



Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction

Lucas M. Seuren^{*}, Joseph Wherton, Trisha Greenhalgh, Sara E. Shaw

Nuffield Department of Primary Care Health Sciences, University of Oxford, United Kingdom



ARTICLE INFO

Article history:

Received 4 June 2020

Received in revised form 22 October 2020

Accepted 11 November 2020

Available online 3 December 2020

Keywords:

Video-mediated interaction

Turn-taking

Conversation analysis

Overlapping talk

Video consultations

ABSTRACT

Latency in video-mediated interaction can frustrate smooth turn-taking: it may cause participants to perceive silence at points where talk should occur, it may cause them to talk in overlap, and it impedes their ability to return to one-speaker-at-a-time. Whilst potentially frustrating for participants, this makes video-mediated interaction a perspicuous setting for the study of social interaction: it is an environment that nurtures the occurrence of turn-taking problems. For this paper, we conducted secondary analysis of 25 video consultations recorded for heart failure, (antenatal) diabetes, and cancer services in the UK. By comparing video recordings of the patient's and clinician's side of the call, we provide a detailed analysis of how latency interferes with the turn-taking system, how participants understand problems, and how they address them. We conclude that in our data latency unnoticed until it becomes problematic: participants act as if they share the same reality.

© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Research on video-mediated interaction has long recognized the problems that technology and a lack of shared physical space (the “fractured ecology”) pose for conversational participants (Heath and Luff, 1993; Hindmarsh et al., 1998; Luff et al., 2003; Rintel, 2013, 2015). In this article we are concerned with *latency*, the technology-generated transmission delay between when a participant produces an action and when the co-participant(s) perceive that action. The effects latency has on smooth turn-taking have featured centrally in a range of studies of social interaction (Olbertz-Siitonen, 2015; Ruhleder and Jordan, 2001; Rusk and Pörn, 2019; Schoenberg et al., 2014; Tang and Isaacs, 1993).

The problems participants experience with turn-taking in video-mediated interaction are of significant benefit to scholars of social interaction. The setting provides something of a *natural breaching experiment* (Garfinkel, 1967): provided the latency is long enough to have a noticeable impact on the interaction—noticeable to the analysts, not the participants—participants find themselves in an environment where unbeknownst to them, their background assumption that turn production and turn

^{*} Corresponding author. Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK.

E-mail addresses: lucas.seuren@phc.ox.ac.uk (L.M. Seuren), joseph.wherton@phc.ox.ac.uk (J. Wherton), trish.greenhalgh@phc.ox.ac.uk (T. Greenhalgh), sara.shaw@phc.ox.ac.uk (S.E. Shaw).

perception occur simultaneously no longer applies (Ruhleder and Jordan, 2001). As a result, they routinely have to solve basic interactional problems. Precisely because latency frustrates smooth turn-taking, we are provided with a treasure trove of phenomena and practices that are rare in instantaneous interaction.

In this paper, we provide a detailed analysis of how participants in video-mediated interaction manage turn-taking problems that happen as a result of latency. In doing so we not only document systematically how latency affects video-mediated interaction (i.e., its procedural consequentiality) (Arminen et al., 2016; Schegloff, 1991), but also contribute to the study of turn-taking and the practices by which participants address problems.

1.1. Turn-taking in co-present interaction

Turn-taking is a basic organizational principle of human social interaction (Levinson, 2016; Sacks et al., 1974). While the particulars of the turn-taking system vary between linguistic cultures, the underlying rules as described by Sacks, Schegloff, and Jefferson hold large cross-cultural sway (Dingemanse and Floyd, 2014; Stivers et al., 2009). Conversational participants organize their talk through sequences of actions in which the norm is one-speaker-at-a-time. The turn-taking system is arguably one of the few true universals of communication. Organizing talk in an orderly and systematic way is a basic condition for people to develop and maintain an intersubjectively shared understanding (Moerman and Sacks, 1988; Sacks et al., 1974; Schegloff, 1992).

The turn-taking system, in combination with the principles of sequence organization (Schegloff, 2007), has implications for our understanding of talk as well as our understanding of silence at points where talk should occur (i.e., non-talk) (Lerner, 2019; Sacks et al., 1974). Whilst some silence is tolerated (Stivers et al., 2009), upon completion of a turn in which the speaker has selected a participant to produce a specific next action (e.g., a question making relevant an answer), silence will often be understood as the noticeable absence of that action (Bolden et al., 2012; Goodwin, 1979; Lerner, 2003; Pomerantz, 1984; Schegloff, 2007; Stivers and Rossano, 2010). Speakers may treat such a *noticeable silence* as, for example, a refusal by the recipient to produce a next action (Schegloff, 2002) or as foreshadowing a dispreferred or non-straightforward response (Davidson, 1984; Kendrick and Torreira, 2015; Robinson, 2020; Sacks, 1987). Even when no next-speaker has been selected, silences between turns where talk should occur will be understood as the noticeable absence of talk: one of the participants should “self-select” to produce a turn (Hoey, 2020).

The turn-taking system is also used to minimize overlapping talk: to maintain the norm of “one-speaker-at-a-time” (Sacks et al., 1974). That is not to say that overlapping talk does not happen—it frequently does. Some forms of overlap are in fact highly affiliative, such as recognition overlap, which can be used to display strong agreement (Vatanen, 2018). However, participants often treat overlap as deviant from the norm and engage in interactional work to resolve it (Drew, 2009; Jefferson, 2004b; Schegloff, 2000, 2001).

1.2. Turn-taking and latency

People acquire the turn-taking system in co-present environments, in which turns will be heard and seen at the same time they are produced (Ruhleder and Jordan, 2001). However, when there is latency, action production and perception no longer co-occur, and this has implications for participants' ability to manage turn-taking (Olbertz-Siitonen, 2015; Rusk and Pörn, 2019; Schoenenberg et al., 2014; Tang and Isaacs, 1993).

Video communication technologies such as Skype inherently suffer from some latency. The length of this transmission delay varies: It can be anywhere in the order of tens to hundreds of milliseconds, depending on the quality of the network and each participants' local internet connection. These delays may seem short, but in social interaction where participants work with split-second timing (Jefferson, 1973) they quickly present problems. Take the following extract from a video consultation between a patient with heart failure and his specialist nurse, where at this point the latency varies around 750 ms both ways. The nurse is inquiring about the patient's medication.¹

¹ Throughout the paper we transcribe gaze of co-participants as directed at the screen, even though participants have no access to their co-participants' gaze orientation, because the camera and screen are not in the same place. For example, in (1a), the nurse cannot know that the patient is looking at the screen, and not something else in his environment. This “Mona Lisa Effect” means that participants cannot precisely track their co-participants' gaze or gestures (Luff et al., 2016). Our way of transcribing therefore partially represents our interpretation of the data: we ascribe an understanding to participants that their co-participant is looking at the screen (i.e., at them). However, in all but one of our recordings (see 4a–4b) the camera is placed adjacent to the screen, either because it is built-in to a smartphone, tablet, or laptop, or because it is placed on a computer monitor. The slight “off gaze” seems to consistently be understood as gaze orientation towards the screen. An analysis of the role of gaze for turn-taking, particularly when the camera is not placed directly adjacent to the screen, is beyond the scope of this paper.

