Navigating the complex social ecology of screen-based activity in video-mediated interaction

Ufuk Balaman and Simona Pekarek Doehler
Hacettepe University | University of Neuchâtel

Task-oriented video-mediated interaction takes place within a complex digital-social ecology which presents, to participants, a practical problem of social coordination: How to navigate, in mutually accountable ways, between interacting with the remote co-participants and scrutinizing one's own screen –which suspends interaction–, for instance when searching for information on a search engine. Using conversation analysis for the examination of screen-recorded dyadic interactions, this study identifies a range of practices participants draw on to alert co-participants to incipient suspensions of talk. By accounting for such suspensions as being task-related through verbal alerts, typically in the form *let me/let's X*, participants successfully 'buy time', which allows them to fully concentrate on their screen activity and thereby ensure the progression of task accomplishment. We discuss how these findings contribute to our understanding of the complex ecologies of technology-mediated interactions.

Keywords: video-mediated interaction, screen-based activity, conversation analysis, L2 interaction, progressivity

1. Introduction

Geographically dispersed participants' video-mediated interactions (henceforth VMIs), require moment-by-moment coordination between individual participants' 'private' (i.e., mutually non-accessible) orientations to screens and their 'public' participation to ongoing talk-in-interaction (Heath and Luff 1993, 2000; Jenks and Brandt 2013; Oittinen and Piirainen-Marsh 2015). Such contextual requirements become particularly significant in online task-oriented settings where task accomplishment is largely dependent on the successful management of the coordination work (Balaman and Sert 2017a; Balaman 2018, 2019), and

where individual participants' orientation to the multisemiotic resources (e.g. texts, images; cf. Goodwin 2013, 2018) for task-accomplishment made available through screens may suspend joint engagement in talk-in-interaction, and therefore possibly cause interactional trouble (Brandt 2011; Brandt and Jenks 2013; Balaman and Sert 2017b; Sert and Balaman 2018). In these situations, participants are faced with the practical problem of navigating, in mutually recognizable ways, between social interaction with remote co-participants and the scrutiny of their own screen, for instance when searching for information on a given search engine (Näslund 2016). While exactly these searches are instrumental for the accomplishment of the joint task, they represent a potential source of interactional trouble, as they typically suspend talk, and hence may impede the progressivity of social interaction (Rintel 2010, 2013; Olbertz-Siitonen 2015).

As a specialized speech exchange system (Arminen, Licoppe, and Spagnolli 2016; Licoppe and Morel 2018, going back to Sacks, Schegloff, and Jefferson 1974), task-oriented VMIs involve an organization of turn-taking that differs from ordinary conversation. In particular prolonged suspension of talk may be instrumental in task accomplishment – and hence mutually agreed upon –, as it allows (individual) participants to fully concentrate on screen activity in view of retrieving task-relevant information. Such silences designed for information retrieval also typically enable the results of these retrievals to be subsequently brought into the 'public' space of the interaction for the purpose of joint task accomplishment. In this sense, while they may suspend talk, silences often contribute to moving the task forward, i.e., to fostering the progressivity of the overall activity. This, however, does not mean that prolonged silences are normatively treated as acceptable by participants. On the contrary, task-oriented video-mediated interaction has been found to include many instances during which suspension of either the progressivity of talk-in-interaction or task-accomplishment are treated as problematic by participants (Brandt 2011; Brandt and Jenks 2013; Balaman and Sert 2017b; Balaman 2018). Accordingly, one issue for the participants in task-oriented VMIs is to consistently display cessation of talk as related to task-accomplishment, rather than as due to a random non-task-related matter.

Against this backdrop, we examine how participants navigate the complex digital-social ecology (Luff et al. 2003) of VMI at the interface of social interaction and individual screen-based activities. Based on screen-recording data collected from task-oriented video-mediated second language interactions that were part of a virtual exchange project between geographically dispersed partners, the current study describes how the participants to dyadic VMIs deploy verbal expressions (most often: *let me X*) in concert with other resources such as non-lexical vocalizations or grammatical projection to alert each-other to incipient suspensions of talk and thereby lay the ground for their own subsequent 'private' screen-based

activities (i.e., activities that are not observable to co-participants; cf. Heath and Luff 1993, 2000). We show how these resources co-operate in ways that respond to locally occasioned needs for individuals' screen orientation and systematically precede their screen-based activities. We demonstrate how participants' signaling of upcoming and ongoing screen-based activities is highly context-bound and context-shaping in VMI, how it pre-emptively accounts for incipient breaks in the progressivity of talk-in-interaction *as being task related*, and how, thereby, it creates affordances for participants to orient to the textual and more generally visual information available on their respective screens in interactionally non-disruptive and mutually recognizable ways. We discuss how these findings contribute to our understanding of the complex ecologies of technology-mediated interactions, and the "new agencies and accountabilities effected through reconfigured relations of human and machine" (Suchman 2007, 12).

2.   Progressivity and video-mediated interaction

Participants to social interaction are centrally concerned with the moving forward of their interaction, that is, with its progressivity. Progressivity is at stake at all levels of organization, including the step-by-step moving forward of words, turns or sequences of actions (Schegloff 1979, 2007). Interactional trouble of various sorts may impede on progressivity at any of these levels (for seminal observations see Schegloff 1979 on repair and Goodwin and Goodwin 1986 on word-searches). Interactional repair provides a paramount example for the fact that the "conjoint operation of the principles of progressivity and intersubjectivity [i.e. establishing mutual understanding]" (Heritage 2007, 260), may result in conflict: In the case of repair, the principle of intersubjectivity may invade the principle of progressivity, as the moving forward of talk is momentarily suspended.

The importance of progressivity for participants is highlighted by the fact that participants employ various means for minimizing disruption when encountering trouble, and thereby maximizing the compatibility between the principles of intersubjectivity and of progressivity. For instance, at the level of turns at talk, they may resort to precise grammatical constructions such as syntactic pivots to minimize the disruptiveness of self-repair by warranting syntactic continuity between the repairable and the repair (Pekarek Doehler 2011; Pekarek Doehler and Horlacher 2013). At the level of sequence progressivity (e.g., Schegloff 2007), questioners may relax or drop the preference for the next selected speaker to respond when progressivity is significantly impeded, allowing non-selected speakers to provide the response (Stivers and Robinson 2006).

Importantly for our purpose here, progressivity at the level of talk may be interfering with or, on the contrary, fostering progressivity at the level of the overall joint activity. Specifically in contexts of multiactivity (Haddington et al. 2014), where the situation at hand requires participants for instance to engage both in interacting with each other and in retrieving information from a screen, participants may orient to the "exclusive orders" (Mondada 2014) of their multiactivity, in which one activity (e.g., talk) is typically suspended in favor of another (e.g., screen-based search). This is exactly what we see participants do in our task-based VMIs (see also Pekarek Doehler and Balaman 2021).

In the data under scrutiny, one potentially central source of disruption and trouble is the mutual non-accessibility (Heath and Luff 1993, 2000; Brandt 2011; Jenks and Brandt 2013; Oittinen and Piirainen-Marsh 2015) of participants' individual screen activities. This entails what Heath and Luff (2000, 86; see also 1993) refer to as "asymmetric access to each other's activities", as one participant's manipulations of the screen may not be inspectable to co-participants for what they are. Mutually non-accessible is not only the nature of these activities, but also the very fact that participants engage in such activity at a given moment in time. That is, others' suspension of talk cannot be inspected by participants for what it is, and such local asymmetry in access may lead to interactional trouble, as illustrated in Extract 1:

Extract 1. Problematization of prolonged silences in task-oriented VMIs

```
1  ZEH:  ↑o::h (.) i found that $↓hahhhh$                        (Balaman 2018)
2        (1.1)
3        in germany (0.5) a city in Germany
4        (1.9)
5        er:
6        (5.3)
7  DEN:  [what is it?
8  NUR:  [yeah?
```

Taken from the very speech exchange system that is the focus of this study, the extract shows how a participant (ZEH) in a hinting/guessing task initiates hinting by delivering a verbal clue (a city in Germany), then pauses (l. 4), and produces a hesitation marker projecting more to come, yet then remains silent (l. 6) while searching for further cues on her screen (not noted in transcript). The co-participants then explicitly problematize this silence through requests for information (what is it?) and for continuation (yeah?), possibly due to the fact that the silence has not been made recognizable to them as being task-related, i.e. being related to ZEH's searching for hinting cues on her screen.

