

ULIANA M. KIMSTACH

Independent Researcher, Moscow, Russia

NILS KLOWAIT

RANEPa, Moscow, Russia

ORCID: 0000-0002-7347-099X

MARIA A. EROFEEVA

RANEPa, Moscow, Russia

ORCID: 0000-0002-0874-5272

Talking without a Voice: Virtual Co-Speakership in an Educational Webinar

doi: 10.22394/2074-0492-2021-4-198-216

198

Abstract:

The following paper analyzes the interactional shifts precipitated by the pandemic induced turn to telepresence. Using the framework of multimodal conversation analysis, we analyze a videorecording of a webinar organized by

Kimstach Uliana Maximovna — BA in Sociology, independent researcher. Research interests: microsociology, multimodal conversation analysis. Email: catoneck@gmail.com

Klowait Nils — Research Fellow at the Center for Innovative Social Research of the Russian Presidential Academy of National Economy and Public Administration (RANEPa), Senior Research Fellow at the International Center for Contemporary Sociological Theory of the Moscow School of Social and Economic Sciences (MSSSES). Research interests: microsociology; multimodal conversation analysis; mixed and virtual reality technologies; education research. E-mail: nils.klowait@gmail.com

Erofeeva Maria Aleksandrovna — Candidate of Sociological Sciences, Researcher at the Center for Sociological Research, RANEPa, Senior Research Fellow at the International Center for Contemporary Sociological Theory (MSSSES). Research interests: microsociology; multimodal conversation analysis; sociology of technology; actor-network theory; education research. E-mail: erofeeva-ma@ranepa.ru

Статья подготовлена в рамках научно-исследовательской работы Центра социологических исследований РАНХиГС «Долгосрочные эффекты технологической стратификации образовательных практик» (2021).

Acknowledgments: This text was prepared in the context of the research project “The long-term effects of the technological stratification of educational practices” (2021) conducted by the Center for Sociological Research, RANEPa.

the Psychological Service of Moscow. In this specific case, webinar participants had unequally distributed interactional resources; only one participant was able to speak, while all other participants could only participate through a text-based chat. We focus on a change of the course of action where the instructor's monologic presentation transitions to a question-answer interaction. We highlight the way the single speaker organizes the transition from these structurally dissimilar participation frameworks. A key feature of the move from monologue to question-response is a self-initiated interruption: another participant's diachronic chat message is deployed as a synchronic overlap by orienting to a virtual second speaker. Thus, we document a case where one speaker chooses to give a voice to a voiceless participant. The work contributes to studies of educational interaction by providing insights on the work that goes into the transition between interactional formats in telemediated asymmetrical ecologies. Our work opens up discussions about the interfacing between different modalities as a locally emergent phenomenon, and how new interactional ecologies create a fertile substrate for hitherto unfamiliar forms of talking, embodiment, and local sequential ordering. The work thus also contributes to research that highlights the non-passive role of the 'listener', which is reflected in the active speaker's orientation to the listener's active contribution to ongoing talk.

Keywords: video-mediated communication, classroom interaction studies, telepresence, multimodality, conversation analysis, co-operative action, distributed speakership, overlap.

199

Ульяна М. Кимстач¹, Нильс Кловайт², Мария А. Ерофеева³

Безмолвный разговор: конструирование виртуального собеседника в образовательном вебинаре

Резюме:

Пандемия интенсифицировала процессы «виртуализации» социальных отношений и взаимодействий. С точки зрения микросоциологии

- 1 Кимстач Ульяна Максимовна — бакалавр социологии, независимый исследователь. Научные интересы: микросоциология; мультимодальный конверсационный анализ. Email: catoneck@gmail.com
- 2 Кловайт Нильс — научный сотрудник Центра перспективных социальных исследований РАНХиГС, старший научный сотрудник Международного центра современной социологической теории МВШСЭН. Научные интересы: микросоциология; мультимодальный конверсационный анализ; технологии смешанной и виртуальной реальности; исследования образования. E-mail: nils.klowait@gmail.com
- 3 Ерофеева Мария Александровна — кандидат социологических наук, научный сотрудник Центра социологических исследований РАНХиГС, старший научный сотрудник Международного центра современной социологической теории МВШСЭН. Научные интересы: микросоциология; мультимодальный конверсационный анализ; социология технологий; акторно-сетевая теория; исследования образования. E-mail: erofeeva-ma@ranepa.ru