Excerpt 1a – Nurse and patient discussing medication – Clinic

01 Nurse: \$yea:h_i
 ns \$>>gz to notes-->
 pt &>>gz to scrn-->
 02 (0.4)
 03 Nurse: .h [a:]:nd
 04 Patient: [yep.]
 05 (.)
 06 Nurse: &you w- \$and is ↑tha\$t in one go&#?:?
 -->\$turns----->\$gz to scrn-->
 pt &gz to left-----&turns-->
 fig #fig1
 07 (0.4)&#(0.8)
 pt -->&gz to scrn-->
 fig #fig2
 08 Nurse: m: is that in [one g]o\$%-yeah,
 09 Patient: [yes:.]
 ns -->\$gz to notes-->
 %nods

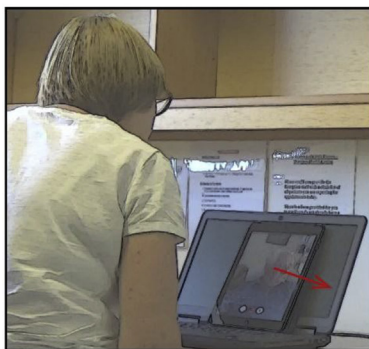


Fig1

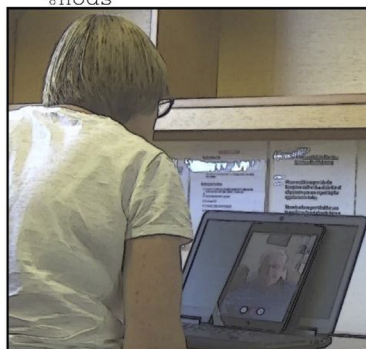


Fig2

At first glance this extract seems unremarkable. The nurse asks a question in line 6. At the point where her turn comes to possible completion, the patient turns his gaze towards the screen, seemingly indicating that he is directing his attention to the nurse in response to the question and that he is about to answer (Rossano, 2013). The nurse maintains her gaze direction and body orientation throughout the subsequent silence, not producing an action and demonstrably awaiting his answer. When the patient does not respond after 1.2 s, the nurse pursues an answer by repeating the question. The patient does eventually answer, but in overlap with the nurse's response pursuit.

This case, however, highlights how latency can interfere with turn-taking. We made the transcript for (1a) using a recording from the nurse's end of the call. If we consider the patient's end, we get a very different picture.

Excerpt 1b – Nurse and patient discussing medication – Home

01 Patient: \$&yep.
 &>>gz to left-->
 ns \$>>gz to notes-->
 02 Nurse: yea:h_i
 03 (0.8)
 04 Nurse: a::nd
 05 (.)
 06 Nurse: y&#ou w- \$an&#d is ↑tha\$t in one go:?
 -->\$turns----->\$gz to scrn-->
 pt &turns-----&gz to scrn-->>
 fig #fig1 #fig2
 07 (.)
 08 Patient: yes.
 09 (0.9)
 10 Nurse: is that in one go\$-yeah_i
 -->\$gz to notes-->>



Fig1



Fig2

The patient actually answers *on time*: there is only a micropause between question and answer (line 6–8). Latency affects the nurse's perception of the patient's behavior. First, she receives his vocal answer significantly later than he produces it: he perceives her turn about 750 ms after she produces it and consequently she perceives his response 1.5s after finishing her turn. Second, she sees him turn his gaze to the screen, directing his attention to her, upon completion of her question (see 1a). When recipients gaze towards the speaker, this “displays reciprocity” (Heath, 1984) and so may indicate an upcoming response (Rossano et al., 2009): the patient's gaze shift at the point of turn completion may thus indicate to the nurse that he is about to answer. However, he had already shifted his gaze at the start of her question (1b, line 6): his gaze reorientation does not indicate incipient speakership, but attention to her question.

The nurse's unnecessary pursuit of a response reveals how participants are often not aware of latency as part of the interactional context. How the nurse makes sense of the patient's vocal and non-vocal actions reveals her assumption that turn production and perception co-occur—that is, she orients to the same turn-taking norms as in face-to-face interaction where co-occurrence of turn production and turn perception is a *seen but unnoticed* background assumption (Garfinkel, 1964).

Whilst recent studies have shown that participants are, at least on some occasions, aware of some of the technical problems that video-mediated interaction poses (Rintel, 2013, 2015; Rusk and Pörn, 2019), in our data they proceed under the same seen but unnoticed background assumptions characteristic of face-to-face interaction. In extracts (1a–b) the nurse does not see, and in fact she cannot see, that she does not receive the patient's turn at the same time at which he produces it. Essentially participants in a video-mediated interaction have different life worlds (Garfinkel, 1967; Luff et al., 2003; Pollner, 1975; Schutz, 1967) and neither is more or less accurate than the other. There are two “non-mutual realities” (Ruhleder and Jordan, 2001)—one for the nurse, and one for the patient—but both orient to their reality as a shared one.

This paper contributes to our understanding of turn-taking in video-mediated interaction in two ways. First, we analyze latency through the lens of the turn-taking system (Sacks et al., 1974): we examine how at points of possible turn transition latency causes (a) silence where talk should occur and (b) overlapping talk. Second, we analyze the embodied behavior of participants in these environments, showing that (a) participants also visually orient to these silences as a failure with turn-taking and (b) latency causes visual cues to be misperceived and misunderstood (e.g., as in (1a–b)).

We argue that how these problems emerge and the resolution practices participants use to address them show that the participants in our data orient to turn-taking in video-mediated interaction as instantaneous: they monitor and interpret both their interlocutor's vocalized and embodied behavior as if production and perception co-occur—that is, as if there were no latency. We thus combine an emic and an etic perspective: demonstrating how participants make sense of these problems (emic), while revealing that this sense-making is discrepant with the actual source of the problems (etic). It is this combined perspective that drives our conclusion that these participants routinely treat the interaction as instantaneous, until latency noticeably interferes with turn-taking. Finally, we discuss the implications of these findings both for participants engaged in video-mediated interaction, as well as scholars interested in researching video-mediated interaction.

2. Materials and methods

We conducted secondary analysis of data that were collected between 2015 and 2018 in two separate studies on video-mediated consultations Shaw et al., 2018. The consultations were recorded in (antenatal) diabetes, cancer, and heart failure services in the UK, and were conducted using either consumer Skype or FaceTime. Patients ranged in age between 21 and 87, with younger patients ($n = 18$, median = 26 years) prevalent in diabetes services and older patients ($n = 19$, median = 72 years) prevalent in cancer and heart failure services. Participants had a range of experience with video-mediated interaction, but due to the variation in latency across conversations, we could not analyze this systematically. Because we were interested in how latency affects the interaction, we only included consultations for which we had recordings from the patient's home and the clinic so we could compare each “reality”. Our final dataset consisted of 25 consultations, totaling 7 h and 57 min of interaction.

We transcribed the data according to conventions by Jefferson (2004a) and Mondada (2018), and compared the two ends of each consultation to collect instances of noticeable, post-turn completion silence and of overlapping talk.

We examined silences at each point where a speaker's turn came to possible completion (the transition-relevance place). However, silence can occur at these points even without latency. We needed to isolate cases in which the speaker attempted

to resolve a post-completion silence, where that silence was caused by latency. In other words, our focus was on cases where latency caused speakers to perceive a noticeable silence, when there was no such silence for the recipient. We therefore excluded cases where the recipient did not take a turn (i.e., where there was an actual failure in turn transition). In these cases, latency would merely prolong the noticeable silence. We did not take the length of a silence as a criterion for inclusion. A silence becomes a noticeable silence when participants treat it as such, that is, when participants orient to that silence as the noticeable absence of talk (e.g., by pursuing a response).

