

## **Audibility and Intelligibility of Talk in Passing-By Interactions: A Methodological Puzzle**

*Esther González-Martínez & Jakub Mlynář*

**Key words:**

conversation  
analysis;  
multimodality;  
video-based field  
study; hospital  
staff; corridor talk;  
passing-by  
interactions; multi-  
location sound  
recordings;  
audibility;  
intelligibility

**Abstract:** A central assumption of conversation and multimodal analysts is that the participants have sensorial access to the acoustic features, most notably talk, of the studied interactions, which are available thanks to the recordings. If speakers stand in close proximity, a microphone is placed between or on one of them, and the analyst considers that the resulting recording captures the sound that is relevant both for the participants and for research purposes. The study of mobile and spatially distributed interactions comes with a challenge: If speakers are not spatially close but apart from each other, what is the relevant sound for the researcher to collect and examine? In this article, we show that prosodic, compositional and sequential features of the same segment of talk sound differently depending on the exact location of the hearer, or recording device, which is relevant for the analyst as well as for the participants. Our study was based on a corpus of hospital staff corridor interactions captured with a set-up composed of four video-cameras and eight wireless microphones operating simultaneously.

### **Table of Contents**

- [1. Introduction](#)
- [2. Audibility and Intelligibility of Talk-in-Interaction](#)
- [3. A Video-Based Field Study in the Corridors of a Hospital Outpatient Clinic](#)
- [4. Editing and Transcribing Sound Recordings of Passing-By Interactions](#)
- [5. Acoustic Features of Three Passing-By Interactions](#)
- [6. Discussion and Conclusion](#)
- [Acknowledgments](#)
- [Appendix: Transcription Conventions](#)
- [References](#)
- [Authors](#)
- [Citation](#)

## **1. Introduction**

Research on language and social interaction (GLENN, LeBARON & MANDELBAUM, 2003) is currently experiencing a turn towards the senses—with groundbreaking studies on the situated organization of seeing, smelling, tasting and touching—, which are all brought together, in their interactional concurrence and synchronicity, under the notion of the "multi-sensorial" (GIBSON & VOM LEHN, 2021; GOICO, GAN, KATILA & GOODWIN, 2021; MONDADA, 2021). The move also includes scrutiny of "hearing-in-interaction" (YAGI, 2021, p.21) going beyond the traditional focus on talk to encompass nonlexical objects such as clicks and whistles (KEEVALLIK & OGDEN, 2020); mechanically generated sounds like car horns and phone notifications (LAURIER, MUÑOZ, MILLER & BROWN, 2020; LICOPPE, 2010); music making (YAGI, 2021) and the soundscape of interactions like surrounding sounds and noises (HEINEMANN &

RAUNIOMAA, 2016; LAURIER, WHYTE & BUCKNER, 2001; MERLINO, MONDADA & SÖDERSTRÖM, 2022; RAUNIOMAA & HEINEMANN, 2014). In the resulting studies, it has been described how participants produce, orient to and handle nonverbal sounds and engage jointly in hearing them as well as in dedicated listening and sharing activities. [1]

In this article, we propose a return to and exploration of the sense of hearing focused on talk, which is central to the analytical approaches of conversation (SACKS, SCHEGLOFF & JEFFERSON, 1974) and multimodal analysis (GOODWIN, 2000; MONDADA, 2007). Indeed, a basic research assumption is that participants have aural access to the talk produced during the studied interactions, which is also available to the analyst thanks to their video or audio recordings. Two recent developments in the field nevertheless encourage greater nuance and further analytical and methodological investigation. Until the turn of the twenty-first century, most research concentrated on telephone calls and face-to-face interaction in static formations and confined spaces (MONDADA, 2009). Since then, studies have expanded to include interactions among participants scattered throughout spatially distributed domestic (LA VALLE, 2014) and professional (GONZÁLEZ-MARTÍNEZ, BANGERTER & LÊ VAN, 2017; HOEY, 2023) spaces, in public open spaces (DE STEFANI & MONDADA, 2014; HOCHULI, 2019; RELIEU, 2023) and in nature (SMITH, 2021), moving around, together or alone. Since participants are not necessarily spatially close, identical aural access to ongoing talk can no longer be taken for granted. Concomitantly, researchers are using more complex recording set-ups. Although in the past a single recording device placed at one end of the telephone line, between two speakers or on one of them would suffice, nowadays scientists use multiple static and portable video cameras and microphones and devices capturing 360° images and spatial audio, and also juggle with multiple synchronized data streams (HADDINGTON et al., 2023; McILVENNY & DAVIDSEN, 2017). [2]