этот сдвиг означает переход от ситуации соприсутствия к телесоприсутствию, который имеет последствия для организации привычных практик, в частности обучения. В статье проводится мультимодальный анализ взаимодействия в рамках образовательного вебинара Московской службы психологической помощи населению. Исследуемый кейс характеризуется асимметрией интеракционных ресурсов участников: лектор может говорить, в то время как слушатели могут общаться только с помощью чата. Анализируемый фрагмент взаимодействия включает в себя смену курса действий лектора — переход от монолога к ответу на вопрос в чате. Из-за асимметрии интеракционных ресурсов единственный спикер вынужден организовывать чередование говорящих. Ключевой особенностью упомянутого перехода является самоперебивание, структурно напоминающее вербальное наложение. Диахронное сообщение в чате задействуется лектором как синхронная речь, что демонстрирует его ориентацию на виртуального собеседника. Таким образом, данный кейс показывает, как происходит передача череда безмолвному участнику. Наша работа вносит вклад в исследования образовательного взаимодействия, поскольку формат дистанционного обучения трансформирует привычные интеракционные феномены. Она открывает дискуссии о пересечении различных модальностей взаимодействия, о влиянии новых интеракционных экологий, таких как вебинары, на формы разговора, воплощения и поддержание локального социального порядка. Тем самым наша работа также вносит вклад в исследования, в которых подчеркивается активная роль «слушателя». Последняя выражается в ориентации единственного говорящего на высказывание безмолвного собеседника.

Ключевые слова: телеконференции, исследования образовательного взаимодействия, телеприсутствие, мультимодальность, конверсационный анализ, ко-операционное действие, распределенный спикер, наложение

In the time we live in right now the need for accomplishing tasks from the comfort and safety of your home is needed more than ever: this is true both due to the immediate changes brought about by the Covid-19 pandemic, but also due to the longer-term effects of a societal shift towards work from home, teleconferencing and—to put it in Zhao's terms [2003]—*telecopresence*, our increasing ability to be socially together at a distance. This shift has been in the making for decades, but it has arguably picked up its pace due to the global crisis [Klowait 2019; Klowait, Erofeeva 2019; Mondada et al. 2020].

The shift to greater digitalisation, virtualisation and telecopresence has affected many spheres of modern life — work, study, food delivery, virtual travels, and other forms of entertainment, such as cinema, bars, libraries, etc. This large-scale, rapid move to telemediation also affected an essential for many — psychological help, in forms of personal therapy

or group therapy, or group informational webinars. While webinars may not be a good fit for people in critical situations, they are still in demand as a means of self-help if a personal therapy is not an option. The following paper will undertake an analysis of the interactional shifts precipitated by global events by investigating, in detail, the phenomenon of virtual psychological webinars.

What's Unique about Virtual Webinars?

Naturally, virtual webinars differ from group meetings in real life in many aspects. Firstly, they are a space where a professional psychologist can shed light on some situation, condition, or specialist term. As such, they are largely educational, rather than therapeutic, in nature. However, many people attending such events have personal questions and problems that they want a professional input on. And with a webinar's capability to have 120 attendees, there arises a problem of organising the participants to minimize disturbances to the speaker whilst allowing attendees to speak their minds, ask relevant questions and share experiences related to the topic of the discussion. In short, the webinar organizers face the need to simultaneously support a one-to-many lecture format whilst enabling a sense of telecopresence among the non-presenting participants. This may be one of the reasons why the webinars typically use a system where only the speaker can turn on their camera and microphone—and share their screen—while others are able to participate through a dedicated text-based chat. But, as effective as this system may be, some problems in communication ensue due to its peculiar configuration and distribution of multimodal interactional resources such as gaze, gesture, prosody, body movement, etc.

201

The case that we are going to analyse in this work is taken from such a psychological webinar. People have already adapted quite well to teleconferencing systems such as Zoom, Skype and Microsoft Teams. Yet, in most of these cases, the interactional resources provided by the system are equal for all involved: participants can speak through their microphone, display their webcam, or may even be allowed to share their screen. This wealth of interactional resources arguably also contributes to the smooth and timely accomplishment of the practical goals of the participants, without the limitation of the interactional tools coming to the foreground of the unfolding action. In our case, matters are quite different: when most of the participants can use neither their body (e.g. gaze, facial expressions, gestures) nor voice (e.g. spoken language, prosody and intonation) to initiate or modify a course of action, the procedures necessary for its production may extend over longer stretches of time (for a programmatic treatment of interactionally constrained situations, see [Goodwin 1995]).