As with silence, overlapping talk can occur in any social interaction (Jefferson, 2004b; Schegloff, 2000). To make sure we only included cases where the emergence of overlap and its resolution were affected by latency, we excluded cases where one participant was making a competitive bid for the turn by launching into a turn when their co-participant's turn was recognizably not reaching completion (Drew, 2009; Jefferson, 1984b; Vatanen, 2018).

Our final collection consisted of 130 cases of noticeable silence and 172 cases of overlap. Since silence resolution strategies almost invariably caused overlapping talk, most cases of noticeable silence are also in the collection of overlap.

We analyzed these data using Conversation Analysis (Ten Have, 2007), focusing on how latency affected the interaction turn by turn (Jefferson, 2004b; Schegloff, 2000). We analyzed the interactional practices by which participants: (a) treated silences as noticeable silences, and (b) attempted to resolve these silences. We subsequently examined all cases of overlapping talk, and then investigated (a) where overlap began for each participant, (b) the strategies they used to try to resolve overlap, and (c) where and how latency caused problems for these strategies.

Both studies from which data were drawn received ethical approval for a detailed analysis of video recordings of video consultations. VOCAL was approved by the National Research Ethics Service Committee London—City Road and Hampstead in December 2014 (14/LO/1883) and OQTS by the South Central—Berkshire Research Ethics Committee in September 2015 (15/SC/0553). All participating staff and patients in both studies gave their informed consent to be audio and video recorded during consultations and for the data to be used for research purposes.

3. Results

The turn-taking system provides for minimal silence between turns and minimal overlapping talk, and it has built-in repair mechanisms to address errors (Jefferson, 2004b; Sacks et al., 1974; Schegloff, 2000). However, the system is built for and applied in a (co-present) context in which recipients perceive turns at the same time as speakers produce them (Ruhleder and Jordan, 2001). When latency is introduced, this causes two problems. First, at a transition-relevance place, even if the recipient takes a turn and produces a next action, latency causes this answer to be delayed and the speaker to perceive silence. Second, when participants talk at the same time, their perception of where in their respective turns the overlap starts (i.e., the point of overlap onset) will differ.

3.1. Silence where talk should occur

In dyadic interaction, turn-transition at a possible completion point can either be *normatively relevant*, when the speaker has selected the recipient to produce a specific “type” of next action (e.g., by asking a question, the speaker selects the recipient to provide an answer), or it can be *possible*, when the speaker has completed their turn without sequentially implicating a specific type of next action (Sacks et al., 1974; Schegloff, 2007). Participants treat silences in these respective environments differently. When turn-transition *should* happen a *gap* emerges (i.e., a failure by the recipient to take a turn in which to produce the projected action). When turn-transition *could* happen, a *lapse* emerges (i.e., the recipient foregoes an opportunity to talk) (Hoey, 2020; Sacks et al., 1974).

Silences thus become problematic and are constituted as a specific type of silence through the behavior of participants. In our data, both the practices used by participants to address silence and the timing with which they implement these practices indicates they do not understand these silences to be the result of latency, at least not initially. They overwhelmingly treat the silences as failures with turn transition,² as silence where either participant should or could have taken a turn. They thus do not orient to these silences as silences that routinely occur in video-mediated interaction as a result of latency.

3.1.1. Understanding and solving gaps

That latency affects turn-taking can best be seen when speakers treat a post-completion silence as the noticeable absence of a projected next action: the recipient is not just heard to be not talking (cf. Hoey, 2020), but to be not responding at a point where a response was due (Sacks et al., 1974; Schegloff, 2007). In all but three cases, we find that they use practices that provide the recipient a new opportunity to respond, that is, they reinstantiate the transition-relevance place. Throughout the post-completion silences, the speaker of the last turn either maintains their gaze if they are looking at the screen, visibly awaiting a response, or they direct their gaze towards the screen after a few hundred milliseconds, before pursuing a response. By using these practices speakers orient to the silence as the noticeable absence of a conditionally relevant response (Schegloff, 2007).

There is a range of practices participants have at their disposal to address these silences such as repeats (extract 1), response prompts (e.g., using a polar response particle such as *yeah* to pursue (dis)confirmation after a yes/no-type question) (Heritage, 1984a) (such as in extract (7) in supplementary material), or adjusting the preference organization of the question, orienting to

² Extract (3a) below is one of three exceptions in our data.

the silence as foreshadowing a dispreferred or non-straightforward response (Davidson, 1984; Kendrick and Torreira, 2015; Robinson, 2020; Sacks, 1987) (e.g., by incrementally adding a tag question to a declarative question, such as in extract (8) in supplementary material).

In the overwhelming majority of our data, speakers correct or clarify (part of) their turn, that is, they perform self-initiated self-repair (Bolden et al., 2012). The speaker identifies a problem (or “trouble”) with their talk and provides the solution for that trouble (Kitzinger, 2013; Schegloff et al., 1977). Take the following extracts from a diabetes consultation in which the doctor asks the patient to confirm that she has been prescribed NovaRapid, a quick-acting insulin that is to be taken with meals.

Excerpt 2a – NovaRapid – Clinic

- 01 Doctor: \$.hh (0.3) that's fine;=how about- so we †also
 \$gz to notes-->
 pt
 &gz to scrn-->
02 Doctor: \$&gave \$you &som:e NovaRapid insulin. didn't we,
 \$turns-\$gz to scrn-->>
 pt
 &gz left--->&gz to scrn-->>
03 #(0.6)
 fig
 #fig1
04 Doctor: %the pen%#: t[type of in%sulins.%]
 %.....%depicts pen-->%,,,,,,%
 fig
 #fig2
05 Patient: [yes: that's that's] ri-right.

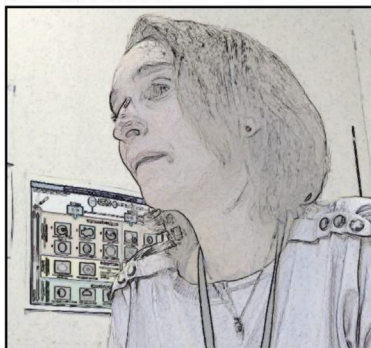


Fig1

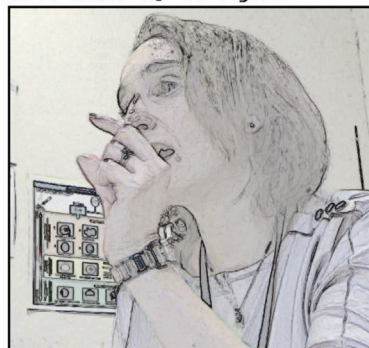


Fig2

Excerpt 2b – NovaRapid – Home

- 01 Doctor: \$.hh (0.3) that's fine;=how about- so& we †also
 \$gz to notes
 pt
 &gz to scrn-----&gz left-->
02 Doctor: \$ga&ve \$you som:e NovaRapid insulin. didn't we,
 \$turns-\$gz to scrn-->>
 pt
 -->&gz to scrn-->
03 Patient: #yes: that's right.=
 fig
 #fig1
04 Doctor: %the &pen%: type of &in%sulins.%
 %.....%depicts pen-->%,,,,,,%
 pt
 ---->&gz down----->&gz to scrn-->>
05 Patient: [°yeah°]

