

Turn taking in student Zoom discussions

Paul Stone and Asa Brinham

1. Introduction

With the sudden shift to online teaching due to the Covid-19 pandemic, teachers and students have had to quickly adapt to new ways of working. This has challenged established interaction practices, developed through years of experience in face-to-face interactions in classrooms. Online, synchronous video-mediated interactions, often conducted via software such as Zoom, have replaced (or are conducted alongside) traditional classroom environments, and this new environment for delivering education has significance for the types of interaction that teachers and learners engage in.

Understandably, there is still little published research that investigates classroom interactions conducted via Zoom. This paper seeks to address this by investigating video-mediated interaction in English communication courses in a Japanese university. In this paper, we focus on how Japanese students participate in English discussions in breakout rooms on Zoom, using Conversation Analysis (CA) to investigate how they organize turn-taking. We hope that the paper will provide potentially useful insights for teachers.

In the following literature review, we will first introduce some key CA concepts that are relevant to the analysis in this paper. We will then introduce CA research into video-mediated interactions and student discussions.

2. Conversation Analysis

2.1 Introduction to CA

CA, which originated as a sub-discipline of sociology in the late 1960s (Hoey &

Kendrick, 2017; Markee, 2000), focuses not on language per se, but is rather the study of how social interactions are organized by participants. CA investigates both ordinary conversations and institutional interactions such as those found in educational settings, and as such the term ‘talk-in-interaction’ is often preferred to the word ‘conversation’ (Markee, 2000).

2.2 Turns and turn constructional units (TCUs)

A key component of the analytical framework of CA is the practice of turn-taking. Although overlapping speech occurs frequently during *talk-in-interactions*, for interactions to succeed, they must be sequential. What then constitutes a turn? In their seminal paper, Sacks et al. (1974) use the term *turn-constructional unit* (TCU) to describe the parts that make up a turn. A TCU is a linguistic unit that can be recognized as a “complete utterance in a given context” (Hoey & Kendrick, 2017). In other words, it could be several sentences or a single sound that holds some communicative significance.

2.3 Transition Relevance Place (TRP)

When a speaker completes a TRP, a *transition relevance place* (TRP) often occurs. TCU is a term first used by Sacks et al. (1974), and it is the point at which an interlocutor’s turn ends, or appears to other interlocutors to be potentially over. It is a window of opportunity for someone else to take their turn in the interaction. It usually involves some signaling beyond a pause or a cessation of speech, such as a change in intonation and/or nonverbal indicators such as eye movement or hand gestures.

2.4 Turn allocation

CA research has revealed talk-in-interaction to be highly organized, as participants in interactions employ certain practices to distribute turns (Hayashi, 2013, p. 167). Turn-allocation refers to the methods used to select the next speaker. Turn-

allocation techniques fall into two categories: (a) the current speaker selects the next speaker, and (b) the next speaker self-selects (Sacks et al., 1974). There is a set of ‘rules’ for *talk-in-interactions* that bestow rights and obligations on the participants. If the current speaker selects a next speaker, then the next speaker has the right and obligation to take their turn, but if a next speaker is not selected then anyone can take their turn (self-selection). If neither of these occur, the current speaker may continue until the next TRP is reached.

2.5 No-gap-no-overlap

Participants in interactions exhibit a normative orientation to only one speaker at a time (Hayashi, 2013), and the turn-taking practices introduced above are used to achieve this. In a study of ten languages, Stivers et al. (2009) found that all of the languages provided evidence for a general avoidance of overlapping talk and silence between turns. This is sometimes referred to as the ‘no gap no overlap’ principle. Even very small silences between turns are noticeable by participants in interaction. In a study of lapses in conversation, Hoey (2018) used 0.5 seconds as the lower limit for identifying a lapse, as this is the point at which the turn-taking option ‘same speaker continues’ (Sacks et al., 1974) tends to cluster.

2.6 Embodied behaviour

While early CA studies tended to focus on spoken language, with the development of video-recording technology, studies began to focus more on the multimodal aspects of interaction. Heath (1984), for example, found that participants display reciprocity through the use of gaze and posture directed towards a co-participant. More recently, Auer (2021) has argued that gaze is the most ubiquitous next speaker selection technique in conversation, as the current speaker often gazes towards a co-participant towards the end of a turn.

2.7 CA research into video-mediated interaction

Mlynar et al. (2018) provide an overview of CA research into video-mediated interaction. Studies have looked at how participants open meetings (e.g. Licoppe, 2015; Mondada, 2015), finding that they engage in a number of practices prior to beginning the main activity, such as deciding whether to use video (Ibnelkaid, 2015) and adjusting the physical environment (Pappas and Seale, 2009). Participants also orient to the quality of the connection, and gaze at the screen to monitor the status of the technology (Mondada, 2015).

Other studies have looked at how participants perform nonverbal actions and present themselves on screen. In video-mediated interactions, Licoppe and Morel (2012) found that participants maintain a ‘talking heads’ arrangement (in which participants are on screen and facing the camera) when there is nothing else relevant to the interaction to show. Video-mediated interaction undermines participants’ ability to perform gestures and establish eye contact (Heath and Luff, 1993), as it is difficult to determine the exact direction of other participants’ gestures and gaze (Luff et al., 2016). Luff et al. (2016) further note that video provides only limited access to others’ physical environments. Luff et al. (2003) refer to this as *fractured ecologies*, as the environment in which an action is produced and in which it is received is fractured, meaning that one participant may be unable to make sense another’s behaviour.