This is exactly the key issue we address here: individual participants' engaging in a concurrent course of action that suspends talk-in-interaction may not be accountable to co-participants as what it is (Whalen 1995; Luff et al. 2016; Hjulstad 2016; Oittinen 2018; for an overview of research on VMI see Mlynar, Gonzalez-

Martinez, and Lalanne 2018), and therefore may create interactional trouble. Yet, specifically in task-oriented interactions where relevant information needs to be retrieved from the screen, such engagement in concurrent courses of action is a prerequisite of task accomplishment: While it halts the progressivity of talk, it affords the progressivity of the overall task-related activity. As such, it is part of the specific type of multiactivity pertaining to task-oriented VMIs.

How concurrent activities on screens can impact on social interaction has been the object of much research. In an early statement, Whalen (1995) for instance observed how, during emergency calls to 911, the position of the cursor on the call-taker's computer menu can influence the order in which she asks questions, and how this may ensue in interactional trouble as the order of the questioning may seem unmotivated or confusing to the caller exactly because of her lack of access to the call-taker's activity on the computer screen. Likewise, concerning telephone conversations in a service encounter setting, Näslund (2016) shows that screen-focused silences that temporarily put the caller on hold for the call-taker to retrieve relevant information from the screen are preceded by explicit and implicit requests for silence and their granting. In other words, call-takers pre-empt the problematization of suspensions of talk by the caller by seeking agreement for the incipient suspension and displaying that suspension as related to the caller's reason-for-the-call. In this sense, silences due to screen orientations are 'contextually conditioned' (Bilmes 1994) and mutually treated as unproblematic.

Especially in video-mediated interactions, further trouble may arise due to distortedness of gaze behaviors that stems from participants' asymmetrical access to the interactional space (Heath and Luff 1993, 2000). Similarly, technological troubles such as transmission delays might impede on progressivity (Olbertz-Siitonen 2015), although such troubles have also been found to operate as affordances for the maintenance of the interaction (Rintel 2010, 2013). Therefore, the overall lack of co-presence in VMI settings occasions the emergence of both context-specific affordances and constraints (Arminen, Licoppe, and Spagnolli 2016). Participants' physical movements (e.g., a mouse click) and orientations to textual and visual objects on the screen (Näslund 2016), as well as ongoing talk, concur as intertwined modalities in VMI (Gardner and Levy 2010; Balaman and Sert 2017a; Knight, Dooly, and Barbera 2018; Balaman 2019) as a highly specialized speech exchange system. Thus, the coordination of multiple activities becomes consequential for the maintenance of the overall joint activity's progressivity (Arminen, Koskela, and Palukka 2014; Due 2015; Näslund 2016).

In this paper we document how participants designedly bypass potential interactional trouble due to breaks in talk by displaying these breaks as being due to screen-based activities, and hence as being instrumental in moving forward towards joint task-accomplishment. They do so by drawing on recurrent

constellations of interactional resources, namely verbal alerts (in the form of *let me X*, and more rarely *let's X*), non-lexical vocalizations such as *hmm*, and the projection potential of grammatical turn-trajectories. They use these resources in situated ways so as to alert co-participants to the incipient (and sometimes ongoing) nature of suspensions of talk, prospectively accounting for these *as being task-related*, and thus pre-emptively displaying cessations of talk-in-interaction as being in the service of the overall progressivity of the joint task-based activity.

3.   Data and procedure

The data for this study come from screen-recorded task-enhanced VMIs among 4 dyads over a period of three weeks (14 hours). The participants use Skype for educational purposes (also known as Virtual Exchange and online intercultural exchange; see O'Dowd and Lewis 2016) as part of a telecollaborative partnership between two universities (one in Turkey and one in Denmark) designed to provide opportunities for intercultural exchange. As such, the data are an integral part of an educational program and represent a type of educational setup that is currently expanding around the world. Each dyad met once a week to complete two online tasks designed in ways that respect the nature of distant settings (i.e., by encouraging screen-based activities), provide opportunities for intercultural awareness (e.g., by addressing issues of food, culture, music, cinema of the respective countries), and facilitate continuous meaning negotiation. Participants received instructions via e-mail, including a written explanation for the task procedure and an instruction video. During each virtual exchange meeting, each dyad went through the instruction e-mails, started the recordings using the online screen-recording software, met on Skype, and completed the tasks based on the instructions. The screen-recording software operated in the background on each participants' computer separately; it successfully captured talk, screen-based activities, and computer sounds, and automatically uploaded the videos to a remote server at the end of each session.

We examined the dataset using conversation analysis and identified that the participants deploy a diverse set of practices to alert each other to their incipient screen-based activities. These include verbal alerts that make explicit the incipient orientation to the screen (e.g. *let me check, let me see, let's find*), the use of non-lexical vocalizations such as *uhhmmm*, as well as grammatical projection (i.e., the suspension of grammatical turn-trajectories at points of maximal grammatical control, cf. Schegloff 1996, 93–94). In this study we focus on the interactions of one dyad (SIN and PAT) for reasons of data quality and completeness (106 min.)

and in order to better show the mutual recognizability of the focal phenomenon in and through one dyad's context-bound interactional history.

Our collection comprises 27 instances in which either one of the participants alerts the other to her incipient screen-based activity, in all cases resorting to a verbal alert in the form of *let me X*, or *let's X* (schematized as *let (me) X* in what follows), which is then followed by other resources (see above). Here, we present four extracts that represent the recurrent features of these alerts. The first three extracts are taken from one task in particular, namely a movie-guessing task. Both participants were given a task-instruction document including a list of 10 movies. Each participant had to pick a movie title from this pre-established list and provide hints for the co-participant so that the co-participant could guess the name of the movie (see Figure 2 below). Once a guessing round was over, the roles switched: The former guesser became the hint-provider, and so forth. Figures 1 and 2 (below) illustrate captures of the dyad's screens (note that Courier New text, such as 'Skype frame' in Figure 1, represents our own labeling of what appears on the screenshots) during task engagement and show how participants' (limited) visual access to each other was provided through a small Skype frame (see above for the distortedness of gaze). The fourth extract is taken from another task from the same dyad in which the participants are expected to move around a street view map to collaboratively create a list of top restaurants for a one-day food trip. As for the asymmetries documented in what follows, we should note that Extracts 2 and 4 show SIN deploying screen-based activities that are not accessible to her co-participants, while 3 and 5 display PAT doing so.
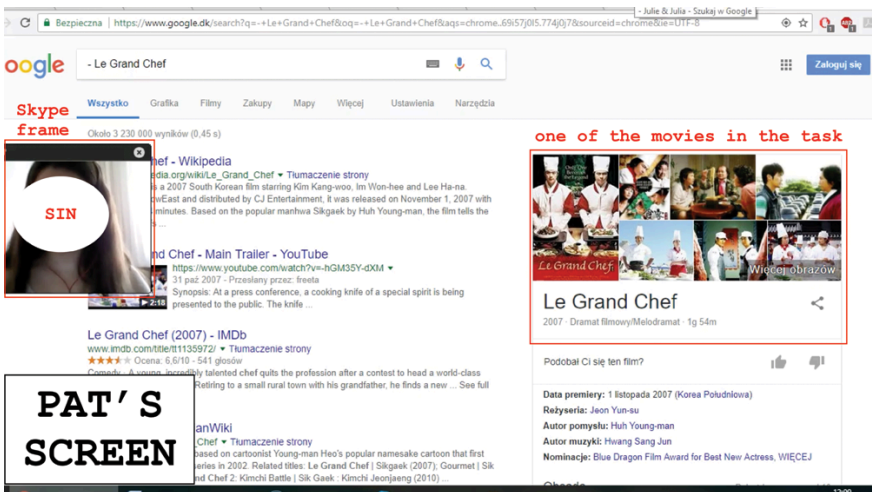