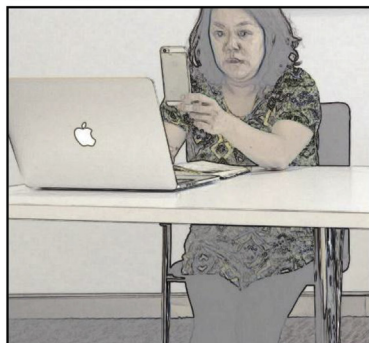


Fig1

Note first that the patient responds on time: there is not even a beat of silence between the doctor's question and her confirming answer (see 2b, line 3). However, the doctor perceives a silence of 0.6 s (2a, line 3). Following syntactic completion of her turn, the doctor adds another constituent that continues that turn, in this case the noun phrase *the pen type of insulin* (i.e., she adds an “increment” to her initial question) (Ford et al., 2002; Schegloff, 2016). With this increment she combines linguistic and embodied resources to depict how the NovaRapid is used. By producing the clarification as an increment, she re-completes her question and provides another opportunity for the patient to provide an answer, following the initial lack of a response. She uses self-initiated self-repair to pursue a response, and by providing a clarification she treats patient's lack of uptake as the result of an understanding problem (Bolden et al., 2012). She thereby reveals her understanding that an answer was due, but missing. In other words, by pursuing a response she treats the silence as a failure with turn-taking, although in actuality it is a result of latency.

In all but three cases, speakers use practices that reconstitute the transition relevance place. In these three exceptions, participants explicitly orient to the technology as the cause of the problem. However, in none of the three do they treat latency as the cause. Consider extracts (3a) and (3b), taken from the start of the consultation in (1a-b).³ The patient asks a question, and when a response is not forthcoming, he provides an account for the silence.

Extract 3a – Cuppa tea - Home

01 Patient: &##.pt have you got a cuppa tea there?
 &>>gz to scrn-->>
 ns %>>gz to scrn-->>
 fig #fig1
 02 (0.8)
 03 Patient: °oh she can't hear me°.=
 04 Nurse: =I wish I ha:d.

Extract 3b – Cuppa tea - Clinic

01 Patient: \$%.pt have you got a cuppa tea there?
 &>>gz to scrn-->>
 ns %>>gz to scrn-->>
 fig #fig2
 02 Nurse: I wish I h[a:d.]
 03 Patient: [°oh she] can't hear me°.=



Fig1

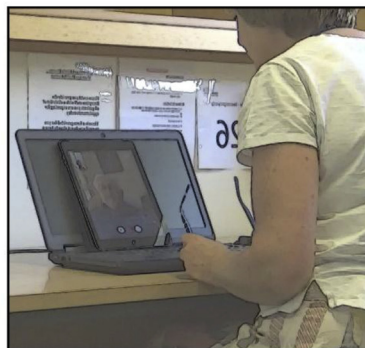


Fig2

The patient produces a question in line 1, making an answer conditionally relevant. The nurse's disconfirmation is significantly delayed and the patient perceives 0.8 s of silence (3a line 2) during which he maintains his gaze towards the nurse, demonstrably waiting for her to take a turn. He then remarks softly—possibly to the GP-researcher—that the nurse cannot hear him. He treats the silence as the noticeable absence of a projected response. He does not, however, pursue a response. Instead he indicates there is a problem with the technology that causes the nurse to not hear him. Note that the account the patient provides for the silence is not latency, but some form of audio trouble. He orients to the silence not as indicating a delay, but as indicating no response at all.

³ The person in front of the patient is a GP-researcher who made the recording. She was not involved in the consultation after setting up.

Recurrently we find that participants in our data treat silence at a point where turn transition is relevant, as a failure in the organization of turn-taking. They use practices to pursue a response, meaning they treat the response not as delayed due to latency, but as missing entirely. By treating these silences as gaps, they show that at least initially they are not aware of latency as the true source of the problem.

3.1.2. Solving lapses

According to the rules of turn-taking, when a speaker reaches possible completion and has not selected a next speaker, turn transition is optional. Recipients will have primary rights to self-select, but they may forego this opportunity. The ensuing silence will still be oriented to as silence where talk should occur (Hoey, 2020), but the turn-taking system provides for its resolution: if none of the recipients self-selects, then the current speaker may continue; and if the current speaker does not continue then “rounds of possible self-selection” (Sacks et al., 1974, p. 715) occur until the silence is resolved (i.e., until either participant has self-selected).

We found that latency interfered with this system. At a point of possible completion where no next speaker has been selected, the current speaker may perceive a lapse even though the recipient has already self-selected. The following example from a diabetes consultation illustrates how speakers perceive and deal with a lapse where one does not occur. The patient has been talking for 1.5 min about a drug trial in which he took part. That story comes to potential completion in line 3, where he returns to describing the study and discusses the outlook (Jefferson, 1978).⁴

Extract 4a – Third part of study - Clinic

01 Patient: &\$and it's the thi:rd part of the study now,
 &gz to left-->
 ns \$gz to scrn-->>
02 Patient &so hopefully .h will continue and see how
 &gz to scrn-->
03 that #goe:s_
 fig #fig1
04 Patient: .hh[h &°and u:h°]
05 Nurse: [%oh ↑right; o]kay:.
 pt --&gz to left-->
 ns %nodes
06 (0.2) &(0.3)
 pt --&gz to scrn-->
07 Patient: ye[ah. and they \$#took- hh]
08 Nurse: [s : o h o w-] ↑how has that has
 pt &gz to left&gz to scrn-->
 fig #fig2
09 an impact on you:?
10 (0.2)
11 Patient: .pt yeah so ↑like u:hm,

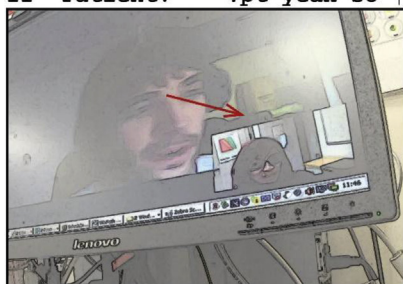


Fig1

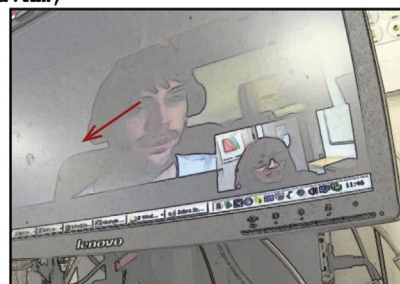


Fig2

⁴ The patient has a dual monitor set-up. Skype is running on the monitor to his left. His camera, however, is positioned so that when he is looking at the Skype screen, the nurse perceives him to be looking to his left (see 4a, Figs. 1 and 2).