Research into video-mediated medical consultations (Shaw et al. 2020) has found that latency (digital delays between one person saying something and another hearing it) of more than 0.5 seconds can disrupt interactions and lead to overlapping talk. This overlapping talk may be resolved quickly, for example by one participant stopping talking, or it may lead to participants competing for the right to talk. While video-conferencing technology may place certain constraints on interactions, it also provides some affordances for interactional behaviour that is less common in face-to-face talk. In Zoom meetings with cameras turned on, all participants’ facial expressions are available to monitor, and Van Braak et al. (2021)

show how non-speaking participants creatively use this affordance to increase participation.

2.8 CA research into student discussions

More ‘traditional’ classrooms have been found to be dominated by the teacher, who for the most part controls turn-taking (Gardner, 2013, p. 594). However, pedagogical changes have led to student-centred approaches that include more task-oriented small-group work. In these classrooms, there is a range of speech-exchange systems, including more conversation-like interaction in which students locally organize turn-taking by themselves (Markee and Kasper, 2004; Seedhouse, 2004). For example, Hauser (2009) describes how university students in a Japanese EFL class engage in extended interactional work to negotiate the next speaker, and how this negotiation includes pointing gestures and gazes.

3. The current study

3.1 Setting

The current study uses CA to investigate turn-taking in student discussions in breakout rooms on Zoom. Breakout rooms are virtual rooms in Zoom, which can be used to split a larger group of people into smaller groups. The data for this study were collected during the second semester of the academic year in two classes at a Japanese university. Both classes were undergraduate English communication courses, which made use of Zoom as the primary method of facilitating spoken interaction. The particular focus of both classes was on developing the students’ English discussion skills, and the students had varying levels of English proficiency, ranging from approximately 500-700 on the TOEIC Listening and Reading Test.

3.2 Data and analytical method

Students were recorded participating in English discussions in breakout rooms

on a weekly basis. The data comprised 15 recordings totaling about 135 minutes, which were made on the researchers' computers using Zoom's recording function. Discussions typically lasted between 8-10 minutes, with students placed into groups randomly. Teachers' instructions depended on the nature of the task, but students were always asked to speak only in English. Discussions centred on handouts that featured discussion prompts (ranging from 1-15 prompts, depending on the task). Zoom allows participants to have their cameras on or off, and while the teachers encouraged students to turn their cameras on, this was not compulsory.

The data were transcribed and analyzed following CA methods. First, the audio-recorded data were transcribed in detail. Multimodal features were added following Mondada's (2019) conventions for multimodal transcription. The transcripts below have been simplified for presentation.

Following the exploratory nature of CA (Sidnell, 2013, p. 77), we approached the data without any preconceived ideas of what we were looking for. Instead, we looked through the data for phenomena of potential interest. Once a particular behaviour was identified, we then looked across the data to find other instances of that behaviour. In this way, we identified recurrent patterns.

4. Analysis

In this section, we explicate our findings by presenting representative examples from the data. We will focus in turn on (1) the recurrence of silence between different speakers' turns, (2) overlapping talk, (3) problems with identifying speakers, and (4) students' resources for dealing with turn-taking problems.

4.1 Silences between turns

Regardless of the size of the group, the topic being discussed, and the proficiency of the students, silence at TRPs was common across the data. Approximately 75% of all TRPs occurred with a silence of more than 0.5 seconds.

4.1.1 Silence in a pair interaction with higher-proficiency speakers

Excerpt 1 provides examples of inter-speaker silences. The excerpt is taken from an interaction in which two higher-proficiency students are discussing a list of questions provided by the teacher in a PDF. In Excerpt 1, the participants are negotiating the transition to the next question on the list. Throughout this paper, symbols in the transcript (e.g. | or \triangle) indicate where embodied actions begin in relation to spoken language.

Excerpt 1

01 M: the next question?

M: gazes@screen

S: gazes@screen

02 (0.5)

M: gazes at and leans over paper

03 S: yeah.

S: nods

04 (0.7)

05 M: do you think you eat |like a horse?

M: |gazes@camera, sits upright

06 (0.7)

07 S: yea:h (.) yeah

S: nods

In line 1, Mai (pseudonyms are used throughout) suggests moving on to the next question, gazing at her screen as she does so (we presume that she is gazing at her screen, as she is gazing near, but not directly at, her camera). This gives the impression that she is gazing towards her interactional partner, Saburo. As Mai speaks, Saburo also gazes at his screen, and both participants are visibly oriented to the interaction.

As Mai's turn is delivered with rising intonation and a gaze towards Saburo (as he appears on her screen), it makes relevant an answer from him. However, there is a 0.5 second silence before Saburo provides this answer. Mai's turn (line 1) requires a yes-or-no response, with a preference for an affirmative answer. That

Mai expects an affirmative response can be seen as she gazes away from her screen to the handout (line 2), which is printed out and visible in her hand, in readiness to read the question, without waiting for Saburo's response. Saburo's affirmative response (line 3) is a simple, preferred action, and he does not display any problems in delivering it (e.g. through false starts or delay markers). Nonetheless, Saburo's response is delayed.