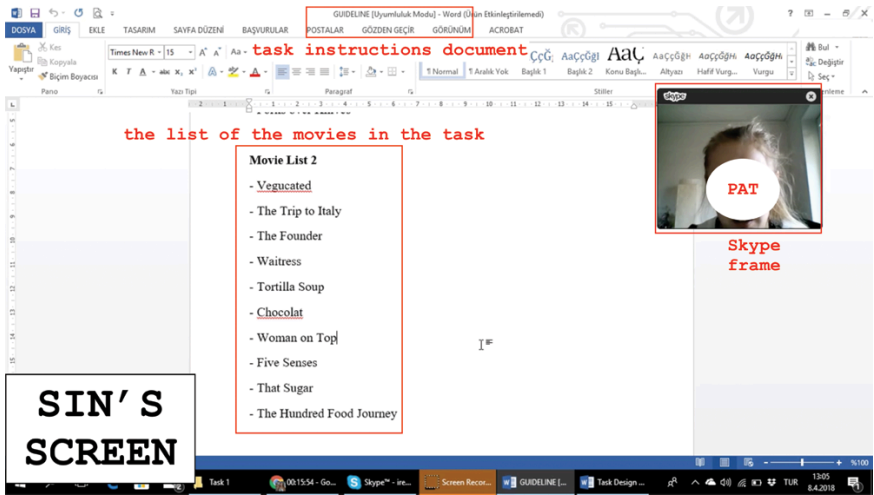**Figure 1.** A screenshot of PAT's screen during task engagement

**Figure 2.** A screenshot of SIN's screen during task engagement

4.   Analyses

In this section we document how the participants use verbal resources, such as *let me X,* to alert their co-participant to their incipient screen-based activity, thereby accounting for an incipient silence while at the same time offering a bid for cessation of talk. We show how such accounting is locally prolonged by subsequent uses of non-verbal-lexicalizations, syllable lengthening and grammatical projections using the analytic tools of conversation analysis. We also provide supplementary observations showing how the delivery of the verbal alert is closely synchronized with the speaker's physically manipulating the screen (not accessibly to the co-participant), through moving the cursor or clicking at certain items, and how it coincides with major steps in the search process, such as shifting from one web-resource to another. We argue that, by preemptively accounting for cessation of talk-in-interaction and thereby creating space for retrieving task-relevant information from the screen without causing interactional trouble, the use of the above resources allows the participants to foster the progression of task-accomplishment.

Extract 2 starts just after SIN has identified the title of a movie "Ratatouille" (line 1) based on PAT's providing hints. Following the closing of the preceding round ([(>correct<), line 4; yes exact↓ly, line 6), SIN signals the transition into a new round in lines 7 (okay) and 9 (>okay<). SIN picks a movie form the list, namely "That Sugar", and immediately turns to the screen to identify possible clues to be offered to PAT by conducting a screen-based web search (1# in line 9).

Extract 2. Movie-guessing task (That Sugar): 18.55–20:37

```
1  SIN: okay I think it is Ra [tatou i $hah hahhh$
2  PAT:                       [(but)
3       (0.7)
4  PAT: yeah [(>correct<)
5  SIN:      [$hah hahhh$
6  PAT: yes exact↓ly=
7  SIN: =okay
8  PAT: .h $↑hah$ (.) nah-
9  SIN: >okay< let me (0.5) 1#just (0.5)
10      make a search#1
       1# - SIN clicks on the browser icon to open Google.
11      2#(2.1)#2
       2# - SIN types "that sugar" and executes the search.
```
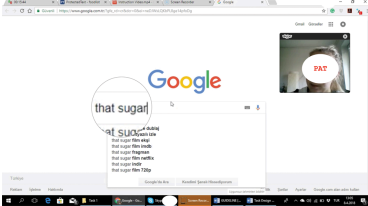


```
12 PAT: uh huh
13      3#(4.0)
       3# - SIN scrolls up and down on the search results page and holds the cursor on the IMDB result at #3 in line 18.
```
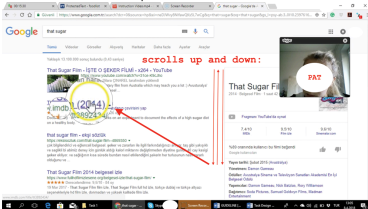


```
14 SIN: uhhmmmm
15      (3.8)
16      errmmmm
17      (3.1)
18      let#3 me (0.7) 4#check (0.2)
19      /I/ /em/ /di/ /bi:/
20      (2.9)
21      >I think #4it's5#< (0.3) er:m (0.7) erm:#5
       4# - SIN clicks on and changes the tab to the IMDB page.
       5# - SIN scrolls down to click the guideline document minimized on the task bar and opens it at #5 in line 21.
```
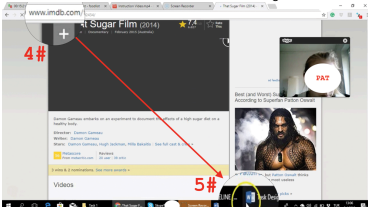


```
22      6#sashhh (.) am I allowed to °say that#6?°
       6# - SIN moves the cursor on to the paragraph on the Word document which includes information about task rules.
23      7#(0.6)
       7# - SIN scrolls up on Word and OKAY in line 26 is produced when the cursor is on the "things to say" list available
       in the Word document.
24      just- >°let° me-<
25      (2.2)
26      OKAY#7
27      (0.5)
28 PAT: yeahh [8#just
29 SIN:       [errm
       8# - SIN returns to the IMDB page, scrolls up and down, holds the cursor still on the part where the director
       information is provided until #8 in line 39.
30 PAT: $sa:: ↑hah hah hah$
```

```
31 SIN:      it's (0.7) originated i:n Australia,
32 PAT: °australia [okay°
33 SIN:           [erm
34     (1.0)
35     and
36     (2.2)
37     the (0.3)
38     <director of the movie also the sta:r (0.3)
39     of the movie#8.>
40     9#(2.1)
       9# - SIN scrolls down to return to Word at #9 in line 45.
41 PAT: oka:y
42 SIN: and ↑in the movie
43     (1.9)
44     er:mmm
45     (2.8)#9
46     °10#let me che:ck°
       10# - SIN opens the guideline Word document and moves the cursor closer to the "things to say" part.
```
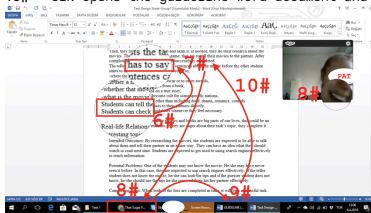


```
47     (2.1)
48     uhmmmmmmm
49     (0.6)
50     i think it's like (0.2) more like a::
51     (1.5)
52     not a do (0.3) cumentary but it's (0.3)
53     more like a documentary because (0.2)
54     it gives you the (0.7)
55     <it gives you the facts> (0.3)
56     about (0.3) the life itself
57     (1.0)
```