We can highlight two principal features of our interactional situation that may require more remedial work on the part of the participants. First, in a virtual meeting the volume of all the speakers does not depend on their proximity to one another — rather on microphone settings, therefore making overlaps potentially more noticeable and more damaging to the conversation. Moreover, the latency and background noise may add to the difficulties. Secondly, a lot of the times participants may not see the other person, due to either their choice to disable the webcam, or due to the system not allowing it. This may increase the difficulties for other participants. This issue has been picked up in a recent review:

Intuitively, a face-to-face setting provides a richer repertoire of cues for coordinating turn-taking than a conversation over the phone and could therefore be expected to be more fluent. For example, seeing each others' faces allows us to perceive gaze direction and facial expressions. However, studies that compare spoken interaction in video meetings with voice-only interactions have not found any substantial differences when it comes to the coordination of turn-taking (O'Connell et al., 1993; Sellen, 1995). But when comparing video conferences to physical face-to-face meetings, O'Connell et al. (1993) found that the former had longer conversational turns, fewer overlaps and backchannels, as well as more formal mechanisms for shifting turns. Thus, it seems like the physical co-presence allows us to more easily pick up these visual cues and coordinate turn-taking. [Skantze 2021: 6]

202

Thus, while the intuitive leap of 'fewer interactional resources = greater interactional effort' appears to be less straightforwardly true, there is some evidence for the greater difficulty of accomplishing certain tasks when faced with diminished resources. That said, Skantze [2021] highlights that the assumption of a greater interactional complexity in mediated environments should not be assumed, but rather substantiated in the analysis of naturally occurring data.

This paper aims to contribute to ongoing research on this matter. The following section will lay the conceptual groundwork for the subsequent empirical analysis.

Conversation Analysis

CA (Conversation Analysis) is a framework of analysing naturally occurring human talk and answering what kind of interactional resources are deployed by participants to accomplish local social action [Sacks, Schegloff, Jefferson 1974]. Building upon the conceptual framework of ethnomethodology [Garfinkel 1967], CA is concerned with a kind of interactional meta-analysis: understanding the methods participants use

to analyse, comprehend, and change the situational normative micro-structures for their own practical purposes. Due to a historic focus on the analysis of recorded voice, CA originally focused on analysing *talk as action*, with lexis, prosody, volume, pace, overlaps, etc. being treated as parts of the interactant's toolkit.

In recent times—and contemporaneously with the emergence of more affordable means of producing videorecording [Heath, Hindmarsh, Luff 2010]—CA has grown to include embodied, non-talk interactional resources in their analyses. This methodological expansion allowed them to analyse silent actions and take into account body movement, gestures, gaze, position of participants in the space and so on, in relation to—but independently from—talk [Goodwin 2000; Mondada 2016, 2019].

More fundamentally, the shift towards *multimodality* [Mondada 2016], i.e. a focus on the study of how discrete interactional toolsets may intersect and be used to accomplish parallel courses of action, had a substantial impact on how CA treated temporality, sequentiality and ultimately the analytical unit called the individual [Goodwin 2018]: if a university professor could use their voice to deliver a lecture whilst using their gaze to manage the students' attention, there would be hardly a good reason to subordinate one to the other. Moreover, if gaze could be analyzed as action, how justified would be the exclusion of the students' gaze as a relevant contributor to the accomplishment of lecturing? Thus, multimodal CA displays a keen awareness of distributed agency — starting from Goffman's initial contributions in *Footing* [Goffman 1979], and continuing in the work on co-operative action pioneered by Charles Goodwin [2006, 2018], and being actively developed by contemporary scholars, particularly in the subfield of atypical interaction analysis, where differently-abled participants often accomplish tasks in unison [Merlino 2018]. For example, when studying everyday activities of people with intellectual disabilities, explore the concept of 'relational autonomy', i.e. "how the ability to have control and agency in one's life requires interdependence, not isolated independence" [Dowling et al. 2019: 2]. In other words, with the methodological loosening of the individual as an acting unit, scholars are increasingly exploring the way participants find themselves in 'intertwined semiosis' [Favareau 2018; Goodwin 2018], or a mutually-oriented, mutually-invoking collaborative meaning-making.

The shift towards this intertwined-ness makes multimodal CA a particularly good fit for telemediated interaction, where participants can oftentimes find themselves equipped with unfamiliar or limited interactional resources, and where resource asymmetries—such as the inability for most participants to speak with a voice—may manifest with regularity.

The Local Ecology of the Webinar

Our focus will be on a case of multimodal interaction in VirtualRoom, an online platform that is currently used by a psychological support service for Moscow residents to hold regular webinars for anyone who wants to attend it. All you need to do is register for a specific webinar of your interest. The main focus will fall upon the situation that exists exclusively in a virtual multimodal environment such as a webinar, where the audio-video modality of the speaker intersects with the modality of chat, in which attendees ask clarifying questions and share personal opinions and experiences related to the webinar topic.

The recording was made on one of the governmental psychological webinars. The website of the Psychological Service of Moscow has a tab with all webinars, which are structured by date and time. When the listener enters the webinar, they see a presentation window, a speaker's window, and a chat room that can be expanded or collapsed. The page view followed below is how listeners see the webinar (fig. 1).