Having received Saburo's affirmative response, Mai reads the question from the handout (line 5) after a silence of 0.7 seconds (line 4). As she reads the question, Saburo continues to gaze towards his screen in a continued state of reciprocity. As Mai approaches the end of the question, she gazes away from her paper and towards her camera, so that she appears to be gazing directly at Saburo. She also shifts posture, sitting back in her chair, which suggests a change in focus away from the handout. This shows, prior to reaching the projected TRP, that an answer from Saburo will be relevant. In other words, she is not simply reading the question, but is addressing it to Saburo. As in line 1, Mai's turn makes relevant a yes-no answer, and Saburo provides an affirmative response to the question in line 7. As in line 3, he apparently has no problem in providing this answer. Nonetheless, his answer starts after a 0.7 seconds of silence.

This short excerpt explicates how silences occurred between speakers' turns. In Excerpt 1, there were only two participants in the discussion, and the excerpt featured yes-no questions that were responded to without any apparent difficulty. As Mai asked the questions, she gazed at her screen or camera, indicating that the questions were directed towards Saburo. There should therefore have been little doubt as to who the questions were directed at. Despite this, there were silences between the questions and their answers.

4.1.2 Silence in a group interaction with lower-proficiency speakers

Excerpt 1 showed how silence occurred between turns in an interaction between two higher-proficiency students who were often gazing towards their screens/

cameras. Silences occurred just as frequently in interactions with lower-proficiency students and larger group sizes, but in these interactions the silences were often longer. Excerpt 2 (presented below) shows three lower-proficiency students engaged in the same interactional practice as Excerpt 1 - negotiating the transition to the next discussion question on a handout.

The excerpt begins in a similar manner to Excerpt 1, with Naoko proposing that they move on to the next question (line 1). However, unlike Excerpt 1, not everyone is gazing at their screen. As she starts speaking, both Naoko and Yuri are gazing off the bottom-right corner of their screens, while Takahiko is apparently gazing towards his computer screen. However, the focus of Takahiko's gaze is somewhat lower than his camera, and we do not have the impression that he is looking at us.

As Naoko finishes her turn in line 1, Yuri gazes up at her screen, and after a short silence she self-selects as the next speaker, responding affirmatively to Naoko's suggestion. Her gaze at the screen, occurring at a TRP, is thus used in her self-selection as next speaker. Yuri then returns her gaze to its former position off the camera. There may have been some expectation that Naoko, having proposed moving on to the next question, would also ask the question by reading it out, as Mai did in Excerpt 1. However, there follows 2.7 seconds of silence. The participants' screen displays at this moment are shown in the images in line 4. Naoko's face is not visible as she gazes off the screen, Yuri is gazing down and to the right, while Takahiko is still gazing towards his screen with his eyes moving right-left, suggesting that he is reading.

In line 5, after 2.6 seconds of silence, Takahiko demonstrates that he is indeed reading, as he starts reading the next question out loud. As the general direction of his gaze has not moved, this suggests that he has been gazing at the PDF since line 1. Therefore, until this moment, there have been no shared gazes towards the screen. The embodied actions of the participants indicate a focus on things in their immediate surroundings and the handout, rather than on one another.

Before Takahiko reaches the end of the question, Yuri also starts to read the question (line 6), starting in overlap with Takahiko, demonstrating a shared focus on the question. Yuri finishes her turn with downward intonation and without gazing at the screen or shifting her posture in a state of reciprocity, and she does not verbally nominate a next speaker. As such, this reading of the question does not make clear who should speak next, and 26 seconds passes before someone starts to answer the question.

First, after a 0.8 second silence, Takahiko receipts Yuri's turn and then leans back in his chair (lines 9-10). While this acknowledges the previous turn, it does not claim a full turn for Takahiko, and the interaction lapses into silence again. Both Yuri and Naoko gaze at their screens (lines 11 and 14), before laughing and immediately gazing away again. By immediately gazing away from their screens they make themselves unavailable for talk, and the gaze away also suggests that the laughter is not an invitation to shared laughter. Takahiko's "eh" in line 16, also performed with a gaze away from his screen, displays that he is thinking about something, but does not develop talk.

Excerpt 2

01 **N: shall we move to the next |question?**
 N: gazes off screen
 Y: gazes off screen |gazes@screen
 T: gazes@screen
 02 (0.6)
 03 **Y: yeah.**
 04 (2.7)
 Y: gazes off screen
 T: gazes@screen - eyes move right-left



Naoko displays that she is thinking about the question, but again does not initiate talk, while Yuri's "hm" in line 20 does similar.

Finally, in line 22, Naoko briefly gazes at her screen before self-selecting to answer the question. This happens 26 seconds after Yuri had finished reading the question in line 7. The longer silences in this excerpt are partly due to the lack of next-speaker nomination and sequence-initiating actions (e.g. a question directed at someone), and a hesitancy to self-select. It is possible that this question may have been a little difficult for these students to answer, which could explain the amount of time it took for someone to answer. However, they had been working on the topic of tattoos for about 50 minutes by this point, and they had just finished discussing a very similar question ("What kind of tattoos do you like?"), during which Naoko had actually given the answer to this question by saying that she would like a tattoo of her family name. She acknowledges this at the beginning of her turn (line 22) by saying "as I said".

In the data collected for this study, longer silences at TRPs often occurred when participants were looking away from their screens, as happens in Excerpt 2. When the question was read (lines 6-7), only Takahiko was gazing at the screen (and he was apparently reading the question rather than gazing at his co-participants), and as gaze has been shown to be important in selecting a next speaker (Auer, 2021) and displaying reciprocity to another speaker (Heath, 1984) this may have impacted the negotiation of who will speak next.

4.2 Overlapping talk

While not as common as silences between turns, there was also frequently overlapping talk in the data, particularly in interactions that included higher-proficiency students. This suggests that timing turns precisely was not always easy for these participants. Excerpt 3 shows two participants negotiating the start of their discussion of the first question on a handout.