Subsequent to her turn-initial >okay< in line 9, SIN explicitly informs about an incipient search: let me (0.5) °just-° (0.5) make a search. She uses the formula *let me X* not to request permission from PAT, but to offer an alert to the recipient as to what she is about to do. Given the fact that SIN's screen is not accessible to PAT, and that her gaze-movement might be hard to spot on the small Skype frame, SIN's verbalization works as a means of making her (upcoming) action, namely a web search, recognizable to PAT. It offers a pre-positioned account for the upcoming silence (2.1 sec.), and, by the same token, blocks self-initiation by the co-participant (see the long silences that follow, e.g., lines 11 and 13, interspersed only by PAT's non-floor-claiming uh huh in line 12). By suspending in this way the turn-taking machinery (Sacks, Schegloff, and Jefferson 1974), the *let me X* projects a slot that affords the possibility for both SIN and PAT to do something else than talking (see 2# and 3#) without their silence being heard as a disruption of the joint activity. It is this framing of the incipient search *as* a search – contextually recognizable as a search related to the joint task at hand – that allows for the ensuing intermittent silences, each of which lasts several seconds (lines 11, 13, 15, 17), to be treated as interactionally non-problematic: By remaining silent, PAT

displays recognition and acceptance of SIN's screen-based activity and SIN herself merely utters two non-lexical vocalizations in lines 14 (uhhmmmm) and 16 (errm-mmm), thereby prolonging the display of her ongoing activity by vocally suggesting that she is still searching (she here possibly orients to the fact that alerts – even overt ones – may have a limited scope, and the prospective accounting needs to be renewed after some time). While this initial alert to an incipient search is here made explicit through the verb-phrase *make a search*, it occurs in the data in various forms, most commonly as *let me check* (also see Extracts 4 and 5, sometimes preceded by *just* (i.e. *just let me check)*, and as *let me see* (Extract 3) and *let's find* (Extract 5).

One such instance occurs in line 18 in Extract 2, where let me (0.7) check (0.2) is followed by the complement IMDB, indicating the object of SIN's checking, namely a precise movie information website. Following a series of prolonged silences, SIN uses the *let me check* construction while privately holding the cursor on the IMDB website (#3, line 18) and then clicking and opening the website (4#, line 18). Similar to the previous instance, the production of the construction is interspersed with short pauses, possibly rhythmed by SIN's concurrent moving of the cursor on the screen. Also similar to the previous instance, it is followed by a significant break in the progressivity of talk, with 2.9 seconds of silence – and is hence treated by PAT as a bid for the suspension of talk. Subsequently, SIN delivers the potential onset of her hinting (>i think it's<) but suspends that trajectory and instead engages in a rule negotiation (Sert and Balaman 2018) in line 22 (am I allowed to say that?). Also note that simultaneously with these verbalizations, but in ways that are not accessibly to the co-participant, she changes the screen to MS Word, opens the guidelines document which includes a list of the pre-established task rules (5#), and moves the cursor closer to the part with the list of rules (6#). Ensuing this prolonged search and the subsequent 0.6s silence (line 23), she then deploys another '*let me*' construction (this time cut off: just- >°let° me-<), again followed by a longer silence (2.2s, line 25), and finally holds the cursor on top of the rules list in the document which is synchronously marked in talk with the loud production of OKAY in line 26, possibly as a way of making public the end of her precedingly 'private' search. It is exactly after this OKAY that PAT starts to produce a response (but is overlapped, line 29), and her preceding long silence can be interpreted as displaying her understanding that SIN is engaged in a screen search. It is also exactly at this moment that SIN moves into the hinting activity, providing the first hint in line 31 (but see already her pre-beginning in line 29; Schegloff 1996, 92–93; pre-hinting, Balaman 2019), and then a subsequent hint in lines 38 and 39, both of which are acknowledged by PAT (lines 32 and 40). While starting a next hint in line 42, however, SIN stalls and suspends talk at a moment

where the syntactic trajectory of her turn-so-far projects more to come. After a 1.9s pause she displays her searching by means of the non-lexical vocalization `er:mmm`, and then, after further suspension of her verbal activity while moving her cursor on the screen (#9, line 45), she ultimately again resorts to `let me che:ck` (line 46), this time without any complement. The construction is once again both preceded and followed by lengthy silences (lines 45 and 47) and functions as a signal of another incipient screen-based activity. SIN also produces a non-lexical vocalization in line 48 (`uhmmmmmm`) and continues hinting with an `i think`-prefaced turn in line 50, which she however suspends at a point of maximal grammatical control (Schegloff 1996, 93–94), thereby deploying a further resource for holding the floor and suggesting that she is engaged in searching during the long ensuing silence (which is further enhanced by the syllable lengthening on `a::` in line 50). PAT's remaining silent (lines 49 and 51) again testifies to the accountability of SIN's non-lexical vocalization and suspension of grammatical trajectories, in the sequential context where they occur, as bids to suspend talk for the purpose of her fully orienting to her screen-based activity.

Extract 2 has shown a range of features that are recurrent in the data: (1) participant's use of verbal alerts in environments of notable suspensions of talk-in-interaction that are mutually treated as unproblematic, (2) the occurrence of non-lexical vocalizations after such initial alerts that prolong the indexing of an ongoing search, sometimes also the occurrence of grammatical projections for the same purpose, (3) a complementary distribution of above elements to cumulatively mark the stepwise progression of the overall task (which presupposes participants' consulting the screen) as well as jointly accounting for breaks in the progressivity of talk. We also observe (4) minute synchronization of the speaker's verbal alerts with her screen-based activity (although this activity is not accessible to co-participants). Before considering these in detail, let us stress that (1) through (3) are analyzed from the participant's perspective, as they are mutually accessible and are observably oriented to as such in the data. By contrast, in (4) we observe individual participants' 'private' on-screen activity that is not accessible to the co-participant. Yet, these observations inform us about how the alerting participant responds to the other's talk and actions and offer an additional take on the inextricable interwovenness of the social and (the partially individual) digital dimensions pertaining to the ecology at hand, showing how the rhythm of individual's social talk is tied to their on-screen activity.

1. **The verbal alerts to a screen-based search occur systematically in environments of notable suspensions of talk-in-interaction**. The *let (me) X* construction is produced in environments of prolonged silences, typically of more than 2 or 3 seconds (see e.g. lines 11, 13, 15, 17, 20, 25, 45, and 46). It is placed

not as a first, but rather as a later resort for displaying task-relevant searching. For instance, between lines 14 and 18, SIN first produces vocalizations indicating her continued searching, interspersed with lengthy pauses, and only then resorts to *let me check*. More strikingly, between lines 42 and 46, she first projects more to come by suspending the grammatical trajectory of her turn mid-way (line 42, followed by a 1.9s pause), then uses a search-indicating non-lexical vocalization (line 44, again followed by a pause), and only then produces *let me check* (line 46). *Let (me) X* is used only after talk has been suspended for quite a while: It occurs in environments of notable breaks in the progressivity of talk-in-interaction, rather than in environments of only short suspension of talk. Concomitantly – and prospectively – the *let (me) X* has a scope that reaches over several seconds (see for instance lines 20, 25, 47, as well as the 13 seconds between lines 3 and 9) – a scope that typically reaches over longer silences than the scope of grammatical projections (points of maximal grammatical control) such as shown in lines 42 and 50 (see also Extract 1 above for projection through a hesitation marker). That is: it is treated by the co-participant, who remain silent, as bidding for a notable suspension of talk.

2. **The verbal alerts are followed by non-lexical vocalizations that prolong the indexing of an ongoing screen-based search**. The sequential environment of these alerts also includes repetitive instances during which non-lexical vocalizations are deployed to display participants' continued orientation to the screen (see e.g. lines 14, 16, 21, 48). Most often, the verbal alert indexing of a screen-based search gets thereby prolonged (lines 9–17, 24–29, 46–48), as observable in the long silences following these vocalizations during which co-participants do not take up a turn (e.g., lines 15, 17, 44). In short, various resources (*let [me] X*, vocalizations, grammatical projections) are deployed in sequentially organized ways to signal continued searching – and are attended to as such by co-participants –, thereby also securing extended preparation time, i.e., 'making time' (Tuncer, Lindwall, and Brown 2020; Garfinkel 2002), for the subsequent hinting turns.