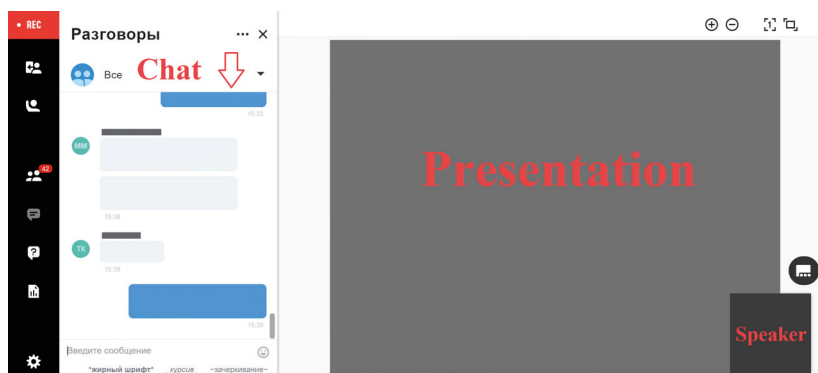


Fig.1. The layout of the webinar.

The number “42” indicates how many listeners there are currently attending the webinar; blue and grey chat clouds are what is written by you as a listener and by other attendees respectively. There are other tools to use on this page, such as a “raise hand” button, a chatroom for questions, settings, a full screen mode, etc. However, in our case only the general chat window and speakers window will be most relevant to the analysis.

The speaker presents their topic of choice, while at the same time reacting to what is happening in the chat window. Only the speaker has the ability to switch their video and audio on and off, switch slides and give permission to download the presentation. Listeners, on the other hand, can only write into the chat and wait for the presenter's reaction once the presenter reads the message. While sometimes specific time

periods are reserved for questions, at other times the questions may be entirely ignored (accidentally or on purpose) — this matter is entirely up to the speaker.

Now, with the understanding of the setting and the approach, we can present and analyse the case at hand itself. In our chosen fragment, the speaker coordinates a sequentially peculiar change to her course of action [Schegloff 2007]. Transitioning from a monologic lecturing format after a 2-second silence, the speaker re-orient to a request for clarification that had appeared in the chat at a prior point. This brings about marked changes in the embodied conduct of the speaker. Most peculiarly, the transition itself appears to be produced as *an interruption by a virtual second speaker*, where the lecturer's re-orientation to the chat 'interrupts' the current course of action. The following section will explore the production of this virtual spoken interruption.

Constructing the Virtual Co-Speaker in Monologic Telecommunication

We are now equipped to tackle our case by demonstrating and explaining how a change in the current course of action occurs. For conventions of transcription see appendix as well as [Jefferson 2004, Mondada 2019a].

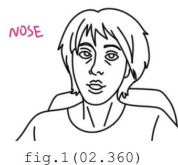
205

First, we are going to look at the features of the embodied production of the monologic lecturing format that preceded the change.

The case starts with the conclusion of the short monologue, a response to the question in line 1. A second after the start, right at the end of the word 'страх' (fear), a comment is sent to the chat, which the speaker, for now, does not see, as she does not shift her eyes from her web camera as if talking directly to the audience.

Transcript 1.

1. >то есть< стр!А!+х+ тор!о!, што м!м:!, а:, п!о!н!у:чим !(!) !йе!сли →
 meaning that the fear of what we will get if
 !hu! !hd! !hd! !hu!!hd! !-hu! !hr!
 +1+
 1: the comment goes at 01.060



Head Movement

The embodied features of the first course of action are expressed in the speaker's cadence and head movement. During the monologic phase, the cadence is regular and supplemented by a rhythmic, regular motion of the head. This can be seen in Transcript 1. A series of head movements indicated by exclamation marks and two letters (hu — head up; hd — head down; hr — head right; hl — head left; hj — head jerk) follow her words. Head movements feature as a frequent embodied component in conversation and can be used by participants to accomplish a multitude of actions, such as emphasis, feedback, turn-taking management and more [Otsuka, Tsumori 2020]. In our case we can see that the head movements follow the speech pattern, with the peak of the movement landing to the stress of some more meaningful words. The head moves from the up position to the down position and then all the way up, even past the initial position. This is called 'batonic' movement: "The speaker's body movements are rhythmically coordinated with the articulated segmentation of his/her speech." [Otsuka, Tsumori 2020: 217172].

206

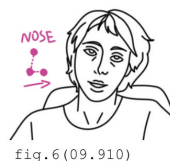
The batonic head movement, however, does not always have the rhythmic 'up and down' pattern. As we can see in the transcript 2 in line 3, the speaker's head moves diagonally down, followed up right after with a jerk to the speaker's left. The head movements, apart from being batonic, visualise the difference between the first part of the recording, intermediate silent sequence and the second part, as the distinct head movements, following the words, is not observed anywhere but the first part of the fragment.

Transcript 2.

2. не сд!е!аем д!ело! с!ли!ш!ко!м вел!ик! для !нас! ! (0.3) ! то →
 don't do the deed is too much for us then
 !hj! !-hd,,!!hl !hj! !hd! !hd! !--hu,!