Extract 4b – Third part of study - Home

```

01 Patient:      &$and it's the thi:rd part of the study now,
                  &gz to right-->
                  ns      $gz to scrn-->
02 Patient:      &so hopefully .h will continue and see how
                  &gz to scrn-->
03              that #goe:s_
                  fig      #fig1
04 Patient:      .hhh &°and u:[h° yeah. ]
05 Nurse:        [%oh ↑right;] &okay:.
                  %nodes
                  pt      &gz to right-----&gz to scrn-->
06 Patient:      yeah. and they &#[took- hh]&
07 Nurse:        [s:o how-] ↑how has that
                  pt      &gz to right&gz to scrn-->
                  fig      #fig2
08              has an impact on you:?
09 Patient:      &.pt yeah so ↑like u:hm,
                  &gz to right-->>

```

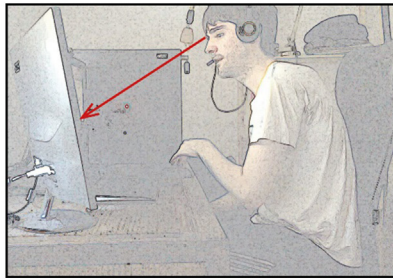


Fig1



Fig2

The nurse receives the patient's story with a claim of understanding *oh right* (Heritage, 1984b) and subsequently uses *okay* to indicate her understanding that his story is done (Beach, 1993). With this turn, she does not select the patient to produce a next turn, and so he can self-select. And indeed, the patient continues. He directs his gaze away from the screen,⁵ and moves into a new TCU (4b, line 6) providing an acknowledgment token (Jefferson, 1984a) before continuing. However, due to latency the nurse perceives 0.5 s of silence, during which she maintains her gaze towards the screen, perceiving the patient to be reorienting his gaze from left to right (i.e., to the screen and thus to her (Luff et al., 2016)), and then asks how the study affected the patient. She thus perceives a lapse where one does not occur and resolves it by self-selecting.

The consequences of this perceived but not actual lapse may seem small initially. There is some overlapping talk before they resume one-speaker-at-a-time with the patient answering the nurse's question (4a, line 11). However, the patient does not return to the turn he had abandoned when the overlap happened. Instead, the talk moves away from the specifics of the study he was talking about. The opportunity to mention what he was going to mention does not arise afterwards and he does not create the opportunity for himself.

We identified 33 lapses that were caused by latency (extract (5) below provides a further example). Lapses provide a different problem for participants than gaps. Latency causes speakers to perceive silence at a point where recipients may self-select. This silence is understood as the recipient foregoing the opportunity to take a turn, even when the recipient has in fact started talking. Because the solution provided by the turn-taking system is for the speaker to continue, the result consistently is overlapping talk, which then has to be resolved by one of the participants dropping out. But that itself is not always straightforward. We now turn to how latency frustrates participants' ability to accomplish overlap resolution.

3.2. Overlapping talk

The turn-taking system provides for minimal overlapping talk (Sacks et al., 1974). When overlap does happen, participants attempt to solve it quickly, applying an “Overlap Resolution Device”: either one of the participants drops out after one or two

⁵ In our data, participants frequently avert their gaze from the screen when launching a turn at talk. The patient here does it three times: in lines 5, 7 and 9 (see 4b). Twice his turn comes off in overlap, and he redirects his gaze to the screen when conceding the turn. The organization of gaze in video-mediated interaction thus seems to be strongly associated with sequence organization, like in face to face (Rossano, 2013). The precise mechanisms and how they are understood by participants, and how these are affected by camera position provide an important area for future research.

Extract 6 – Playing Croquet

Recorded at Patient's End

01 Pt: uhm I told you about playing
 02 croquet? as you know I
 03 played croquet last year;
 04 .h[hh this year .h I'm
 05 Ns: [yeah,
 06 Pt: struggling to complete
 07 one round.
 08 (0.5)
 09 Pt: without having to sit d[own.
 10 Ns: [are
 11 you really;
 12 Pt: yea:h.
 13 (0.5)
 14 Pt: u:hm I'm gonna keep on at
 15 it, I'[ve got a-]
 16 Ns: [and is i]t a pos-
 17 (1.0)
 18 Pt: sorry,
 19 (1.5)
 20 Pt: .hh[h
 21 Ns: [yeah go on;
 22 Pt: I:'ve g[ot-]
 23 Ns: [>y'r gonna] keep
 24 on<,
 25 (0.4)
 26 Pt: yeah I've got a .hh
 27 competition at ((place
 28 name)) tomorrow.



Fig1

Recorded at Nurse's End

Pt: uhm I told you about playing
 croquet? as you know I
 [played croquet last year;
 Ns: [yeah,
 Pt: .hhh this year I'm
 struggling to complete
 one round.
 Ns: are you really;
 Pt: without having to sit down.
 (0.6)
 Pt: [yeah.
 Ns: [(hm)
 (0.4)
 and is i[t a p o s-]
 Pt: [I'm gonna kee]p
 on at it, I've-
 (1.9)
 Pt: sorry,
 Ns: g- yeah go on;
 (0.5)
 Ns: y'r gonna keep on,
 (0.2)
 Pt: I've got-
 (1.0)
 Pt: yeah I've got a .hh
 competition at ((place
 name)) tomorrow.



Fig2

struggles to play croquet.⁸ Throughout the sequence, both participants maintain a posture with focus on their screens (see Figs. 1 and 2, both taken following the overlap at line 16).

The patient's turn reaches possible completion for the first time at line 7. At the patient's end (left column), the nurse does not take a turn and silence ensues. The patient then recompletes his turn by providing an increment (Ford et al., 2002). In overlap with that increment, he hears the nurse asking for confirmation, which he then provides. His confirming *yeah* (line 12 at the patient's end, line 11 at the nurse's end) completes the sequence, and since neither subsequently takes a turn, a lapse ensues. Both participants attempt to resolve this lapse by self-selecting: the nurse begins to offer a potential solution to some of the patient's problems, and the patient starts to say that he will keep trying. But because of latency, neither participant can immediately notice that the other has self-selected. As a result, each has already produced a partial turn when they hear the other's turn, leading to interjacent overlap at both ends (i.e., the patient hears the nurse begin a turn in overlap with his, whereas the nurse hears the patient begin a turn in overlap with hers (lines 14–16)).

⁸ We present the patient's and nurse's version side-by-side to facilitate comparison. To maintain legibility, we did not add Mondada (2018) conventions here. Detailed transcripts are provided in supplementary material. Note that we did not transcribe according to chronological time, and so the line numbers do not show what happens at a specific point in time, simultaneously, for both participants.

In this example, latency has a significant effect on participants' perception of where overlap onset happens. It also frustrates their attempts to figure out who gets to talk next. First when the participants attempt to give the right to talk to the other party by abandoning their turn, subsequently when they explicitly grant the other the right to talk through what Jefferson (2004b) calls a *repeat request*, and finally when the patient attempts to self-select.

Both participants initially attempt to resolve the overlap by abandoning their turn after a few overlapping beats, but due to latency these strategies create a new problem. From the patient's perspective (left column), he drops out quickly after the nurse starts her turn, and so she continues talking (line 16) without him talking in overlap. However, from the nurse's perspective (right column), she stops talking after about two beats of overlapping talk, and it is in fact the patient who can continue talking (line 16) without her talking in overlap. Whilst they solve the overlapping talk by both dropping out, they are faced with another problematic silence (line 17).

To move out of this silence, one participant needs to self-select or grant the turn to the other, but again latency frustrates their moves for a solution. The patient first attempts to give the turn to the nurse: by saying *sorry* (line 18) he orients to his prior overlapping talk as an interruption and acknowledges that the nurse has the right to talk. The nurse responds with *go on* (line 19), thereby asking the patient to continue his turn. Here the problem could have been resolved. After he has heard the nurse's turn, the patient recognizably restarts his turn by repeating *I've got* (line 15 and 22). However, because latency delays the nurse's perception of his restart, she perceives 0.5 s of silence and then more actively prompts him to continue by repeating part of his turn: *y'r gonna keep on*. This prompt comes off in overlap the patient's restart. He abandons his turn and another silence ensues.