3. **The verbal alerts, non-lexical vocalizations and grammatical projection work in concert and in sequentially sensitive ways to structure the very search process and the progression of the task as well as the progressivity of talk-in-interaction**. The verbal alerts occur at pivotal moments where the participants transition from one step to the other in the accomplishment of their private search, and more generally of the task at hand. This is most obvious in lines 9 to 10, where let me (0.5) °just-° (0.5) make a search coincides with the very start of a new round of task engagement. It is also evidenced in the let me (0.7) check in line 18, which marks a shift from SIN's scrolling up

and down the search results page (3# from line 13 to 18) to holding the cursor on a precise result and clicking the corresponding tab (#3, #4, line 18). And, as mentioned above, her `let me che:ck` in line 46 follows her shifting from one Word document to another and coincides with moving the cursor closer to the rules list, which eventually shapes her subsequent hinting behavior. So, the verbal alerts articulate different steps in SIN's own screen-based search in Extract 1. By contrast, the non-lexical vocalizations are deployed as means within these steps to prolong the signaling of her being engaged in searching, while the grammatical projections are put to use only once SIN engages in the hinting activity (lines 21, 42). In other words, there is a complementary distribution – a division of labor – between the three resources.

4. **The delivery of the *let (me) X* verbal alert is closely coordinated with the speaker's screen-based activity.** For instance, the `let me (0.5) °just-° (0.5) make a search` in lines 9 and 10 is synchronized with SIN's own bodily conduct relating to her manipulation of the computer: The delivery of `°just-° (0.5) make a search` coincides exactly with her moving the cursor to the Google icon and clicking on it. Such synchronization is achieved through the speed of delivery of the verbal segment, with its two interspersed 0.5 pauses, which allows the stretch of talk to be minutely articulated to SIN's screen-based activity and to end exactly with her click on the search icon (#1), and the related bodily conduct (moving the mouse across the table and clicking it): It is as if her search on the screen rhythms SIN's speech. The very same synchronization is observable with the `let me (0.7) check (0.2)` in line 18, in which the *let* coincides with SIN's holding the cursor on the result on the IMDB page (#3; beforehand, she was scrolling up and down), and the *check* is delivered simultaneously with SIN's clicking on the IMDB page (4#). Again, such close coordination is made possible by the precise synchronization of the construction's start (*let*) and its course (see the 0.7s pause in line 18) with SIN's movement of the mouse and the corresponding cursor across the screen. Similarly, in line 38, SIN's *let* coincides exactly with her opening of the task-guideline document after she had previously been inspecting another word document. In a nutshell, then, there is close synchronization of SIN's publicly (recognizably) issued verbal alert with her private (not mutually recognizable) inspection of the screen and the related manipulation of mouse and cursor. And this is symptomatic for the inextricable interwovenness of the participants' individual on-screen activities and social coordination with the co-participants for the purpose of joint task accomplishment.

In what follows, we present further representative cases. We start with one instance of PAT producing *let me see* in a similar sequential environment as what

we saw SIN do in Extract 2. Extract 3 is taken from the same movie guessing task as Extract 2.

Extract 3. Movie-guessing task (Tampopo): 28.17 – 28:38

```
1 SIN: 1#they (.) [try to help you.
2 PAT:             [waitress
      1# – SIN holds the IMDB page of the previous movie on until #1 in line 8.
3     (0.9)
4 SIN: yeah
5     (3.5)
6 PAT: 2#u:hmmm#2 (.) 3#okay#3↓ (0.3)
      2# – PAT highlights the current movie, Tampopo on the movies list document.
```



```
      3# – PAT moves the cursor towards the browser and opens it up.
7     ↓4#let me see yeah i definitely need to look up
      4# – PAT conducts web search on the browser and the results appear at #4 in line 9.
8     (2.7)#1
9     °#4hmmmmm°
10    5#(1.9)
      5# – SIN returns to MS Word (task rules), closes it, and returns to IMDB web page at #5 in line 13.
11    it's a comedy
12    (1.1)
13 SIN: #5okay
```

The extract starts with PAT's responding to SIN's hinting, ensuing in the completion of one round of movie guessing and the switch of roles between the participants (i.e., PAT becomes hint-provider and SIN becomes guesser). Just like in SIN's example in Extract 2, the *let (me) X* construction in line 6 (here combined with some further specification as to what PAT will do) is followed by a long silence during which PAT deploys screen-based search (#4). The silence itself shows that SIN treats PAT's *let me X*-initiated turn as a bid for suspension of talk, just like PAT did in regard to SIN's bids in Extract 1. Also, in a similar way as SIN, PAT here uses a non-lexical vocalization (line 9) to prolong the indexing of her being busy with searching on the screen. Following that, PAT starts hinting in line 11 (it's a comedy) which is acknowledged by SIN in line 13. The prolonged silences and non-lexical vocalizations in the public interactional space mark the moments of coordination of screen-based activities (see 4#) on the private space and the participants move the task forward while also carefully avoiding causing breakdowns in talk, which align with all four of our observations on the sequential environment of verbal alerts.

After seeing a variant of verbal alerts (*let me see*), we now turn to the most commonly used format, namely *let me check* from the same movie-guessing task.

Extract 4. Movie-guessing task (Chocolat): 25:22 – 26:15

```
1  SIN: 1#okay
         1# – SIN is on the movies list document.
2        (1.3)
3        mm↑mine i:s (0.7) °hmmm°
4        (1.0)
5        okay#1.
6        2#(1.1)#2
         2# – SIN moves the cursor down towards the task bar.
7        3#let me just↓ (.)#3 check it.
         3# – SIN opens the browser at #3 in line 7.
8        4#(7.1)#4
         4# – SIN types "chocolat" to the address bar, executes the search, and clicks the IMDB result.
9        5#the year of the movie is (0.6) two thousand
         5# – SIN opens up the IMDB page precisely at 5# in line 9 and same frame remains open until #5 in line 12.
10       (1.8)
11       a:nd
12       (2.4)#5
13       6#i think it's (0.3) based on a novel↑
         6# – SIN slightly scrolls down with frame open until #6.
14       (0.4)
15 PAT: ohka:y
16       7#(2.9)#6
         7# – PAT returns to the Google results on the query "chocolat" and opens another page with a query on "vegucated"
         at #7.
17 SIN: 8#it's drama and ro:↑mance#7.
         8# – SIN scrolls down slowly until #8 in line 18.
18       (1.0)#8
19       ↑9#and it's nominated for five oscars.
         9# – SIN moves the cursor on top of "nominated for 5 oscars" on the page and holds it there until #9 in line 22.
```



```
20       (1.7)
21 PAT: huh huhmm
22       (0.6)#9
```

Extract 4 starts with SIN's marking (okay, line 1) her transition into a next guessing round of the task, including a switch of roles. In line 3, she shows her readiness to start telling (mm↑mine i:s), which she suspends while grammatically projecting more to come. Following the ensuing 0.7s pause, SIN produces a minimal vocalization (°hmmm°), followed by a further silence in line 4, further displaying that SIN is busy retrieving task-relevant information. Noteworthy is again the fact that, the multiple breaks in talk (see the silences in lines 2, 3, 4, 6), are not taken as opportunities to access the floor by PAT, nor are they observably treated as problematic. Rather, the co-participant's remaining silent indicates that she is monitoring SIN's vocalizations as displays of her being engaged in task-related search.

