the interviews clarify that the interactional work starts before the VC session: participants work to choose a time and location where disruptions can be minimised. As evidenced by the interviews, for some young adults this interactional work feels very demanding.

In the analysis I take an interactional view of attention, as conceptualised in multimodal (inter)action analysis (Norris 2004, 2016). This approach makes it possible to account for multiple simultaneous courses of action, making it particularly well suited for analysing interaction in environments that are rich in potential distractions, such as a video chat. In addition to speech, gaze, and posture; the analysis of which is well-established within multimodal (inter)action analysis; I also examined camera movement and screen layout (see also Cserző 2016, 2020), two modes that are central to VC interactions. A multimodal (inter)action analysis (Norris 2004, 2016) of these modes together with the interview analysis has illuminated the types of actions which can be smoothly interwoven with VC interactions.

The expectations surrounding attention during a VC can be expressed through a maxim: *focus your attention on the VC interaction*. This maxim applies to all participants, regardless of their momentary status as speaker or listener. Like Grice’s maxims of conversation (1989), the maxim is not always adhered to. However, such cases are framed as exceptional in the interviews and the videos show that deviation from the maxim requires further interactional work (i.e.: giving an account for shifts in the focus of attention). All interactions entail some degree of attention, but in other forms of distance communication and in physical presence there is greater license to move the body, delay a response, and attend to unrelated activities. What is more, VC interactions are prone to disruptions through people entering the space or incoming messages on the very screen mediating the encounter. Since the attention costs of these *mediational means* are so high, it is no surprise if participants expect a return on their investment. After all, if the interlocutor’s attention is occupied elsewhere, you might as well switch to instant messaging.

The *affordances* (Hutchby 2001) of VC have also made it possible to engage in the counter-normative practice of lapsed VC encounters. In such interactions, participants withdraw from the joint attentional frame but continue to allow themselves and their environment to be observed and can be summoned to shift their focus back to the VC partner. When participating in *lapsed encounters*, VC users can engage in activities that are incompatible with focused interaction (such as studying, writing, or watching TV). These *lapsed encounters* are the clearest examples of VC as a self-contained meaningful act: in the absence of words, gestures, or images, all that is exchanged is the live video feed, without the pressure to communicate. Such practices have been reported in previous VC studies; however, they have not yet been studied in detail. In this paper I have examined their relation to focused VC encounters based on the accounts from the VC user interviews. Further insights could be gained by analysing naturally occurring *lapsed encounters* in VCs.

VC can be intertwined with other activities if participants agree on the *interaction order*. In *lapsed encounters*, the relationship is so important that there is value in keeping the channel of communication open for hours at a time just to make interaction possible. This is in stark contrast with focused encounters, where pursuing other courses of action is taken as a lack of investment in the interaction. For some participants, any sign of diverted attention is undesirable. For others, engaging in parallel activities requiring little attention (for example tidying) makes the interaction more enjoyable. Activities can be smoothly intertwined with VC if they do not interfere with the modes that are central in VC: speech, posture, gaze, and facial expressions. Actions that create noise or take the participant too far away from the device are not compatible with a focused VC interaction.

The use of video chat and video conferencing has massively increased in the wake of COVID-19. However, the changes brought about in 2020 also provide support for the arguments laid out in this paper. Spending more time in video calls has made the attention costs of this mode of communication more visible to the public. Although we could easily reach hundreds of people through VC, our capacity to pay focused attention has not increased. Therefore, in our personal lives we reserve this costly mode of communication for those who most deserve our attention.

Acknowledgements: I would like to thank the anonymous reviewers for their insightful comments. I also want to thank my generous mentors and colleagues who have given me feedback on earlier drafts: Virpi Yläne, Jaspal Singh, and Ruth Page. Any remaining inadequacies are my own. This research would not have been possible without financial support from The Sidney Perry Foundation, The Gen Foundation, The Sir Richard Stapley Educational Trust, The Humanitarian Trust, and The Allan and Nesta Ferguson Charitable Trust and the time and trust of my participants.

Appendix: Participant descriptions

Mark is a 22 year old undergraduate student from the UK. He described his use of VC as ‘infrequent’, about once every 3 months. There was a 5 month period when he was travelling overseas and used VC about twice a month. He first tried VC at around the age of 14. He uses VC primarily with friends and occasionally family.

Gemma is a 21 year old undergraduate student from the UK. She uses VC about 3–4 times a week. She started using VC when she moved away from home for her university course. She uses VC to keep in touch with her mum, her sisters, and her boyfriend.

April is a 27 year old postgraduate student from Germany. She uses VC at least 5 times a week, often daily. She has been using VC for 7 years. She keeps in touch with her friends, family, and her boyfriend through VC (among many other forms of communication). She also uses videoconferencing for work purposes.