Excerpt 3

01 S: okay so let's talk about the question.
S: gazing bottom-right
N: gazing bottom-right
02 (0.8)
03 N: [yeah.]
04 S: [first of] all we have to answer the (.) question one.

Saburo suggests that they start the discussion (line 1), and Naoko responds affirmatively to this (line 3). Naoko starts this response after 0.8 seconds of silence (a similar length of inter-turn silence to those we saw in Excerpt 1 and lines 1-3 of Excerpt 2), and the response occurs in overlap with Saburo continuing to speak (line 4). According to Hoey (2018, p. 331), when a current speaker continues after a short silence at a TRP, the continuation tends to start after around 0.5 seconds of silence. In line 4, after 0.8 seconds of silence, Saburo self-selects to continue speaking, and this occurs in overlap with Naoko's "yeah". Overlap of this sort, in which a recipient provides a response (e.g. *yeah*, *nn*, *I see*, etc.) as another speaker continues, is frequent in the data, and usually occurs after a silence of more than 0.5 seconds. It appears that the timing of short responses is difficult to perform accurately on Zoom, leading to overlapping talk, which can hinder the smooth progression of the interaction.

Excerpt 4 provides another example of overlapping talk. Unlike Excerpt 3, in which the overlapping talk occurred after a silence, this overlapping talk occurs towards the end of an ongoing turn. It also occurs as one participant shifts her posture and gaze away from the screen, and there is an apparent lack of understanding between the participants. Yuri and Naoko have been asked to decide what the best five cuisines in the world are, and just prior to Excerpt 4, Naoko asked how they should choose the criteria to make this decision.

Yuri suggests that they brainstorm some ideas, and then laughs (line 1). Her rising intonation and gaze towards the screen make relevant a response from Naoko. As Yuri speaks, Naoko displays reciprocity by gazing at the screen. Naoko

does not, however, immediately reply and there follows 1.9 seconds of silence (line 2) during which Naoko leans forward and gazes down (as shown in the image in line 2). At this moment, Naoko is not observably focused on interacting with Yuri, and does not respond to the suggestion.

Recipients of yes-no questions often perform disagreements or rejections with delays (Pomerantz and Heritage, 2013, p. 214; Stivers et al., 2009), and Yuri might see Naoko's delay as being indicative of an upcoming negative response, as she justifies her suggestion (or perhaps appeals to Naoko to provide an alternative idea) by saying that she does not have any other ideas (line 3).

Excerpt 4

01 Y: so: (.) like |(1.0) brainstorming? uhuh hah hah hah

Y: gazing@screen

N: gazing@screen

02 (1.9)

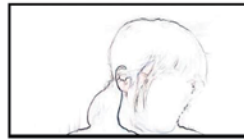
N: gazes down and leans forward



03 Y: I don't know (0.4) other (1.2) |uhuh solu[tion.

N

|sits back, gazes right



04 N:

[I think (0.9)

05 N: I think it's important to how look like.

However, before Yuri finishes this turn, Naoko sits up and gazes right (shown in the image in line 4). She then starts speaking in overlap with Yuri, following Yuri's suggestion that they brainstorm ideas by offering "how to look like" (i.e. appearance) as one idea. It had apparently not been necessary for Yuri to justify her

suggestion in line 3, as Naoko had already accepted it.

While Yuri treated Naoko's delay in responding as a potential problem, Naoko herself did not observably orient to it as such, and she did not attempt to account for it (e.g. by saying something like "sorry, just a moment"). Nor was the delay a hesitation before a negative response. The two participants, therefore, treat this delay differently. The lack of physical co-presence may have contributed to this difference. We are not able to see what Naoko is gazing at in lines 2-3, and it is unclear what she is doing. Participants in an interaction may not necessarily treat silence as a problem if some observable action can account for it (such as a dictionary check [Stone, 2019, p. 5-6]). The actions that Naoko was performing out of view of the camera might have accounted for her silence, and had Yuri been able to see what Naoko was doing, then possibly the silence may have been less of a problem.

Silences and overlaps often occurred in the data when participants shifted their posture and gaze away from the screen, and this excerpt provides an example of this happening. Not being physically co-present, and having a limited view of a co-interactant's space, can cause difficulties in understanding what the other person is doing. Further, not gazing at the screen to monitor an interactional partner may make it harder to achieve shared understandings of the progress of the interaction.

4.3 Identifying recipients and speakers

Another problem that occurred when taking turns, particularly in groups of three or more, was deciding who should speak next. Excerpt 5 shows a problem in determining who an initiating action is intended for. In this interaction, three students were discussing their attitudes to food. These students had their cameras turned off and so only the audio is transcribed.

Excerpt 5

01 Y: how about (.) you? (1.5) ahm (1.0) did you (0.6) eat unfamiliar
02 Y: foods (0.7) have you (.) eat (1.4) unfamiliar foods?
03 (3.0)
04 C: ah me?
06 (0.5)
07 Y: yeah.
08 C: ah. a:::hm (.) yeah I (0.7) actually when I went to Taiwan

Just prior to this excerpt, Chihiro had asked Yuri if she had eaten unfamiliar food, and Yuri answered this question. Yuri then proceeds to ask the same question (lines 1-2). Yuri's "how about you" can be understood as directed at Chihiro, as she had been addressing her turn to Chihiro until now. However, there follows a 3-second silence, indicating some trouble. Chihiro initiates repair of this trouble (line 4), asking if the question is intended for her. After a 0.5 second silence, Yuri confirms that the question is addressed to Chihiro, who then provides her answer.