As a next step – but non-accessible to PAT –, SIN engages in a screen-based activity (she clicks on the web browser, (3#) that is finely coordinated with the onset of the *let me check* construction (line 7). The physical movement of clicking in line 7 results in the opening of the browser that coincides precisely with the

intra-turn micro-pause in `let me just↓ (.) check it` and SIN's subsequent starting to type (4#) occurs immediately after the end of her uttering that verbal alert and leads to her consulting the IMDB page (#4), based on which she will deliver hints. The prolonged silence of 7.1s in line 8, during which SIN privately executes her IMDB page search, is not sanctioned by PAT, which again testifies to the fact that the *let (me) X* (here: *X=check*) construction works as a pre-positioned account for lengthy breaks in the progressivity of talk-in-interaction – an account that is understood as such by the co-participant. As in all the other cases in the collection, the construction marks SIN's progressing through the task and, as its mutually recognized as such, creates space for retrieving task-related information that is subsequently used in hinting turns.

Accordingly, SIN starts hinting in line 9 and continues doing so in an environment of prolonged silences (lines 10, 12). Given their sequential context, these silences, however, are different from the ones observed above: Here, the silences provide opportunities for PAT to place potential candidate solutions (which PAT does not do), and SIN's adding further hints (lines 17, 19) can be heard as displaying her understanding that PAT does not yet have enough information to propose a candidate. This is further corroborated by PAT's producing an `ohka:y` in line 15, thereby herself using a non-lexical means as in Extract 3 – syllable lengthening (see also her `huh huhmm` in line 21) – to possibly project her own incipient search (see her screen based-activity line 16–17, 7#-#7), which is partially granted by SIN (see the 2.9s of silence, line 15).

The final extract both further illustrates the diverse features that we observe in the sequential environment of *let (me) X* verbal alerts and presents evidence from another task to show that the focal interactional phenomenon of the current study is not necessarily related to specific task types, but is recognizably larger in scope as part of the specialized speech exchange system, task-oriented VMIs. The participants are instructed to engage in an online wayfinding task that encourages intercultural exchange (i.e., food culture). Therefore, the participants move around StreetView to spot restaurants on their screens and prepare a one-day food trip.

Extract 5. Spotting restaurants on street view task: 47:22 – 48–46

```
1  SIN: 1#i found another place too (.)
          1# – SIN holds the cursor still on the Vin and Maree restaurant.
```



```
2        let me: ↓just- describe it (0.4)
3        it's (0.8) vin and maree and it's (0.5) something li:ke,
```

```
4        (0.9)#1
5        2#let me just check (0.2) ops ↑U:OP no:::
         2# – SIN clicks to zoom in and the street view starts moving to another location. She moves around the street
         view until #2 in line 34 and find the Ruc restaurant there.
6        (4.8)
7        wait
8        (1.4)
9   PAT: what happened↑
10       (2.3)
11  SIN: o:p ↑sorry (0.2) just wait for a sec-
12       >↑oh< i am lost again
13       (0.8)
14       perfect
15       (0.7)
16  PAT: $hah$
17       (1.1)
18       it's fine (0.3) just
19       (1.0)
20       let's def- (.) let's (0.3) err find >/cuz/<
21       we found a place where >we can eat des↑sert,<
22       ↑we found a place (.) a restaurant↓ (0.4)
23       so I think we:: (.)
24       <just missed the place> we can eat breakfast↑
25       (1.0)
26  SIN: <yeah>
27       (0.5)
28  PAT: so (0.2) lethh just find a ↓place <°where we can 3#eat↓ breakfast↓°>
         3# – PAT changes the tab to street view page.
29       (1.1)#3
30  SIN: <°okay#3°>
31  PAT: 4#uhmmm
         4# – PAT moves around the street view until the end of the extract.
```



```
32       (6.6)
33       °uh huh huhhh°
34       (7.5)#2
35  SIN: and i found a place called
36       (2.8)
37       ruc#4
38       (2.0)
```

The final extract starts with an announcement, by SIN, of finding a potential task-relevant place of interest (l. 1). The first occurrence of *let (me) X* emerges in line 2; however, it does not necessarily work as a verbal alert due to the specified nature of the screen-based activity, that is to describe which would not require suspension of talk but to verbally deliver a screen-based hint (Balaman 2019). Accordingly, SIN delivers the hint in line 3 and in what follows, she deploys the most commonly used verbal alert in the dataset (i.e. *let me (just) check*). On the private space, she zooms in to access further details related to focal task-relevant place (2#) yet zooming in causes disruption on the StreetView and leads her to move around the page until she finds another place (#2 in line 34). SIN brings this trouble to the public space of interaction in line 5 (ops ↑U:OP no:::) and 6 (wait). Her problematization and divergence from the recurrently observed sequential environment of verbal alerts are attended to by PAT (what happened↑) in line 9, which is

responded to by SIN initially with request for wait time (just wait for a sec-) and an account giving (i am lost again). PAT accepts this account (it's fine) and proposes an alternative course of screen-based activity (let's def- (.) let's (ø.3) err find). Her deployment of the *let us def-/find* structure here marks the beginning of an extended proposal that is inclusive in structure (*let's*). PAT primarily elaborates on the nature of the proposed incipient activity and deploys another inclusive construction, this time functioning as a verbal alert (lethh just find a ↓place) which are both acknowledged by SIN in line 26 (<yeah>) and 30 (<°okay°>). Also note that PAT's completion of the verbal alerting practice is marked with her turn-final production that is slow paced and low voiced and includes gradually falling intonation. In what follows, we see repetitive occurrences of prolonged silences and non-lexical vocalizations on the public space and rhythming of screen-based activities with talk in the private space. All in all, PAT's inclusive and collaborative verbal alerting practice is recognized by SIN and lead them to perform screen-based activities that observably moves the task forward (and i found a place called).

5.   Discussion and conclusions

In this paper we have documented how participants successfully navigate the complex social ecology of task-oriented VMI as a specialized speech exchange system by closely coordinating their own (mutually non-accessible) on-screen search with their 'public' participation in social interaction with the remote co-participant. We have shown that this coordination centrally involves the searching participant's use of diverse resources to make the search publicly inspectable *as* a search to her co-participant, and hence to deal with the asymmetry of access pertaining to their mutual screen-based activities: Overt verbal alerts in the form *let (me) X,* combined with non-lexical vocalizations and the suspension of talk in mid-syntactic trajectory are used as means to account for halts in talk-in-interaction as being due to the speaker's on-screen search. These displays are not distributed randomly over the search episodes, but are closely synchronized with the speaker's physical (moving of mouse; see Knight, Dooly, and Barbera 2018), visual (interacting with the content on the screen; see Balaman and Sert 2017a) and cognitive (e.g., retrieving task-related information from the screen, see Arminen 2005; Näslund 2016) activities involved in task accomplishment. And this is symptomatic of the interwovenness of social-interaction and individual screen activity in the social-digital ecology at hand. The social significance of the search-displays lies hence both in accounting for breaks in talk-in-interaction by

displaying that the speaker is engaged in a concurrent line of action (Näslund 2016), and in centrally showcasing every line of action as being a constitutive part of the social encounter at hand, i.e., as being instrumental for the joint task to move forward: The search for clues is a pre-condition for providing hints to the guessing other. In this sense, the way the screen-based activity is made mutually manifest and hence publicly accountable is reflexively related to the task at hand (see Garfinkel 1967, on accountability and reflexivity). It is part of the myriad of new agencies and accountabilities involved in humans interacting with and through machines (cf. Suchman 2007).