Here, we present a study on corridor interactions in a hospital outpatient clinic based on a corpus of audio and video data collected with four action cameras and eight wireless microphones functioning simultaneously to capture the activities of the staff walking in and out of the rooms and through the hallways (GONZÁLEZ-MARTÍNEZ et al., 2017). In the next sections, we first discuss how research on language and social interaction deals with issues related to the audibility and intelligibility of talk for the participants and the analyst (Section 2). We next present the studied interactions (Section 3) and the methodological puzzle that they give rise to: what is the relevant sound to be considered when examining talk produced by participants who are not spatially close? (Section 4). We then analyze six fragments, drawn from three different videoclips in our corpus, showing how sound recordings of a single talk fragment may differ depending on the location of the microphone (Section 5). In the discussion, we consider the methodological and analytical implications of the aural equivalent of the problem of multiple views of a "single" scene (LUFF & HEATH, 2012) and underline the significance of investigating mobile interactions to reach an understanding of them (Section 6). [3]

## 2. Audibility and Intelligibility of Talk-in-Interaction

When examining the recordings of interactions under study, practitioners of multimodal and conversation analysis sometimes encounter difficulties with perceiving sonic phenomena (whether talk, nonverbal tokens or other sounds) and identifying their source and what they might mean (see JEFFERSON, 2010 for a classic account). Sometimes, it is an audibility issue: The analysts may have the impression that they can hear a sound in one fragment of the recording but the next moment they may no longer be sure; participants in a data session may not agree whether there is something to be heard at some specific point but after some time passes, cannot fail to hear sounds they could not initially discern (PSATHAS & ANDERSON, 1990). The ability to perceive recorded sounds depends on their acoustic features and the acuteness of the senses as well as contingent factors like the degree of attentiveness, degree of familiarity, surrounding noises and external guidance. Recurrent examination of recordings trains participants to discern and differentiate subtle acoustic features that would otherwise go completely unnoticed. The praxeology of perception (COULTER & PARSONS, 1990; SHARROCK & COULTER, 1998) also serves as a reminder that the very possibility of discerning a sensorial "stimulus" depends on knowing that it can indeed take place and having a concept for it, even if only a concept like unidentified sound object. Audibility is thus intrinsically related to intelligibility: The possibility of identifying the source of the sound and, in the case of a symbolic acoustic object, its meaning. Is this a fragment of talk? Who is speaking? What is the person saying? What does she mean by that and what is she doing? At this level, the analysts rely on resources gathered through direct and indirect contact with the participants, continued examination of the recordings, their own professional training, discussions with other practitioners and their general cultural knowledge (LEE, 1987; PSATHAS, 1990). [4]

Addressing audibility and intelligibility issues is central to conversation and multimodal analysis, as well as the object of specific methodological and analytical practices. At the first level, JEFFERSON (2004), in her transcription conventions, instructed transcribers to use empty parentheses to mark the fact that speech is likely being heard even if the transcribers are unable to reproduce what they hear or determine what may have been said. Parentheses are also used when proposing a candidate transcription of an unclear sound or alternative hearings of a single fragment. At the analytical level, the study of repair is related to "problems in speaking, hearing, or understanding, the talk" (SCHEGLOFF, 2007, p.100) that the participants address in observable, recurrent ways in order for the interaction to progress and intersubjectivity to be maintained. The examined practices are evidence of the relevance for the participants themselves of issues of audibility and intelligibility, and of their distinction and subtle aspects. For instance, a hearing-impaired adolescent may indicate that the trouble he is experiencing is not related to his medical condition but rather caused by the complexity of the teacher's instruction (GROEBER & PEKAREK DOEHLER, 2012). DREW (1997) provided another example, noting that some repair practices address trouble that is not related to understanding, word by word, a

fragment of talk but rather grasping its connection to what has previously been said or done. [5]