The intended recipient of Yuri's question (lines 1-2) could be understood from the previous talk. Nonetheless, Yuri's use of "you" creates some ambiguity, which leads to the problem in determining the next speaker. As the participants' cameras were turned off, there were also no visual clues as to who should speak next. It is, however, worth noting that similar problems occurred when participants had their cameras on, as it was not always clear who a speaker was gazing at (and therefore who they intended as the recipient of their turn).

In Excerpt 6, the problem is not determining who should speak next, but rather who has just spoken. This excerpt is taken from the same interaction as Excerpt 5. The handout used in this discussion had been used in the previous week's class, and so the participants had already discussed some of the questions. Excerpt 6 shows the start of the interaction, as the speakers negotiate which question from the handout to discuss.

The participants establish that Yuri was absent during the last class (lines 1-3), and so she has not discussed any of the questions yet. In line 6, Chihiro begins a

turn, which she abandons as Yuri speaks in overlap with her (line 7). However, there is computer distortion on Yuri's voice, and only the "yeah" in line 7 is clearly audible. This creates trouble, indicated by the subsequent silence and Chihiro's repair initiation (line 9). Chihiro's "eh" with rising intonation indicates some uncertainty, before she says "Yuri" also with rising intonation. She is displaying an understanding that it is possibly Yuri who has just spoken, and is seeking confirmation of this. Once she has this confirmation (line 11), she continues her abandoned turn in line 13.

Excerpt 6

```

01 Y: sorry I was (0.8) abse[nt]
02 C:                                [oh] you were absen[t
03 Y:                                [last class (.]
04 Y: so (0.6) ahuhu huhu
05   (0.3)
06 C: okay so may[be a:h           ]
07 Y:          [((inaudible))] ((inaudible)) yeah
08   (1.0)
09 C: eh? Y?
10   (0.5)
11 Y: yes.
12   (0.4)
13 C: oh okay. (0.7) a:h (0.5) ja Y (1.0) you can ask any questions

```

Excerpts 5 and 6 provide examples of problems in identifying speakers and recipients. Gaze plays an important role in selecting next speakers, and Zoom interactions make the use of this important resource difficult. It may also at times be difficult to determine who is speaking, especially when more than one person speaks at the same time, or when there is distortion on the audio. In a face-to-face interaction, different speakers' voices will reach us from the different places in which those speakers are. In a computer-mediated interaction, however, all voices reach us from the same amplification point, and perhaps this also makes distinguishing between speakers more difficult.

4.4 Students' actions for managing turn-taking


Excerpts 1-6 provide examples of silence, overlapping talk, and difficulties in determining who has spoken and who will speak next. While these issues were common across the data, the participants were nonetheless able to participate in the discussions and complete their tasks. Often, as in Excerpts 1 and 3, the participants did not observably orient to silence or overlaps as problematic, while in Excerpts 5 and 6 we saw how participants were able to attend to problems actively. The students also displayed competency in dealing with turn-taking issues in a number of other ways. It is worth noting that the interactional practices for managing turn-taking seen in the next three excerpts were almost exclusively performed by the more proficient learners in the study (with TOEIC scores of around 700).

4.4.1 Negotiating the next speaker after overlapping talk

In Excerpts 3 and 4, we saw examples of overlapping talk. Higher-proficiency speakers displayed a practice for negotiating the next speaker after overlapping talk, as explicated in Excerpt 7.

In Excerpt 7, a group of students is discussing tattoos. Saburo asks the other members of the group what they think the most popular tattoo in the world is (lines 1-2). After a 0.8 second silence, Mai starts to answer the question (line 4), but as she does so Saburo also continues speaking. Both Mai and Saburo abandon their turns, and after a short silence (line 6) Mai apologizes, demonstrating that she sees the overlapping talk as a problem. After apologizing, she lifts her right-hand palm-upwards towards the camera while laughing, in a gesture that offers the next turn to Saburo. In line 8, Saburo responds to Mai's apology by saying it is "okay", and offers her the next turn, while performing a similar hand gesture to Mai's. In this way, using both verbal and embodied resources, the problem is resolved, and Mai starts answering the question (line 9).

Excerpt 7

01 S: what do you guys think the most (.) I mean best (0.7) most
02 S: popular (.) tattoo in this world.
03 (0.8)
04 M: actua[lly] I
05 S: [I mean]
06 (0.6)
07 M: okay sorry | (.) uh huhuh
M: |palm-up gesture towards camera

08 S: it's okay ^ (.) go ahead.=
S: ^gestures with hand towards camera
09 M: =I don't see a lot of tattoo in my life

4.4.2 Verbal nomination of next speaker at turn-beginning

Participants sometimes nominated a next speaker by name. This not only resolved issues with deciding who should speak next, but could also help to eliminate silence between turns. Excerpt 8 is taken from the same interaction as Excerpts 5 and 6, and it shows a speaker nominating a next-speaker by name in turn-initial position.