Camille is a 30 year old undergraduate student from the UK. In addition to her studies, she also works in the hospitality industry part-time. She uses VC every week and has first used VC about 3 years ago. She talks primarily with her new boyfriend, who lives in another city, and some friends who live far away.

Bryn is a 31 year old postgraduate student from the UK. He provided three recordings of his VC sessions with his boyfriend and participated as an interviewee. Bryn first used VC 5 years before the interview, and was initially using it about every 6 weeks for work purposes. This changed when he got a new job in a different city and he started using VC with his boyfriend every day when they were apart. The recordings were made during this period, but by the time I interviewed him they were living together again and Bryn reverted to using videoconferencing for work as required.

Sian provided a video recording but she was not interviewed. The VC was recorded as part of the coursework for an undergraduate module on Digital Literacies. Students were asked to conduct two interviews: one via VC and one via IM. They summarised the responses and reflected on the differences between the two experiences. Recordings were compulsory for the module, and students could opt in to share the recordings for

research purposes. **Sian** interviewed her friend **Tracy** who lives in a different city. In addition to answering Sian's interview questions they also discussed personal matters and made plans for future events.

References

- Ames, M.G., Go, J., Kaye, J., and Spasojevic, M. (2010). Making love in the network closet: the benefits and work of family videochat. *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work - CSCW '10, May 2014*, p. 145.
- Baron, N.S. (2008). Controlling the volume: everyone a language czar. In: *Always on: language in an online and mobile world*. Open University Press, London, pp. 31–44.
- Bernal, M. (2008). Do insults always insult? Genuine Impoliteness versus non-genuine Impoliteness in colloquial Spanish. *Pragmatics* 18: 781–802.
- Cserző, D. (2016). Nexus analysis meets scales: an exploration of sites of engagement in videochat interviews. In: Singh, J.N., Kantara, A., and Cserző, D. (Eds.). *Downscaling culture: revisiting intercultural communication*. Cambridge Scholars, Cambridge, pp. 337–365.
- Cserző, D. (2020). Intimacy at a distance: multimodal meaning making in video chat tours. In: Thurlow, C., Dürscheid, C., and Diémoz, F. (Eds.). *Visualizing digital discourse: interactional, institutional and ideological perspectives*. Mouton DeGruyter, Berlin, pp. 151–169.
- de Fornel, M., and Libbrecht, L. (1996). The interactional frame of videophonic exchange. *Réseaux. The French Journal of Communication* 4: 47–72.
- Develotte, C., Guichon, N., and Vincent, C. (2010). The use of the webcam for teaching a foreign language in a desktop videoconferencing environment. *ReCALL* 22: 293–312.
- Fish, R.S., Kraut, R.E., Root, R.W., and Rice, R.E. (1993). Video as a technology for informal communication. *Commun. ACM* 36: 48–61.
- Geenen, J. (2017). Show and (sometimes) tell: identity construction and the affordances of video-conferencing. *Multimodal Communication* 6: 1–18.
- Gibbs, G. (2007). Flick, U. (Ed.), *Analysing qualitative data*. SAGE, London.
- Goffman, E. (1963). *Behavior in public places: notes on the social organization of gatherings (First edit)*. The Free Press, New York.
- Goldhaber, M.H. (1997). The attention economy and the Net. *First Monday* 2, Available at: <http://firstmonday.org/ojs/index.php/fm/article/view/519/440> (Accessed 18 May 2012).
- Grice, P. (1989). Logic and conversation. In: *Studies in the way of word*. Harvard University Press, Cambridge, pp. 22–40.
- Heath, C.C., and Luff, P. (1992). Collaboration and control: crisis management and multimedia technology. *Cscw* 1: 69–94.
- Hutchby, I. (2001). Technologies, texts and affordances. *Sociology* 35: 441–456.
- Jaworski, A. (2000). Silence and small talk. In: Coupland, J. (Ed.), *Small group research*. Longman, Harlow, pp. 110–132.
- Jiang, M. (2020). *The reason zoom calls drain your energy*. BBC Worklife, Available at: <https://www.bbc.com/worklife/article/20200421-why-zoom-video-chats-are-so-exhausting>.
- Jones, R.H. (2004). The problem of context in computer mediated communication. In: LeVine, P., and Scollon, R. (Eds.), *Discourse and technology: Multimodal discourse analysis*. Georgetown University Press, Washington, D.C., pp. 20–33.
- Jones, R.H. (2005). Sites of engagement as sites of attention: time, space, and culture in electronic discourse. In: Norris, S., and Jones, R.H. (Eds.), *Discourse in action: introducing mediated discourse analysis*. Routledge, London and New York, pp. 141–154.
- Jones, R.H. (2010). Cyberspace and physical space: attention structures in computer mediated communication. In: Jaworski, A., and Thurlow, C. (Eds.), *Semiotic landscapes: language, image, space*. Continuum, London and New York, pp. 151–167.
- Kirk, D.S., Sellen, A., and Cao, X. (2010). Home video communication: mediating “closeness”. *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work - CSCW '10*, pp. 135–145.
- Licoppe, C. (2017). Skype appearances, multiple greetings and ‘coucou’. *The sequential organization of video-mediated conversation openings*. *Pragmatics* 27: 351–386.
- Licoppe, C., and Morel, J. (2012). Video-in-Interaction: “talking heads” and the multimodal organization of mobile and skype video calls. *Res. Lang. Soc. Interact.* 45: 399–429.
- Longhurst, R. (2013). Using skype to mother: bodies, emotions, visuality, and screens. *Environ. Plann. Soc. Space* 31: 664–679.
- Longhurst, R. (2017). *Skype: bodies, screens, space*. Routledge, London and New York.
- Mason, J. (2002). *Qualitative researching*, 2nd ed Sage Publications, London.
- Miller, D., and Sinanan, J. (2014). *Webcam*. Polity Press, Cambridge.
- Nellis, S. and Menn, J. (2020). *Demand for video calling continues to surge, Microsoft and others say*. Reuters, Available at: <https://www.reuters.com/article/us-health-coronavirus-software/demand-for-video-calling-continues-to-surge-microsoft-and-others-say-idUSKCN21R20P>.
- Neustaedter, C., Pang, C., Forghani, A., Oduor, E., Hillman, S., Judge, T.K., Massimi, M., and Greenberg, S. (2015). Sharing domestic life through long-term video connections. *ACM Trans. Comput. Hum. Interact.* 22: 1–29.