This silence, while still problematic, is more easily resolved. The nurse has now twice explicitly prompted the patient to continue his turn, and the patient has made a first attempt at doing so. Both have thus oriented to the patient as having the right to talk. The patient acknowledges the nurse's second prompt with *yeah* (Jefferson, 2004b) before restarting his turn for the third and final time. Although the nurse still perceives a silence of 1 s between his second and third restart, she does not make a move to do anything in this silence: she waits for the patient to continue.

The above example is perhaps extreme: there were only 15 cases in our dataset with so many overlaps and repeated attempts to fix them. We have included it here, because it reveals in detail how latency causes problems for the basic mechanisms of organizing social interaction. In chronological time it does not take the participants long to fix the problem of whose turn it is. Between the first point of overlap onset and the patient resuming his turn, only 12.7 s have elapsed. However, in interactional time it takes a lot of work. There are multiple extensive silences and each participant uses multiple practices to resolve the halt in progressivity, running into new problems each time. This reveals how fundamentally latency frustrates smooth turn-taking and participants' ability to solve problems as they arise.

4. Discussion

We demonstrated in this paper in detail how latency causes problems for turn-taking. Building on previous work on latency (Olbertz-Siitonen, 2015; Ruhleder and Jordan, 2001; Rusk and Pörn, 2019; Schoenenberg et al., 2014; Tang and Isaacs, 1993), we explored how latency in video-mediated interaction matters for the talk by focusing on how participants manage silences in transition relevance places and overlapping talk. First, by causing a delay between turn-production and turn-perception, latency causes speakers to perceive silence where talk should, and in fact does happen. This problem can be exacerbated by re-orientation of recipient gaze, which is also delayed and therefore can be understood differently (e.g., gaze can signal incipient speakership, whereas it was done to signal attention to a new action). Second, when participants talk in overlap, latency causes them to have different perceptions of where overlap starts and it delays the overlap resolution strategies they implement. Often these problems are quickly resolved, but sometimes they lead to extended sequences before participants manage to resume one-speaker-at-a-time.

These findings show that the participants in our data were generally not aware of latency: they used overlap and silence resolution practices that treated these problems as turn-taking problems, not as latency problems. Latency is thus well and truly part of the context of video-mediated interaction (Arminen et al., 2016): it has *procedural consequences* (Schegloff, 1991). An upshot is that video-mediated interaction can provide a natural breaching experiment (Garfinkel, 1967) for turn-taking. Participants are recurrently faced with a basic structural-organizational problem, providing us as analysts with an invaluable resource to understand how participants address these problems.

Note, however, that while latency is a clear problem in our data, we conducted secondary analysis on a small data set, meaning that the data were collected for a different purpose, and that they involved many older people who had limited experience with video-mediated interaction. A study designed to investigate latency specifically could provide a more detailed understanding of (a) how long a delay needs to be before it starts to disrupt turn-taking, and (b) to what extent the length of the delay matters for different types of actions and turn-transitions. Furthermore, the COVID-19 pandemic has led to an enormous increase in the use of video-mediated communication. Early indications are that people are now far more aware of the problems that latency can cause. Indeed, it seems to be one of the main frustrations. Whether and if so how this emerging experience re-shapes the turn-taking rules for video-mediated interaction is unclear. Participants may become better at addressing the problems, and develop new turn-taking conventions that are optimized for the medium.

4.1. Upshot for study of video-mediated interaction

Video-mediated interaction has been a subject of research for decades, and much of our social interaction has moved online as a result of the COVID-19 pandemic and social distancing protocols. Nonetheless, the literature around turn-taking in video is scarce and primarily about institutional talk. Since latency is omni-present, inter-turn silences will always be longer in video-mediated interaction than they are for in-person or telephone interaction. However, what kind of silences are tolerable and when video-mediated interaction can be understood as “smooth” has so far not been investigated. Our data suggests that with latency around 100 ms, turn-taking is barely affected, whereas with 700 ms, participants struggle routinely. Studies need to be set up to systematically address the nature of video-mediated turn-taking, focusing on its vocal, embodied and technological components in both dyadic and multi-party interaction. Given the challenge posed by social distancing protocols for data collection, such studies may prove difficult during the pandemic, but with screen capture software it should be possible to start exploring these questions in a more systematic way.

Our analysis indicates that in order to get an adequate understanding of the sequential organization of video-mediated interaction, it is often necessary to have recordings of both ends of the conversation. However, this is not always practically feasible. Participants may be far apart geographically (e.g., Ekberg et al., 2019; Licoppe, 2017; Licoppe and Morel, 2012; Rintel, 2013, 2015), and it places a significant demand on the research team to travel to multiple locations to make recordings. In addition there are methodological considerations: participants do not have access to each other's reality (Luff et al., 2003) and in order to get an *emic* perspective, we should focus on the individual's life world (Olbertz-Siitonen, 2015).

For many questions, however, it is crucial that researchers have access to each side of the conversation, both what each participant sees, hears and does.⁹ Participants are continuously asking the question “why that now” (Schegloff and Sacks, 1973), and while there is one answer for each participants, for us as analysts there are as many answers as there are participants. An analysis of how a sequence unfolds and how structure emerges cannot be conducted when we rely on one side of the interaction.

4.2. Conclusion

Immediacy is a taken for granted background assumption of communication. We have demonstrated in this paper that in video-mediated interaction, time and again latency disrupts the structural organization of the interaction. These problems also happen in face-to-face and other forms of synchronous, instantaneous interaction. Indeed, it is precisely because these problems are familiar, that they are not easily recognized as technological problems: there is no way for participants to distinguish between silence that indicates a withheld response, and silence that indicates a delayed response. All silences sound the same. That participants manage to overcome these problems at all is then a testament to the robustness of the turn-taking system. No matter how significant the impact of latency is on the interaction, participants manage to resume normal turn-taking at some point, even if it takes some effort and frustration.

Funding

This work was supported by the National Institute for Health Research (NIHR) [grant number PB-PG-1216-20012]; the NIHR Biomedical Research Center, Oxford, UK [grant number BRC-1215-20008]; and the Wellcome Trust [grant number WT104830MA]. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Author contributions

Lucas Seuren: Writing – Original Draft, Data Curation, Investigation **Joseph Wherton:** Investigation, Writing – Review & Editing **Trisha Greenhalgh:** Funding acquisition, Writing – Review & Editing **Sara Shaw:** Supervision, Writing – Review & Editing, Project Administration.

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.pragma.2020.11.005>.

⁹ What happens off-screen may be consequential for what happens on-screen. Recordings should ideally capture as much of the environment as possible to facilitate in-depth analyses.