Excerpt 8

01 C: Y have (.) have you ever:: tried anything (.) unfamiliar
02 C: when you (0.6) went (0.5) travelling [or
03 Y: [ah a::h.
04 (1.5)
05 Y: I (0.9) a::h when I went to (0.6) Aizu in Fukushima

Prior to asking her question (lines 1-2), Chihiro specifies that the question is addressed to Yuri at the beginning of line 1. By selecting the next speaker in turn-initial position, she allows Yuri to pay attention and claim the next turn at the earliest opportunity. Although there is a silence in line 4, there is no silence

between speakers, as Yuri uses the non-lexical token *ah* (line 3) to project her upcoming turn and claim speakership while Chihiro's turn is nearing completion. Using *ah* avoids the negative effects of overlap (Hayashi, 2013, p. 174) while displaying that Yuri has understood the question, so that Chihiro can stop speaking and allow Yuri to take her turn.

4.4.3 Turn-final laughter

Inter-speaker silence was sometimes avoided with the use of turn-final laughter, which filled a transition space that might otherwise have been silent. Excerpt 9 provides an example of this.

Excerpt 9

01 S: when we went to the French restaurant? (.)
 S: gazing@screen
 M: gazing@screen
 02 M: a[ha]
 03 S: [he] ate snail but you know (0.6) [(inaudible)]
 04 M: [he::?]
 05 S: ^that's not you know (.) food for humans (.)
 M: ^covers mouth with left hand--->
 06 S: |uh huh heh huh
 S: |covers mouth with left hand--->
 07 M: yeah I can't eat like |disgusting \$things\$.
 S: --->|

Saburo is telling the story of his brother eating snail at a French restaurant (lines 1-5). Both participants gaze at their screens, suggesting engagement in the interaction, and Mai receipts Saburo's story with the non-lexical tokens "aha" and "hee" (a Japanese token that can be used to express surprise), following which she places her hand in front of her mouth in a gesture of surprise that she maintains throughout the rest of the excerpt. This further demonstrates her engagement with what Saburo is saying. After Saburo reaches a TRP (line 5), and so possibly brings his turn to completion, he laughs. Mai then self-selects as next speaker (line 7) just

as Saburo finishes laughing, without any inter-turn silence occurring.

Shaw et al. (2013, p. 102) found that laughter that occurs after a turn is complete may fill a transition space where a delay might otherwise be expected to ensue. This laughter can soften interactional trouble that might be suggested by a delay. By laughing at the end of their turns, participants were able to cover short inter-speaker silences. It may also help to promote affiliation between speakers. As Saburo laughs (line 6), he places his hand over her mouth in a very similar way to Mai (line 5), so that both participants are performing almost identical gestures simultaneously.

5. Discussion and conclusions

In this study, we have investigated how Japanese university students in two English classes organized turn-taking in pair and small-group discussions in breakout rooms on Zoom. The above analysis shows how organizing interactions on Zoom provides some challenges that are not normally present in face-to-face interactions.

5.1 Timing

Precisely timing turns on Zoom seems to be challenging, evidenced by the frequent silences between turns and overlapping talk in our data. While the participants' English proficiency may play some role in this, the time lag that occurs in online video-mediated interactions most likely also plays a role. Research has found, that while periods of latency up to 0.2 seconds are unlikely to have much effect on the progress of a video-mediated interaction, gaps of 0.5 seconds or more can result in overlap (Shaw et al. 2020). In the current study, we observed frequent gaps of 0.5 seconds or more between speakers' turns, and although these did not always lead to overlapping talk (and when overlapping talk did occur it was not always treated as a problem, as in Excerpts 3 and 9), there were examples of problematic overlapping talk occurring after a silence (see Excerpt 7).

Sacks et al. (1974, p. 705) argue that “self-selection, to be assured, must be

done at the transition relevance place”. In other words, in order to claim a turn as next speaker, participants should aim to start speaking as soon as the previous speaker finishes. Slight delays at a TRP, which are very common on Zoom, may therefore cause problems for selecting the next speaker, as the current speaker may see the silence as an opportunity to continue while other participants simultaneously attempt to claim the next turn.

5.2 Turn allocation

There were also issues in negotiating who should speak next in the discussions. Gaze is argued to be essential for next-speaker selection (Auer, 2021), and in face-to-face EFL classroom interactions Japanese university students have been found to use pointing gestures and gazes towards one another when negotiating a next speaker (Hauser, 2009). However, it is difficult to use these embodied resources in the same way on Zoom, as we cannot be sure exactly who co-participants are looking or pointing at. It therefore seems likely that, in groups of more than two, there is a greater need for verbal nominations or self-selection on Zoom than in face-to-face talk.

5.3 Fractured ecologies

In Zoom interactions, participants are only able to see a selected area of the other participants' space. It is therefore often difficult to determine precisely what others are gazing at and what they are doing. In Excerpt 4, for example, we cannot be sure what Naoko is doing when she leans forward and gazes off screen, and this causes problems in understanding the significance of her actions. In a shared space, we may be able to determine that someone is gazing at a dictionary (for example), and therefore most likely searching for a word. This information, which can help us to decide how to interact with others, is often not available to us on Zoom.

5.4 Students managed interactions by themselves

While the participants in this study faced challenges in organizing their discussions, for the most part they were able to successfully complete tasks and informally reported that they enjoyed these discussions. That the students enjoyed these discussions is also observable in the data. We can see in Excerpt 9 how the participants are clearly engaged with one another and enjoying the discussion. Without input from the teacher, the students performed interactional work to manage their discussions and were often successful in maintaining one speaker at a time. In doing so, they managed and resolved silences, overlapping talk, and issues with determining previous and next speakers.