- Norris, S. (2004). *Analyzing multimodal interaction: a methodological framework*. Routledge.
- Norris, S. (2016). Concepts in multimodal discourse analysis with examples from video conferencing. *Yearbook of the Poznan Linguistic Meeting* 2: 141–165.
- Norris, S. (2019). Focused attention in focus: crossing micro-analytical boundaries. In: Gnosa, T., and Kallass, K. (Eds.). *Grenzgänge. Digitale Festschrift für Wolf-Andreas Liebert*, Issue July, pp. 1–13.
- Norris, S., and Jones, R.H. (Eds.) (2005). *Discourse in action: introducing mediated discourse analysis*. Routledge, London and New York.
- Norris, S., and Pirini, J. (2017). Communicating knowledge, getting attention, and negotiating disagreement via videoconferencing technology: a multimodal analysis. *Journal of Organizational Knowledge Communication* 3: 23.
- O'Hara, K., Black, A., and Lipson, M. (2006). Everyday practices with mobile video telephony. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 871–880.
- Pirini, J. (2016). Producing shared attention/awareness in high school tutoring. *Multimodal Communication* 3, <https://doi.org/10.1515/mc-2014-0012>.
- Rettie, R. (2007). Texters not talkers: phone call aversion among mobile phone users. *PsychNol. J.* 5: 33–57.
- Rettie, R. (2009). Mobile phone communication: extending Goffman to mediated interaction. *Sociology* 43: 421–438.
- Rintel, S. (2013). Video calling in long-distance Relationships : the opportunistic use of audio/video distortions as a relational resource. *Electron. J. Commun.* 23.
- Rosenbaun, L., and Licoppe, C. (2017). Showing 'digital' objects in web-based video chats as a collaborative achievement. *Pragmatics* 27: 419–446.
- Rosenbaun, L., Rafaeli, S., and Kurzon, D. (2016). Blurring the boundaries between domestic and digital spheres: competing engagements in public Google hangouts. *Pragmatics* 26: 291–314.
- Scollon, R. (2001). *Mediated discourse: the nexus of practice*. Routledge, London and New York.
- Scollon, R., and Scollon, S.W. (2003). *Discourses in place: Language in the material world*. Routledge, London and New York.
- Scollon, R., and Scollon, S.W. (2004). *Nexus Analysis: Discourse and the emerging internet*. Routledge, London and New York.
- Stefani, E. De, and Horlacher, A.-S. (2017). Mundane talk at work : multiactivity in interactions between professionals and their clientele. *Discourse Stud.* 1–25, <https://doi.org/10.1177/1461445617734935>.
- The British Association for Applied Linguistics (2016). Recommendations on good practice in applied Linguistics, Available at: https://www.baal.org.uk/wp-content/uploads/2016/10/goodpractice_full_2016.pdf.