References

- 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>.
- Beach, W.A., 1993. Transitional regularities for 'casual' "Okay" usages. *J. Pragmat.* 19 (4), 325–352.
- Bolden, G.B., Mandelbaum, J., Wilkinson, S., 2012. Pursuing a response by repairing an indexical reference. *Res. Lang. Soc. Interact.* 45 (2), 137–155. <https://doi.org/10.1080/08351813.2012.673380>.
- Davidson, J., 1984. Subsequent versions of invitations, offers, requests, and proposals dealing with potential or actual rejection. In: Atkinson, J.M., Heritage, J. (Eds.), *Structures of Social Action*. Cambridge University Press, Cambridge, pp. 102–128.
- Dingemanse, M., Floyd, S., 2014. Conversation across cultures. In: Enfield, N.J., Kockelman, P., Sidnell, J. (Eds.), *The Cambridge Handbook of Linguistic Anthropology*. Cambridge University Press, Cambridge, pp. 447–480.
- Drew, P., 2009. Quit talking while I'm interrupting": a comparison between positions of overlap onset in conversation. In: Haakana, M., Laakso, M., Lindström, J. (Eds.), *Talk in Interaction: Comparative Dimensions*. Finnish Literature Society, Helsinki, pp. 70–93.
- Ekberg, S., Danby, S., Theobald, M., Fisher, B., Wyeth, P., 2019. Using physical objects with young children in 'face-to-face' and telehealth speech and language therapy. *Disabil. Rehabil.* 41 (14), 1664–1675. <https://doi.org/10.1080/09638288.2018.1448464>.
- Ford, C.E., Fox, B.A., Thompson, S.A., 2002. Constituency and the grammar of turn increments. In: Ford, C.E., Fox, B.A., Thompson, S.A. (Eds.), *The Language of Turn and Sequence*. Oxford University Press, Oxford, UK, pp. 14–38.
- Garfinkel, H., 1964. Studies of the routine grounds of everyday activities. *Soc. Probl.* 11 (3), 225–250.
- Garfinkel, H., 1967. *Studies in Ethnomethodology*. Prentice-Hall, Englewood Cliffs.
- Goodwin, C., 1979. The interactive construction of a sentence in natural conversation. In: Psathas, G. (Ed.), *Everyday Language: Studies in Ethnomethodology*. Irvington Press, New York, NY, pp. 97–121.
- Heath, C., 1984. Participation in the medical consultation: the co-ordination of verbal and nonverbal behavior between the doctor and patient. *Sociol. Health Illness* 6 (3), 311–338. <https://doi.org/10.1111/1467-9566.ep10491964>.
- Heath, C., Luff, P.K., 1993. Disembodied conduct: interactional asymmetries in video-mediated communication. In: Button, G. (Ed.), *Technology in Working Order: Studies of Work, Interaction, and Technology*. Rank Xerox Research Centre, London, UK, pp. 35–54.
- Heritage, J., 1984a. Garfinkel and Ethnomethodology. Polity Press, Cambridge.
- Heritage, John, 1984b. A change-of-state token and aspects of its sequential placement. In: Atkinson, J. Maxwell, Heritage, John (Eds.), *Structures of Social Action*. Cambridge University Press, Cambridge, UK, pp. 299–345.
- Hindmarsh, J., Fraser, M., Heath, C., Benford, S., Greenhalgh, C., 1998. Fragmented interaction: establishing mutual orientation in virtual environments. In: *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*. ACM, New York, NY, pp. 217–226.
- Hoey, E.M., 2020. When Conversation Lapses: the Public Accountability of Silent Copresence. Oxford University Press, Oxford, UK.
- Jefferson, G., 1973. A case of precision timing in ordinary conversation: overlapped tag-positioned address terms in closing sequences. *Semiotica* 9 (1), 47–96. <https://doi.org/10.1515/semi.1973.9.1.47>.
- Jefferson, G., 1978. Sequential aspects of storytelling in conversation. In: Schenkein, J. (Ed.), *Studies in the Organization of Conversational Interaction*. Academic Press, New York, pp. 219–248.
- Jefferson, G., 1984a. Notes on a systematic deployment of the acknowledgement tokens "Yeah"; and "Mm Hm". *Pap. Linguist.* 17 (2), 197–216. <https://doi.org/10.1080/08351818409389201>.
- Jefferson, G., 1984b. Notes on some orderlinesses of overlap onset. In: D'Urso, V., Leonardi, P. (Eds.), *Discourse Analysis and Natural Rhetorics*. Cleup Editore, Padua, IT, pp. 11–38.
- Jefferson, G., 1986. Notes on 'latency' in overlap onset. *Hum. Stud.* 9 (2/3), 153–183.
- Jefferson, G., 2004a. Glossary of transcript symbols with an introduction. In: Lerner, G.H. (Ed.), *Conversation Analysis: Studies from the First Generation*. John Benjamins, Amsterdam, NL, pp. 13–31.
- Jefferson, G., 2004b. A sketch of some orderly aspects of overlap in natural conversation. In: Lerner, G. (Ed.), *Conversation Analysis: Studies from the First Generation*. John Benjamins, Amsterdam, NL, pp. 43–59.
- Kendrick, K.H., Torreira, F., 2015. The timing and construction of preference: a quantitative study. *Discourse Process* 52 (4), 255–289. <https://doi.org/10.1080/0163853X.2014.955997>.
- Kitzinger, C., 2013. Repair. In: Sidnell, J., Stivers, T. (Eds.), *The Handbook of Conversation Analysis*. Wiley-Blackwell, Chichester, UK, pp. 229–256.
- Lerner, G.H., 2003. Selecting next speaker: the context-sensitive operation of a context-free organization. *Lang. Soc.* 32 (2), 177–201. <https://doi.org/10.1017/S00474045032202X>.
- Lerner, G.H., 2019. When someone other than the addressed recipient speaks next: three kinds of intervening action after the selection of next speaker. *Res. Lang. Soc. Interact.* 52 (4), 388–405. <https://doi.org/10.1080/08351813.2019.1657280>.
- Levinson, S.C., 2016. Turn-taking in human communication – origins and implications for language processing. *Trends Cognit. Sci.* 20 (1), 6–14. <https://doi.org/10.1016/j.tics.2015.10.010>.
- Licoppe, C., 2017. Skype appearances, multiple greetings and 'coucou': the sequential organization of video-mediated conversation openings. *Pragmatics* 27 (3), 351–386. <https://doi.org/10.1075/prag.27.3.03lic>.
- Licoppe, C., Morel, J., 2012. Video-in-Interaction: "talking heads" and the multimodal organization of mobile and Skype video calls. *Res. Lang. Soc. Interact.* 45 (4), 399–429. <https://doi.org/10.1080/08351813.2012.724996>.
- Luff, P.K., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., Oyama, S., 2003. Fractured ecologies: creating environments for collaboration. *Hum. Comput. Interact.* 18 (1–2), 51–84. https://doi.org/10.1207/S15327051HCI1812_3.
- Luff, P.K., Heath, C., Yamashita, N., Kuzuoka, H., Jirotko, M., 2016. Embedded reference: translocating gestures in video-mediated interaction. *Res. Lang. Soc. Interact.* 49 (4), 342–361. <https://doi.org/10.1080/08351813.2016.1199088>.
- Moerman, M., Sacks, H., 1988. On 'understanding' in the analysis of natural conversation. In: Moerman, M. (Ed.), *Talking Culture: Ethnography and Conversation Analysis*. University of Pennsylvania Press, Philadelphia, pp. 180–186.
- Mondada, L., 2018. Multiple temporalities of language and body in interaction: challenges for transcribing multimodality. *Res. Lang. Soc. Interact.* 51 (1), 85–106. <https://doi.org/10.1080/08351813.2018.1413878>.
- Olbertz-Siitonen, M., 2015. Transmission delay in technology-mediated interaction at work. *PsychNol. J.* 13 (2–3), 203–234.
- Pollner, M., 1975. The very coinage of your brain': the anatomy of reality disjunctures. *Philos. Soc. Sci.* 5 (3), 411–430. <https://doi.org/10.1177/004839317500500304>.
- Pomerantz, A., 1984. Pursuing a response. In: Atkinson, J.M., Heritage, J. (Eds.), *Structures of Social Action: Studies in Conversation Analysis*. Cambridge University Press, Cambridge, UK, pp. 152–163.
- Rintel, S., 2013. Video calling in long-distance relationships: the opportunistic use of audio/video distortions as a relational resource. *Electron. J. Commun./La Revue Électronique de Communication* 23 (1&2).
- Rintel, S., 2015. Omnirelevance in technologised interaction: couples coping with video calling distortions. In: Fitzgerald, R., Housley, W. (Eds.), *Advances in Membership Categorisation Analysis*. SAGE, London, UK, pp. 123–150.
- Robinson, J.D., 2020. Revisiting preference organization in context: a qualitative and quantitative examination of responses to information seeking. *Res. Lang. Soc. Interact.* 53 (2), 197–222. <https://doi.org/10.1080/08351813.2020.1739398>.
- Rossano, F., 2013. Gaze in conversation. In: Sidnell, J., Stivers, T. (Eds.), *The Handbook of Conversation Analysis*. Wiley-Blackwell, Chichester, UK, pp. 308–329.