While embodied resources, such as gaze, were not available to the same extent as in face-to-face talk, participants could nonetheless make use of them to help organize their interactions. For example, students could help to facilitate orderly turn-taking by visibly displaying reciprocity (e.g. directing body posture and gaze purposefully towards the screen at the end of a turn). Some students were more successful than others in managing their interactions, and perhaps predictably it was the more proficient students who displayed most competency in this. For example, more proficient learners made use of hand gestures, apologies and phrases such as “go ahead” to resolve instances of overlapping talk, which demonstrated their competence in managing interactions in English, as these are precisely some of the practices that L1 speakers have been found to use (Shaw et al. 2020). More proficient learners also made use of verbal nomination of the next speaker, with every instance of this practice in the data coming from one of the more proficient students.

5.5 Limitations of study

Latency issues mean that we cannot be sure exactly when each participant in an interaction receives information, and the data recorded on an analyst’s computer are just one perspective on what happened (van Braak et al., 2021). This is

something that we need to remember when performing a precise analysis of timings in an online video-mediated interaction. It is probable that latency issues contribute to many of the frequent silences between turns in our data, but while participants did likely experience similar silences to those that we observed, without access to recordings made on their computers we cannot be certain of this. This analytical problem emphasizes the difficulties in timing that participants in a Zoom interaction are constantly dealing with.

The current paper is not able to address a number of potentially interesting and compelling lines of investigation. For example, how do participants new to Zoom develop interactional practices over time, how do students' relationships with one another affect their Zoom interactions, and how are interactions in the L1 and L2 similar and different? These all seem to be fruitful topics for future studies.

We also need to bear in mind that this is just one study of two classes, and as such we should be wary of generalizing the findings to other classrooms. Having said this, the findings of the current study are largely consistent with the findings of other CA studies into video-mediated interactions, which does suggest there may be some case for generalizing.

5.6 Ideas for teaching

With the above caveats in mind, we end this paper by presenting some ideas for teaching based on the observations made in this paper. The findings of this study would suggest that, compared with face-to-face discussions, students may need different types of support to help them organize their online interactions.

To help resolve issues with turn allocation, it may be useful to encourage students to verbally nominate the next speaker. In particular, this may be done in the turn-initial position to help the next speaker take their turn at the earliest opportunity. Guidance from the teacher may also help to avoid the problems in deciding the next speaker seen in Excerpt 2. Students may be advised to only have one student read each question from the handout, or choose a group leader to read

the questions and nominate speakers. They may further be instructed in how to read the questions in a way that makes them sequence initiating actions. Students could be advised to shift posture and gaze towards the screen as they finish reading, and verbally nominate the next speaker.

Practices for resolving overlapping talk used by the more proficient students in the study were not used by the lower-proficiency students. Teachers of lower-level classes may therefore want to teach some of these practices, such as saying “sorry, go ahead” while performing a hand gesture. Also, the problems posed by being in different physical spaces may also be helped by encouraging students to verbally explain what they are doing. For example, phrases like “I’m just checking my notes” or “bear with me a moment” may help students to reach better understandings of one another’s embodied actions.

We saw in Excerpt 9 how turn-final laughter can cover silences between different speakers’ turns. In Excerpt 9, the participants were both observably engaged in the talk with one another, and as well as laughter there were expressions of surprise, shared gestures that indicated affiliation, and mutual gazes towards the screen. This suggests, perhaps unsurprisingly, that interactions with higher levels of affiliation between participants may have a better interactional flow, and that choosing engaging and enjoyable topics may be useful for promoting smooth interaction.

A further point to consider in relation to Excerpt 9, is that it involves only two higher-proficiency students. Van Braak et al. (2021) found that turn-taking in Zoom interactions became more difficult with larger numbers of participants, while Stivers (2021) argues that the conversational turn-taking system may be best-suited to dyadic interaction. As such, when trying to facilitate smooth interactions on Zoom, it may be worth keeping lower-level students in pairs when putting them in breakout rooms. It might also be useful to include higher-level students in larger groups where possible to help facilitate talk.

The above are just some initial ideas formed in response to the current study.

Clearly, more research is needed into how students organize their interactions on Zoom, and as our understandings of this develop, we will hopefully achieve better understandings of how to best prepare students for these interactions.

References

- Auer, P. (2021). Turn-allocation and gaze: A multimodal revision of the “current-speaker-selects-next” rule of the turn-taking system of conversation analysis. *Discourse Studies*, 23(2), DOI: 10.1177/1461445620966922
- Gardner, R. (2013). Conversation Analysis in the classroom. In J. Sidnell & Stivers, T. (eds.), *The handbook of Conversation Analysis*, 593-611. Oxford: Blackwell.
- Hauser, E. (2009). Turn-taking and primary speakership during a student discussion. In H. Nguyen and Kasper, G. (eds.), *Talk-in-interaction: Multilingual perspectives*, 216-244. Hawaii: University of Hawaii.
- Hayashi, M. (2013). Turn allocation and turn sharing. In J. Sidnell & Stivers, T. (eds.), *The handbook of Conversation Analysis*, 167-190. Oxford: Blackwell.
- Hoey, E. M. (2018). How speakers continue with talk after a lapse in conversation. *Research on Language and Social Interaction*, 51(3), 329-346.
- Hoey, E. M. and Kendrick, K. H. (2017). Conversation analysis. In A. M. B. De Groot & Hagoort, P. (eds.), *Research methods in psycholinguistics and the neurobiology of language: A practical guide*, 151-173. Chichester: Wiley & Sons.
- Heath, C. (1984). Talk and reciprocity: sequential organization in speech and body movement. In Atkinson, J.M. and Heritage, J. (eds.), *Structures of social action: Studies in conversation analysis*, 247-265. Cambridge: Cambridge University Press.
- Heath, C. and Luff, P. (1993) Disembodied conduct: interactional asymmetries in video-mediated communication. In Button, G. (ed.), *Technology in Working*