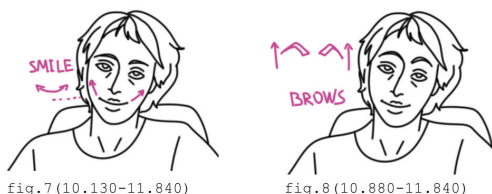
fig.4 fig.5 fig.6

3. м!от!ивация у нас б!удет (0.3) ! н!из!к!ой!
 motivation will be low
 !hd! !hu-----! !hd! !hj!



We enter a silent sequence in line 4 (transcript 3) which later proved to be transitional.

4. (0.3)*(0.8)&(1.0)&
 fig7
 smiles-----[------
 fig8
 [-raises brow-----&



207

Silence can hold different meanings and achieve different actions, depending on the locality of the case.

Within Conversation Analysis, silence has been categorized in different ways, as the absence of talk at different sequential locations with respect to the organization of turns (Hoey, 2017). An intra-turn silence occurring in the middle of a turn, before its completion and before a transition-relevance point TRP¹ is reached, is a pause; an inter-turn silence that occurs within a TRP is a gap; and an extended silence after a TRP is a lapse. [Mondada 2019b]

While a prolonged silence of two seconds may be treated as problematic in everyday interactions, the institutional character of our case casts doubt on this silence's threat to progressivity — an orientation to make progress in conversation with each turn [Schegloff 1980]. The instructor

1 TRP (transition-relevance place) is a place where the utterance comes to a possible completion, and the conversation can move to the next point [Sacks et al. 1974].

holds the floor, both in their role as the instructor, and in the asymmetrical distribution of interactional resources.

The silence may be explained with a reference to a number of situational features. Firstly, the talk reached a presumable sequence completion — in line 3 the intonation drops, followed by a still pose of anticipation. The speaker has answered the question, stated their point, and is now waiting for the response that the answer is received and accepted. She locks eyes with the camera, as if to look like she is locking eyes with the audience, which creates a mutual gaze called a “gaze window”, which is “a strong cue for eliciting responses from listeners” [Ferre, Renaudier 2017]. Moreover, she smiles and shortly after raises brows (transcript 4), while holding still and maintaining the gaze window. These facial expressions, combined with the pose, indicate that she is assuming a waiting mode.

Backchanneling

Backchanneling describes feedback from the listeners that is used to indicate, usually with short utterances, such as ‘uh huh’ and ‘mm hmm’, and nods, that the speaker is being listened to [Clark, Krych 2004]. Thus, a backchannel is an essential communication channel between a speaker and a listener. Under most circumstances, speakers tend to orient to some form of backchannel to know that they are not talking aimlessly into the void, especially when the speaker’s purpose is to educate.

In our specific case, there were diminished resources for backchanneling due to the sparse set of interactional resources made available to the non-talking participants: with their resources being concentrated in the written chat, their ability to generate immediate backchanneling is hindered, with a textual response typically taking more than two seconds to compose. One way for the chat to provide some sort of backchannel is to present relevant questions and comments when they are due, and to provide responses when the speaker prompts the chat directly (and specifically chooses to look at the responses).

It seems that in the present data, the lack of verbally produced backchanneling is compensated by the aforementioned animation of a speaker on the basis of a non-voiced textual contribution by a silent participant.

Change of a Course of Action

The visible re-orientation initiates a change in the course of action in line 5. After the disruption in the flow of talk — the prolonged silence in the place where the backchannel should have been — the speaker’s talk, cadence, and body work is notably transformed (transcript 4).

Transcript 4.

- fig.9
5. И ЭТО очень важный мо=мент (0.4)
and that's a really important moment
6. =gaze----->9



With an increased volume, that can be seen in the first two words in line 5, the speaker launches into a generic gap-filling phrase “and that’s a really important moment”. In the middle of the word “moment” in line 5 her gaze shifts to the approximate direction of the chat window on her screen (line 6). She then proceeds to hold the gaze there for quite a while.

209

While the gaze is on the chat (lines 6-9) marks the likely moment when the speaker becomes aware of the comment that was sent in the first second of the recording. Briefly, she still continues the phrase, saying “это в этом случае” (this in this case) in line 7, followed by a cough. As we can see in transcript 5, the cough occurs in the middle of the phrase (line 7) while the gaze is still locked on the chat, as she, presumably, starts reading the comment, pursing her lips. The cough may be taken as a token necessary for resolving a disjunctive change of a course of action [Couper-Kuhlen 2004], that is an abrupt transition to a next sequence.

There is a long 2-second pause as she reads the comment. She does not indicate what she is doing and why is there a pause. We may assume that—at that point—she is still undecided whether she will continue speaking or change her course of action towards answering the question in the chat. After the silence, she proceeds with answering the question, indicating that its content was taken as requiring elaboration.