- Rossano, F., Brown, P., Levinson, S.C., 2009. Gaze, questioning, and culture. In: Sidnell, J. (Ed.), *Conversation Analysis: Comparative Perspectives*. Cambridge University Press, Cambridge, UK, pp. 187–249.
- Ruhleder, K., Jordan, B., 2001. Co-constructing non-mutual realities: delay-generated trouble in distributed interaction. *Comput. Support. Coop. Work* 10 (1), 113–138. <https://doi.org/10.1023/a:1011243905593>.
- Rusk, F., Pörn, M., 2019. Delay in L2 interaction in video-mediated environments in the context of virtual tandem language learning. *Ling. Educ.* 50, 56–70. <https://doi.org/10.1016/j.linged.2019.02.003>.
- Sacks, H., 1987. On the preference for agreement and contiguity in sequences in conversation. In: Button, G., Lee, J.R.E. (Eds.), *Talk and Social Interaction. Multilingual Matters*, Clevedon, UK, pp. 54–69.
- Sacks, H., Schegloff, E.A., Jefferson, G., 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50 (4), 696–735.
- Schegloff, E.A., 1991. Reflections on talk and social structure. In: Boden, D., Zimmerman, D.H. (Eds.), *Talk and Social Structure: Studies in Ethnomethodology and Conversation Analysis*. University of California Press, Berkeley, CA, pp. 44–70.
- Schegloff, E.A., 1992. Repair after next turn: the last structurally provided defense of intersubjectivity in conversation. *Am. J. Sociol.* 97, 1295–1345. <https://doi.org/10.1086/229903>.
- Schegloff, E.A., 2000. Overlapping talk and the organization of turn-taking for conversation. *Lang. Soc.* 29 (1), 1–63.
- Schegloff, E.A., 2001. Accounts of conduct in interaction: interruption, overlap, and turn-taking. In: Turner, J.H. (Ed.), *Handbook of Sociological Theory*. Springer, Boston, MA, pp. 287–321.
- Schegloff, E.A., 2002. Opening sequencing. In: Katz, J.E., Aakhus, M. (Eds.), *Perpetual Contact: Mobile Communication, Private Talk, Public Performance*. Cambridge University Press, Cambridge, UK, pp. 326–385.
- Schegloff, E.A., 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis I*. Cambridge University Press, Cambridge, UK.
- Schegloff, E.A., 2016. Increments. In: Robinson, J.D. (Ed.), *Accountability in Social Interaction*. Oxford University Press, Oxford, UK, pp. 238–263.
- Schegloff, E.A., Jefferson, G., Sacks, H., 1977. The preference for self-correction in the organization of repair in conversation. *Language* 53 (2), 361–382. <https://doi.org/10.1353/lan.1977.0041>.
- Schegloff, E.A., Sacks, H., 1973. Opening up closings. *Semiotica* 8, 289–327.
- Schoenenberg, K., Raake, A., Koeppe, J., 2014. Why are you so slow? – misattribution of transmission delay to attributes of the conversation partner at the far-end. *Int. J. Hum. Comput. Stud.* 72 (5), 477–487. <https://doi.org/10.1016/j.ijhcs.2014.02.004>.
- Schutz, A., 1967. In: Walsh, G., Lehnert, F. (Eds.), *The Phenomenology of the Social World*. Northwestern University Press, Evanston, IL.
- Shaw, Sara, Cameron, Deborah, Wherton, Joseph, Seuren, Lucas, Vijayaraghavan, Shanti, Bhattacharya, Satyajit, A'Court, Christine, Morris, Joanne, Greenhalgh, Trisha, 2018. Technology-Enhanced Consultations in Diabetes, Cancer, and Heart Failure: Protocol for the Qualitative Analysis of Remote Consultations (QuARC) Project. *JMIR Research Protocols* 7 (7). <https://doi.org/10.2196/10913>.
- Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al., 2009. Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. USA* 106 (26), 10587–10592. <https://doi.org/10.1073/pnas.0903616106>.
- Stivers, T., Rossano, F., 2010. Mobilizing response. *Res. Lang. Soc. Interact.* 43 (1), 3–31. <https://doi.org/10.1080/08351810903471258>.
- Tang, J.C., Isaacs, E.A., 1993. Why do users like video? *Comput. Support. Coop. Work* 1 (3), 163–196.
- Ten Have, P., 2007. *Doing Conversation Analysis: A Practical Guide*. SAGE, London.
- Vatanen, A., 2018. Responding in early overlap: recognitional onsets in assertion sequences. *Res. Lang. Soc. Interact.* 51 (2), 107–126. <https://doi.org/10.1080/08351813.2018.1413894>.

Lucas M. Seuren is a Health Services Researcher at the University of Oxford. His research interests include the impact of communication technologies on the delivery and quality of healthcare services. The main focus is on how professionals and patients adapt their vocal and embodied interaction to the video-mediated context, what this means for the feasibility of using remote consultations as a service model, and how technology can be designed to better support their needs.

Joseph Wherton is a Senior Researcher at the University of Oxford and a Research Member at Wolfson College. His research interests include the use of technologies to support older people and vulnerable groups, including people living with dementia, loneliness, restricted mobility and chronic illness. In particular, his research seeks to support the co-production of personalised care solution that can address the complex and diverse needs of service users.

Trish Greenhalgh is Professor of Primary Care Health Sciences at the University of Oxford and a Senior Fellow at Green Templeton College. She leads a programme of research at the interface between social sciences and medicine, with strong emphasis on the organisation and delivery of health services. Her research seeks to celebrate and retain the traditional and humanistic aspects of medicine while also embracing the unparalleled opportunities of contemporary science and technology to improve health outcomes and relieve suffering.

Sara E. Shaw is Associate Professor of Health & Social Policy at the University of Oxford and a Fellow at Green Templeton College. Her particular interests lie in the development of health and social policies, how these are understood and interpreted by patients, practitioners, professionals and policymakers, and how policies shape organisational processes, routines and decision-making. She has considerable experience of developing and applying qualitative methods and has undertaken work on topics ranging from digital health and integrated care to health and research policy.