- Order: Studies of Work, Interaction and Technology*, 35–54. Routledge, London/New York.
- Ibnelkaid, S. (2015) Scénographie d’une ouverture d’interaction vidéo. In *Réseaux*, 6, 125–168.
- Licoppe, C. (2015) ‘Apparitions’, multiples salutations et ‘coucou’. In *Réseaux*, 6, 85–124.
- Licoppe, C., & Morel, J. (2012). Video-in-Interaction: “Talking Heads” and the Multimodal Organization of Mobile and Skype Video Calls. *Research on Language and Social Interaction*, 45(4), 399-429.
- Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K. and Oyama, S. (2003) Fractured ecologies: creating environments for collaboration. *Human Computer Interaction*, 18(1), 51–84.
- Luff, P., Heath, C., Yamashita, N., Kuzuoka, H. and Jirotko, M. (2016) Embedded reference: translocating gestures in video-mediated interaction. *Research on Language and Social Interaction*, 49(4), 342–361.
- Markee, N. (2000). Conversation analysis. Mahwah, NJ: Lawrence Erlbaum.
- Markee, N., & Kasper, G. (2004). Classroom talks: An introduction. *The Modern Language Journal*, 88, 491–500.
- Mlynar, J., Gonzalez-Martinez, E. & Lalanne, D. (2018) Situated Organization of Video-Mediated Interaction: A Review of Ethnomethodological and Conversation Analytic Studies. *Interacting with Computers*, 30(2), 73–84.
- Mondada, L. (2015) Ouverture et préouverture des réunions visiophoniques. *Réseaux*, 6, 39–84.
- Mondada, L. (2019) Conventions for multimodal transcription. Accessed 30 June 2021. <https://www.lorenzamondada.net/multimodal-transcription>
- Pappas, Y. and Seale, C. (2009) The opening phase of telemedicine consultations: an analysis of interaction. *Social Science and Medicine*, 68(7), 1229–1237.
- Pomerantz, A. and Heritage, J. (2013). Preference. In J. Sidnell & Stivers, T. (eds.), *The handbook of Conversation Analysis*, 210-228. Oxford: Blackwell.

- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, 50(4), 696–735.
- Seedhouse, P. (2004). *The interactional architecture of the language classroom: A conversation analysis perspective*. Oxford: Blackwell.
- Shaw, C., Hepburn, A., & Potter, J. (2013). Having the last laugh: On post-completion laughter particles. In P. Glenn & Holt, E. (eds.), *Studies of laughter in interaction*, 91-106. London: Bloomsbury.
- Shaw, S. E., Seuren, L. M., Wherton, J., Cameron D., A’Court, C., Vijayaraghavan, S., Morris, J., Bhattacharya, S., & Greenhalgh, T. (2020) Video Consultations Between Patients and Clinicians in Diabetes, Cancer, and Heart Failure Services: Linguistic Ethnographic Study of Video-Mediated Interaction. *Journal of Medical Internet Research*, 22(5). DOI: 10.2196/18378
- Sidnell, J. (2013). Basic Conversation Analytic methods. In J. Sidnell and Sivers, T. (eds.), *The handbook of Conversation Analysis*, 77-99. Oxford: Blackwell.
- Stivers, T. (2021). Is conversation built for two? The partitioning of social interaction. *Research on Language and Social Interaction*, 54(1), 1-19.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), 10587-10592.
- Stone, P. (2019). Repair sequences in ‘off-task’ conversations in an EFL university classroom in Japan: Japanese language resources and learning opportunities. *Classroom Discourse*, 10(2), 188-208.
- van Braak, M., Huiskes, M., Schaepkens, S., & Veen, M. (2021). Shall we all unmute? A conversation analysis of participation in online reflection sessions for General Practitioners in Training. *Languages*, 6(72). DOI: 10.3390/languages6020072

Zoom での学生ディスカッションにおける話者交替

Paul Stone
Asa Brinham

Covid-19の影響により授業が突然オンラインに移行することとなり、教員と学生の両方がその新しい学習方法に迅速に適応しなければならない状況になった。本稿ではそのような状況の中で大学の英語コミュニケーションクラスにおいて、日本人大学生がオンライン上でビデオ討論をする際に、そこでどのようなことが起きていたのかを調査した結果について詳述する。学生がZoomの「breakout room」で英語による会話にどのように参加しているかを詳細に把握するため、討論に参加している様子を録画し、その後、会話分析の手法を用いてその録画データの分析を行った。

分析の結果、学生がオンライン上で同時に会話に参加する際に次のような問題が生じていることが分かった。具体的には(1)話者交替のタイミングをどのように正確に知るか(2)次に誰が話すべきかをどのようにして理解するか(3)他の参加者の身体動作の意味をどのようにして理解するかの3つであった。学生たちの中にはこれらの問題に対処するため、次に示すような対応を行った学生がいたことが分かった。(1)話者交替を容易にするためウェブカメラを注視する(2)身振り表現を使って発話の重複を回避する(3)話者交替の際に次の話者を口頭で指名することの3つである。この研究によってオンラインでの英語によるビデオ討論を行う際に生ずる問題点が明らかになった。また、そのような問題が生じているにもかかわらず、学生たちは討論に首尾よく参加できていたことも明らかになった。