Regarding specifically the verbal alerts, as mentioned earlier, while pervasive in the data, their precise lexico-syntactic constituency slightly varies among participants. SIN's case is particularly interesting, as her verbal alerting practices consist routinely of *let me check*. As Hoey (2020) has recently documented, *let me X* shows a distinct interactional working: Through its use, the speaker, rather than doing an ordinary request, builds on an assumption of permission by the recipient; thereby, she carefully ensures attainment of self-authorization in talk-in-interaction through apparently soliciting the recipient's consent but at the same time minimizing the potential emergence of resistance in the next sequential slot. This is exactly what we see at work in our participants' verbal alerting practices. While SIN does so repeatedly drawing on the same format (*let me check*) without specifying the nature of the incipient activity, PAT uses diverse resources (*let me see, let's find*) and specifies (even elaborates on) the incipient activity. This aligns with what we found elsewhere: SIN's verbal alerts to incipient screen-based activities included diverse resources when she first participated to task-oriented VMIs, yet over four years her practices routinized into the passe-partout formula *let me check* (Pekarek Doehler and Balaman 2021). It might therefore be that a similar longitudinal examination of PAT's practices would indicate a similar routinization over time. In any case, we see the mutual recognizability of both participants' verbal alerts based on the *let (me) X* (e.g. *search, check, find* etc.) structure. Similarly, Näslund (2016) has found that call-takers put the callers on hold while retrieving information from the screen after requesting for wait time and by doing so, they collaboratively project incipient silences. In our study, the screen-oriented silences are not jointly decided on but unilaterally signaled through the sequential positioning of the verbal alerts.

Participants' deployment of the alerts to incipient or ongoing search as well as their responses to these are reflexively tied to the situation at hand; they index participants' understanding of the goals of their joint activities, of the potential multitasking these involve, and of the pre-conditions individuals need to establish in real time in order to meet the requirements of task-accomplishment so as to move their encounter ahead. In this sense, the hinting participant's public display of her

incipient or ongoing screen-based activities reflects the very nature of the speech exchange system under scrutiny, including its specific normativity regarding the disruption of talk-in-interaction: The progressivity of task-accomplishment is observably prioritized over the progressivity of talk. This is part of the complexity of VMI, and in particular of task-oriented VMI, in that talk represents only one layer in this multi-faceted ecology, in which talk is temporally coordinated in subtle ways with participants' inspection of text or images on the screen and related movement of the cursor (Heath and Luff 1993, 2000; Jenks and Brandt 2013; Oittinen and Piirainen-Marsh 2015; Luff et al. 2013; Luff et al. 2016; see Mlynar, Gonzalez-Martinez, and Lalanne 2018 for an overview).

Our findings further add to the current understandings in VMI research by showing that participants draw on multiple resources, including syntactic ones (Sacks and Schegloff 1979; Wilkinson 2009; Pekarek Doehler 2011; Pekarek Doehler and Horlacher 2013), for achieving mutual coordination and avoiding progressivity related trouble. They highlight the fact that in VMI visibility of mutual actions cannot be taken for granted despite the existence of the Skype frame throughout, yet silences do not seem to be attended to as signs of interactional trouble (Brandt 2011; Brandt and Jenks 2013; Balaman and Sert 2017b; Sert and Balaman 2018; Rintel 2013). This is so because participants actively work to make such silences publicly recognizable as displays of their ongoing screen-based activities that otherwise are not inspectable as such to co-participants. To these ends, the technicalities and related constraints of VMI settings create opportunities for the emergence of context-specific resources, and shape into being affordances that are specifically functional for managing digitally mediated interactions (Heath and Luff 1993; Arminen, Licoppe, and Spagnolli 2016).

Funding

References

Arminen, Ilkka. 2005. *Institutional Interaction: Studies of Talk at Work.* Routledge.
Arminen, Ilkka, Inka Koskela, and Hannele Palukka. 2014. "Multimodal Production of Second Pair Parts in Air Traffic Control Training." *Journal of Pragmatics* 65: 46–62.
    https://doi.org/10.1016/j.pragma.2014.01.004
Arminen, Ilkka, Christian Licoppe, and Anna Spagnolli. 2016. "Respecifying Mediated Interaction." *Research on Language and Social Interaction* 49 (4): 290–309.
    https://doi.org/10.1080/08351813.2016.1234614

Balaman, Ufuk. 2018. "Task-Induced Development of Hinting Behaviors in Online Task-Oriented L2 Interaction." *Language Learning*, 21. 10125/44640

Balaman, Ufuk, and Olcay Sert. 2017a. "The Coordination of Online L2 Interaction and Orientations to Task Interface for Epistemic Progression." *Journal of Pragmatics* 115 (July): 115–29. https://doi.org/10.1016/j.pragma.2017.01.015

Balaman, Ufuk, and Olcay Sert. 2017b. "Development of L2 Interactional Resources for Online Collaborative Task Accomplishment." *Computer Assisted Language Learning* 30 (7): 601–30. https://doi.org/10.1080/09588221.2017.1334667

Balaman, Ufuk. 2019. "Sequential Organization of Hinting in Online Task-Oriented L2 Interaction." *Text & Talk* 39 (4): 511–34. https://doi.org/10.1515/text-2019-2038

Bilmes, Jack. 1994. "Constituting Silence: Life in the World of Total Meaning." *Semiotica* 98 (1–2): 73–88. https://doi.org/10.1515/semi.1994.98.1-2.73

Brandt, Adam. 2011. "The Maintenance of Mutual Understanding in Online Second Language Talk." PhD Thesis, Newcastle University.

Brandt, Adam, and Christopher Jenks. 2013. "Computer-Mediated Spoken Interaction: Aspects of Trouble in Multi-Party Chat Rooms." *Language@Internet* 10: 1–21.

Due, Brian L. 2015. "The Social Construction of a Glasshole: Google Glass and Multiactivity in Social Interaction." *PsychNology Journal* 13 (2): 149–178.

Gardner, Rod, and Mike Levy. 2010. "The Coordination of Talk and Action in the Collaborative Construction of a Multimodal Text." *Journal of Pragmatics* 42 (8): 2189–2203. https://doi.org/10.1016/j.pragma.2010.01.006

Garfinkel, Harold. 1967. *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice Hal.

Garfinkel, Harold. 2002. *Ethnomethodology's Program: Working out Durkheim's Aphorism*. Rowman & Littlefield Publishers.

Goodwin, Charles. 2013. "The Co-Operative, Transformative Organization of Human Action and Knowledge." *Journal of Pragmatics* 46 (1): 8–23. https://doi.org/10.1016/j.pragma.2012.09.003

Goodwin, Charles. 2018. *Co-Operative Action*. Cambridge University Press.

Goodwin, Marjorie Harness, and Charles Goodwin. 1986. "Gesture and Coparticipation in the Activity of Searching for a Word." *Semiotica*, 62(1–2): 51–76.

Haddington, Pentti, Tiina Keisanen, Lorenza Mondada, and Maurice Nevile, eds. 2014. *Multiactivity in Social Interaction: Beyond Multitasking*. John Benjamins Publishing Company. https://doi.org/10.1075/z.187

Heath, Christian, and Paul Luff. 1993. "Disembodied Conduct: Interactional Asymmetries in Video-Mediated Communication." In *Technology in Working Order: Studies of Work, Interaction, and Technology*, edited by Graham Button, 35–54. Routledge.

Heath, Christian, and Paul Luff. 2000. *Technology in Action*. Cambridge University Press. https://doi.org/10.1017/CBO9780511489839

Heritage, John. 2007. "Intersubjectivity and Progressivity in Person (and Place) Reference." In *Person Reference in Interaction: Linguistic, Cultural and Social Perspectives*, edited by Nick J. Enfield and Tanya Stivers, 255:255–80. Cambridge University Press.