The Answer

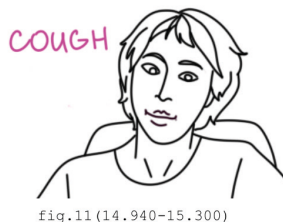
She rapidly answers “of course you can” in line 9, and without missing a beat she incorporates the likely reference of the question (line 10) (transcript 6). This incorporation may be necessary because 17 seconds have passed from when the message was initially sent. Everyone can see the chat and the messages within it, yet the speaker still feels the need to read it out loud, likely to ensure that everyone can understand the new course of action.

Transcript 5.

fig.11
↓

7. >эт!о! в этом случае< (cou%gh) (2)
this in this case
 !hu!

8. %spinning in a chair----->12



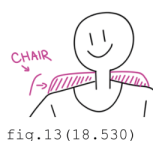
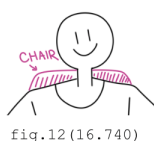
Transcript 6.

9. >коне=чно можно в=от да вы пишите А мож=но выставить
of course you can here yeah you ask can I create
 --= =gaze-----=

10. самому о=птиум мо=тивации конечно можно< (inhale)
my own optimum motivation of course you can
 =gaze----

11. мы с вами об Этом и гов=%
we are talking about that

12. --%



In short, there is a marked shift in the speaker's conduct—from regular, rhythmic narration towards 'compressed' talk, from a comparatively stable bodily position to a swaying chair motion, and with a noticeable engagement with the writing produced by the silent co-participant. In the subsequent section, we will explore the moment of transition in greater detail.

Self-Interruption: A Multimodally Produced 'Overlap'?

The transition from one course of action to the other structurally resembles a verbal overlap, despite this being impossible under the di-

minished and asymmetrically distributed interactional resources in our data. Under regular circumstances, an overlap characterizes a moment in interaction when two or more participants talk at the same time. Shegloff [2000] states that even when the main principle of turn-taking (one party at a time) is mostly achieved, the overlap is inevitable. Jefferson [1973] drew attention to overlaps and stated that the appearance of overlaps is not a trivial error in opening and closing sequences, but a rather organised phenomenon. Even though such a problem as an overlap is an obvious issue for the flow of the talk, overlaps are a part of the turn-taking system, the occurrence of which allows some participants to take the next turn.

Jefferson stated that overlaps usually occur after or during the closing sequence; and once this has happened, the participants resolve the problem in a quite orderly fashion [Jefferson 1973].

There are

three main positions or locations where overlapping talk begins, or onsets; in the transition space, just before the transition space (last-item onset), and just after the transition space (post-transition). Instead of chaos, we find that overlapping talk is systematically associated with participants' close, fine-grained orientation to one another's talk, and particularly to when and how another's turn at talk might be complete [Drew 2009].

211

Yet there are also exceptions. "It does happen that speakers begin speaking whilst another is speaking, at points where the 'current' speaker cannot be close to completing their turn (i.e. at a point which is distant, in some fashion, from a completion or transition point)" [Drew 2009]. This is called an 'interjacent' overlap. In this case one speaker interrupts the other in the middle of the ongoing turn construction unit.

In our data, the speaker interrupts herself in this fashion. The 'overlap' can be seen, but not heard. The modality of the chat is overlapping the modalities of the speaker, who can use video and audio. The text in the chat that presumably overlaps and disrupts the flow of the speech was produced *17 seconds* prior to the actual 'disruption'. Too much time has passed before the text being typed and the response; even if there was some sort of notification system on the speaker's screen, it would have disappeared by the time the speaker turned her gaze towards the chat's location.

We may thus describe this case as a self-interruption, a multimodally produced 'overlap' through the vocal animation [Goodwin 2006] of a second speaking participant: in line 5, the speaker turns to the chat at the last item of an ongoing turn, prompting the change of the course of action discussed previously (lines 6, onward). While, in a vocal multi-party

interaction, the ‘overlap’ between ongoing talk and the vocal contribution of another speaker at a TRP may have unproblematically produced a shift in speakership—with the other participant taking the floor to ask a question, for example—the ‘overlap’ is produced by the speaker, for a non-voiced participant; a kind of vicarious interruption is thus the visible outcome.

This is notable on its own, but becomes a greater interactional challenge when the new (re-oriented) talk is produced at an interjacent on-set (line 7), i.e. when a new speaker begins a turn at a point where the current speaker is recognizably not approaching completion. In other words, the only speaker not only produces an animated second speaker by attending to non-voiced chat messages and recruiting them as spoken, transition-able overlaps, but does so in a sequential position that requires a greater amount of work to retain the floor — as was demonstrated in our analysis of the cadentially-modified, compressed, bodily-supported nature of the talk after line 5.

Thus, the disjunctive production of the change in the course of action orients to the silent agency of the non-speaking participants by ‘resurrecting’ their ability to take the speaking floor. The interactional work expended by the lecturer for the accomplishment of this resurrection provides evidence for the sustained mutual orientation of the participants, despite the technical circumstances that make the non-speakers appear entirely passive. In sum, the co-operative nature of the organization of the one-to-many speaker-listener format powerfully highlights the distributed nature of speaking, even in situations of ‘technologically diminished’ interactional resources.

Conclusion

Our case highlights several features typical of video-mediated communication, where not only several modalities are deployed, but they differ in what resources the speaker and the listeners can use. This asymmetry creates a situation where familiar actions—such as a question asked at a TRP—become problematic due to the modal disparities across participants (voice vs. textual chat). As a consequence, classical features of face-to-face conversations such as verbal overlaps, which cannot naturally arise due to the aforementioned asymmetry, are reinvented through the other-initiated animation of a second speaker.

In our data the self-interruption of the only speaker works as a multimodally produced ‘overlap’. The talk and the abrupt stop of it looks like the speaker has been stopped by someone, as if actual verbal overlap has happened. Although, the text has been sent long before the response. The speaker deliberately chose to be interrupted. She suspended

her activity to give a chat question a turn, in the middle of her ongoing contribution. In other words, our case features the creation of a virtual co-speaker, who is made to competently deploy spoken turn-taking resources.

Our work opens up discussions about the interfacing between different modalities as a locally emergent phenomenon [Sicoli 2016], and how new interactional ecologies create a fertile substrate for hitherto unfamiliar forms of talking, embodiment, and local sequential ordering [Mlynář, González-Martínez, Lalanne 2018]. The work thus also contributes to research that highlights the non-passive role of the ‘listener’ [Goodwin 1981], which is reflected in the active speaker’s orientation to the listener’s active contribution to ongoing talk.

References

- Clark H. H., Krych M. A. (2004) Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50 (1): 62–81. <https://doi.org/10.1016/j.jml.2003.08.004>
- Couper-Kuhlen E. (2004) Prosody and sequence organization in English conversation. E. Couper-Kuhlen & C. E. Ford (Eds.), *Typological Studies in Language. Sound Patterns in Interaction* (Vol. 62, pp. 335–376). Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.62.17cou>
- Dowling S., Williams V., Webb J., Gall M., Worrall D. (2019) Managing relational autonomy in interactions: People with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities*, 32 (5): 1058–1066. <https://doi.org/10.1111/jar.12595>
- Drew P. (2009) Quit talking while I’m interrupting: a comparison between positions of overlap onset in conversation. M. Haakana, M. Laakso, & J. Lindström (Eds.), *Comparative Aspects of Conversation Analysis* (pp. 70–93). Finnish Literature Society. Retrieved from https://repository.lboro.ac.uk/articles/chapter/Quit_talking_while_I_m_interrupting_a_comparison_between_positions_of_overlap_onset_in_conversation/9472022
- Favareau D. (Ed.) (2018) *Tartu semiootika raamatukogu: Vol. 19. Co-operative engagements in intertwined semiosis: Essays in honour of Charles Goodwin*. Tartu: University of Tartu Press.
- Ferre G., Renaudier S. (2017, August 15) Unimodal and Bimodal Backchannels in Conversational English. In *SEMDIAL 2017 (SaarDial) Workshop on the Semantics and Pragmatics of Dialogue* (pp. 20–30). ISCA: ISCA. <https://doi.org/10.21437/SemDial.2017-3>
- Garfinkel H. (1967) *Studies in ethnomethodology*. Englewood Cliffs: Prentice-Hall.
- Goffman E. (1979) Footing. *Semiotica*, 25(1-2):1–30. <https://doi.org/10.1515/semi.1979.25.1-2.1>

Goodwin C. (1981) *Conversational organization: Interaction between speakers and hearers*. New York, London: Academic Press.

Goodwin C. (1995) Co-Constructing Meaning in Conversations With an Aphasic Man. *Research on Language and Social Interaction*, 28(3): 233-260. https://doi.org/10.1207/s15327973rlsi2803_4

Goodwin C. (2000) Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32(10): 1489-1522. [https://doi.org/10.1016/S0378-2166\(99\)00096-X](https://doi.org/10.1016/S0378-2166(99)00096-X)

Goodwin C. (2006) Interactive Footing. E. Holt & R. Clift (Eds.), *Reporting Talk* (pp. 16-46). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486654.003>

Goodwin C. (2018) *Co-operative action. Learning in doing*. Cambridge: Cambridge University Press.

Heath C., Hindmarsh J., & Luff P. (2010). *Video in qualitative research: Analysing social interaction in everyday life*. Los Angeles, London: SAGE.