Hjulstad, Johan. 2016. "Practices of Organizing Built Space in Videoconference-Mediated Interactions." *Research on Language and Social Interaction* 49 (4): 325–41. https://doi.org/10.1080/08351813.2016.1199087

Hoey, Elliott M. 2020. "Self-Authorizing Action: On Let Me X in English Social Interaction." *Language in Society*, 1–24. https://doi.org/10.1017/S0047404520000779

Jefferson, G. (2004). Glossary of Transcript Symbols with an Introduction. In. *Conversation Analysis, Studies from First Generation*, edited by Gene Lerner, 13–34. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.125.02jef

Jenks, Christopher Joseph, and Adam Brandt. 2013. "Managing Mutual Orientation in the Absence of Physical Copresence: Multiparty Voice-Based Chat Room Interaction." *Discourse Processes* 50 (4): 227–48. https://doi.org/10.1080/0163853X.2013.777561

Knight, Janine, Melinda Dooly, and Elena Barberà. 2018. "Multimodal Meaning Making: Navigational Acts in Online Speaking Tasks." *System* 78: 65–78. https://doi.org/10.1016/j.system.2018.07.007

Licoppe, Christian, and Julien Morel. 2018. "Visuality, Text and Talk, and the Systematic Organization of Interaction in Periscope Live Video Streams." *Discourse Studies* 20 (5): 637–65. https://doi.org/10.1177/1461445618760606

Luff, Paul, Christian Heath, Hideaki Kuzuoka, Jon Hindmarsh, Keiichi Yamazaki, and Shinya Oyama. 2003. "Fractured Ecologies: Creating Environments for Collaboration." *Human–Computer Interaction* 18 (1–2): 51–84. https://doi.org/10.1207/S15327051HCI1812_3

Luff, Paul, Christian Heath, Naomi Yamashita, Hideaki Kuzuoka, and Marina Jirotka. 2016. "Embedded Reference: Translocating Gestures in Video-Mediated Interaction." *Research on Language and Social Interaction* 49 (4): 342–61. https://doi.org/10.1080/08351813.2016.1199088

Mlynář, Jakub, Esther González-Martínez, and Denis Lalanne. 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, Lorenza. 2014. "The Temporal Orders of Multiactivity." In *Multiactivity in Social Interaction: Beyond Multitasking*, edited by Pentti Haddington, Tiina Keisanen, Lorenza Mondada, and Maurice Nevile, 33–75. John Benjamins Publishing. https://doi.org/10.1075/z.187.02mon

Näslund, Shirley. 2016. "Tacit Tango: The Social Framework of Screen-Focused Silence in Institutional Telephone Calls." *Journal of Pragmatics* 91 (January): 60–79. https://doi.org/10.1016/j.pragma.2015.10.008

O'Dowd, Robert, and Tim Lewis. 2016. *Online Intercultural Exchange: Policy, Pedagogy, Practice*. Routledge. https://doi.org/10.4324/9781315678931

Oittinen, Tuire. 2018. "Multimodal Accomplishment of Alignment and Affiliation in the Local Space of Distant Meetings." *Culture and Organization* 24 (1): 31–53. https://doi.org/10.1080/14759551.2017.1386189

Oittinen, Tuire, and Arja Piirainen-Marsh. 2015. "Openings in Technology-Mediated Business Meetings." *Journal of Pragmatics* 85: 47–66. https://doi.org/10.1016/j.pragma.2015.06.001

Olbertz-Siitonen, Margarethe. 2015. "Transmission Delay in Technology-Mediated Interaction at Work." *PsychNology Journal* 13 (2–3): 203–34.

Pekarek Doehler, Simona. 2011. "Emergent Grammar for All Practical Purposes: The On-line Formating of Dislocated Constructions in French Conversation." In *Constructions: Emerging and Emergent*, edited by Peter Auer and Stefan Pfänder, 46–88. Mouton de Gruyter. https://doi.org/10.1515/9783110229080.45

Pekarek Doehler, Simona and Anne-Sylvie Horlacher. 2013. The Patching together of Pivot-patterns in Talk-in-interaction: On 'Double Dislocations' in French. *Journal of Pragmatics* 53: 92–108. https://doi.org/10.1016/j.pragma.2013.04.002

Pekarek Doehler, Simona and Ufuk Balaman. 2021. "The Routinization of Grammar as a Social Action Format: A Longitudinal Study of Video-mediated Interactions". *Research on Language and Social Interaction*. https://doi.org/10.1080/08351813.2021.1899710

Rintel, E. Sean. 2010. "Conversational Management of Network Trouble Perturbations in Personal Videoconferencing." In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, 304–11. https://doi.org/10.1145/1952222.1952288

Rintel, Sean. 2013. "Video Calling in Long-Distance Relationships: The Opportunistic Use of Audio/Video Distortions as a Relational Resource." *The Electronic Journal of Communication/La Revue Electronic de Communication (EJC/REC)* 23.

Sacks, Harvey, and Emanuel A. Schegloff. 1979. "Two Preferences in the Organization Ofreference to Persons in Conversation and Their Interaction." In *Everyday Language: Studies in Ethnomethodology*, edited by George Pathas, 15–21. New York: Irvington Publishers.

Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. "A Simplest Systematics for the Organization of Turn-Taking for Conversation." *Language* 50 (4): 696–735. https://doi.org/10.2307/412243

Sert, Olcay, and Ufuk Balaman. 2018. "Orientations to Negotiated Language and Task Rules in Online L2 Interaction." *ReCALL* 30 (3): 355–74. https://doi.org/10.1017/S0958344017000325

Schegloff, Emanuel A. 1979. "The Relevance of Repair to Syntax-for-Conversation." In *Discourse and Syntax*, edited by Talmy Givon, 261–86. New York: Academic Press. https://doi.org/10.1163/9789004368897_012

Schegloff, Emanuel A. 1996. "Turn Organization: One Intersection of Grammar and Interaction." In *Interaction and Grammar*, edited by Elinor Ochs, Emanuel A. Schegloff, and Sandra A. Thompson, 52–133. Cambridge University Press. https://doi.org/10.1017/CBO9780511620874.002

Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis I*. Cambridge University Press. https://doi.org/10.1017/CBO9780511791208

Stivers, Tanya, and Jeffrey D. Robinson. 2006. "A Preference for Progressivity in Interaction." *Language in Society* 35 (3): 367–92. https://doi.org/10.1017/S0047404506060179

Suchman, Lucy A. 2007. *Human-machine Reconfigurations. Plans and Situated Actions*. Cambridge University Press.

Tuncer, Sylvaine, Oskar Lindwall, and Barry Brown. 2020. "Making time: Pausing to coordinate video instructions and practical tasks". *Symbolic Interaction*. https://doi.org/10.1002/symb.516

Whalen, Jack. 1995. "A Technology of Order Production: Computer-Aided Dispatch in Public Safety Communications." In *Situated Order: Studies in the Social Organization of Talk and Embodied Activities*, edited by Paul ten Have and George Psathas, 187–230. Washington DC: University Press of America.

Wilkinson, Ray. 2009. "Projecting a Reference in Aphasic Talk and Normal Talk." *Discourse Processes*, 46(2–3), 206–225. https://doi.org/10.1080/01638530902728892

Appendix. Transcription conventions

1#	Onset point of the screen-based activity surrounding the talk that is marked along with the lines of the transcript
#1	Offset point of the screen-based activity surrounding the talk that is marked along with the lines of the transcript
1#...	Continuation of the screen-based activity (used only within the screen-based activity illustrations)
Illustrations	Current screen of the participants who perform the screen-based activities
Circles	Points on the screen where the participants either click or hold the cursor still
Arrow	Direction of the cursor movements within the screen-based activity illustrations
Lines 2–5	Duration of screen-based activity represented across lines in order to indicate the scope of each description
Descriptions	Unanalytical descriptions of the illustrated screen-based activities

These notations are used in addition to Jeffersonian (2004) transcription conventions below.

[	start of overlap
]	end of overlap
=	latching (no pause, no overlap)
&	turn continuation after overlap
(0.7)	measured pause in seconds and tenths of seconds
wo-	truncated word
wo:rd	syllable lengthening
?	rising final intonation
¿	mid-rise intonation
.	falling final intonation
,	continuing intonation
<u>word</u>	emphasis
°word°	softer than surrounding speech
WORD	louder than surrounding speech
↑word	marked high rise in pitch (refers to the next syllable)
word^word	phonetic 'liaison'
.h	in-breath
((laughter))	transcriber's comment

Publication history