Jefferson G. (1973) A Case of Precision Timing in Ordinary Conversation: Overlapped Tag-Positioned Address Terms in Closing Sequences. *Semiotica*, 9(1). <https://doi.org/10.1515/semi.1973.9.1.47>

Jefferson G. (2004) Glossary of transcript symbols with an introduction. Lerner, G. H. (Ed.), *Conversation Analysis: Studies from the First Generation*. John Benjamins, Amsterdam, NL: 13-31.

Klowait N. O. (2019) Interactionism in the age of ubiquitous telecommunication. *Information, Communication & Society*, 22(5): 605-621. <https://doi.org/10.1080/1369118X.2019.1566487>

Klowait N. O., Erofeeva M. A. (2019) Work in the Age of Intelligent Machines: The Rise of Invisible Automation. *Philosophical Literary Journal Logos*, 29(1): 53-84. <https://doi.org/10.22394/0869-5377-2019-1-53-80>

Merlino S. (2018) Co-Constructing Wor(l)ds in Aphasia Speech Therapy. D. Favareau (Ed.), *Tartu semiootika raamatukogu: Vol. 19. Co-operative engagements in intertwined semiosis: Essays in honour of Charles Goodwin* (pp. 287-303). Tartu: University of Tartu Press.

Mlynář J., González-Martínez 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. <https://doi.org/10.1093/iwc/iwx019>

Mondada L. (2016) Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, 20(3): 336-366. https://doi.org/10.1111/josl.1_12177

Mondada L. (2019a) Conventions for multimodal transcription. Available at: <https://www.lorenzamondada.net/multimodal-transcription>

Mondada L. (2019b) Transcribing silent actions: a multimodal approach of sequence organization. *Social Interaction. Video-Based Studies of Human Sociality*, 2(1). <https://doi.org/10.7146/si.v2i1.113150>

Mondada L., Bänninger J., Bouaouina S., Gauthier G., Hänggi P., Koda M.

- Tekin B. S. (2020) Changing social practices. Covid-19 and new forms of sociality. *Etnografia e ricerca qualitativa*. (2): 217-232. <https://doi.org/10.3240/97807>
- Otsuka K., & Tsumori M. (2020) Analyzing Multifunctionality of Head Movements in Face-to-Face Conversations Using Deep Convolutional Neural Networks. *IEEE Access*, 8: 217169-217195. <https://doi.org/10.1109/ACCESS.2020.3041672>
- Sacks H., Schegloff E. A., Jefferson G. (1974) A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50 (4): 696. <https://doi.org/10.2307/412243>
- Schegloff E. A. (1980) The Relevance of Repair to Syntax-for-Conversation. T. Givon (Ed.), *Discourse and Syntax* (pp. 261-286). BRILL. https://doi.org/10.1163/9789004368897_012
- Schegloff E. A. (2000) Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29 (1): 1-63. <https://doi.org/10.1017/S0047404500001019>
- Schegloff E. A. (2007) *Sequence organization in interaction: A primer in conversation analysis*. Cambridge, New York: Cambridge University Press.
- Sicoli M. A. (2016) Repair organization in Chinantec whistled speech. *Language*, 92 (2): 411-432. <https://doi.org/10.1353/lan.2016.0028>
- Skantze G. (2021) Turn-taking in Conversational Systems and Human-Robot Interaction: A Review. *Computer Speech & Language*, 67: 101178. <https://doi.org/10.1016/j.csl.2020.101178>
- Zhao S. (2003) Toward a Taxonomy of Copresence. *Presence: Teleoperators and Virtual Environments*, 12 (5): 445-455. <https://doi.org/10.1162/105474603322761261>

Appendix

Multimodal transcribing conventions [Mondada, 2019a]

* * Descriptions of embodied actions are delimited between
+ + two identical symbols (one symbol per participant and per type
of action)

Δ Δ that are synchronized with correspondent stretches of talk or
time indications.

*---> The action described continues across subsequent lines
---->* until the same symbol is reached.

>> The action described begins before the excerpt's beginning.

--->> The action described continues after the excerpt's end.

..... Action's preparation.

---- Action's apex is reached and maintained.

,,,, Action's retraction.

fig — The exact moment at which a screen shot has been taken

Рекомендация для цитирования:

Kimstach U. M., Klowait N., Erofeeva M. A. (2021) Talking without a Voice: Virtual Co-Speakership in an Educational Webinar. *Социология власти*, 33 (4): 198-216.

For citations:

Kimstach U. M., Klowait N., Erofeeva M. A. (2021) Talking without a Voice: Virtual Co-Speakership in an Educational Webinar. *Sociology of Power*, 33 (4): 198-216.

Поступила в редакцию: 08.09.2021; принята в печать: 11.11.2021

Received: 08.09.2021; Accepted for publication: 11.11.2021