More fundamentally, the question is which data and related media and knowledge should be considered as analytically consequential? For instance, is it the sound as it was produced and perceived by the participants during the event, the way they observably orient to it, the sound recording of the event, the transcripts, or other graphic representations? (ASHMORE & REED, 2000; WEEKS, 2002) The transcript has traditionally been considered a constructed object grounded in the analysts' transcription practices, their approach to the data and their analytic interest (OCHS, 1979; PSATHAS & ANDERSON, 1990). Before that, the analyst would decide at the time of data collection how much importance to give to the phenomenal background of what was being heard (BJELIĆ, 2019). If studying summoning phones, for instance, the analyst can take into consideration the singularity of the telephone ring—how it is being heard and understood by that particular person, in that particular place, at that particular moment—or instead concentrate on the recordable features of the call: The telephone tone, the voices of the interlocutors and other sounds referred to or heard in the recording playing a role in the sequential organization of the conversation. ASHMORE and REED (2000) argued that even the first hearing of the recording is not an immediate grasping experience but the result of multiple mediations providing for something to be heard in the first place. Instead of "a direct and evidential record of a past event" (ASHMORE, MacMILLAN & BROWN, 2004, p.349), the recordings are the result of multiple practical decisions about what to record, when to start and stop, which equipment to use, how to handle and place it, how to edit the files, and how to pick what is to be listened to (PSATHAS & ANDERSON, 1990; see also MONDADA, 2006). MODAFF and MODAFF (2000) cautioned analysts to be attentive to how the recording device, its settings and location influence how the talk event is documented, and can be heard and understood. They showed, for instance, that "recording both ends of a telephone conversation provides a more complete record of the interactional details of the conversation than does recording on just one end" (p.102). Nevertheless, LUFF and HEATH (2012) noted that when examining multiple views of a single event, it is sometimes not clear whether or how what the analyst can see has also been seen by the different participants. It can also be unclear whether or how a particular conduct has been produced by the participant in respect to the others: Was it oriented towards another participant and if so, to whom? Informed by the turn towards multisensoriality, MERLINO (2021) reflected on the need to pose these questions when examining not only the visual but also the auditory details of interaction, and their interrelation. This is especially relevant when examining interactions among participants scattered throughout a space and captured with complex recording set-ups. [6]

### 3. A Video-Based Field Study in the Corridors of a Hospital Outpatient Clinic

The present study was grounded in data collected for a research project on staff mobility practices in a hospital outpatient clinic in the French-speaking part of Switzerland (GONZÁLEZ-MARTÍNEZ et al., 2017).<sup>1</sup> On an average working day, the clinic's team was composed of fourteen people: Senior and junior doctors, a head nurse, nurses, aides, a nursing trainee and the secretary-receptionist. The clinic provided both scheduled and walk-in healthcare services. Patients reached the clinic through the main entrance next to the reception desk (see Figure 1). Inside, the main centers of activity were the day hospital room, used mainly for administering treatment of chronic patients, and the urgent-care room, reserved for non-life-threatening emergencies. There were also several other consultation, examination and treatment rooms as well as general use and clinical support spaces (a break room, restrooms, stock and sluice rooms, etc.). The rooms were served by two parallel double-loaded corridors connected in the middle and at the ends by short hallways. During the day, the staff kept most of the doors open and relied on curtains surrounding the patients' beds which they occasionally drew closed for reasons of privacy. [7]

Over four months, the research team conducted fieldwork in the clinic, including observations, shadowing nurses, formal and informal interviews and collecting documentation. At the end of the fieldwork period, the team recorded activity taking place in the clinic's corridors over seven consecutive days, twelve hours a day, from open to close. With the help of a technician, the researchers had put together a recording set-up composed of four Sony HDR video cameras suspended from the ceiling with their corresponding internal microphones, eight omnidirectional Sennheiser ME 2 microphones suspended from mural light fixtures with their corresponding wireless Sennheiser ew 500-G3 transmitters,<sup>2</sup> and a reception/mixing/editing station. This recording set-up allowed staff members to carry on their activity as usual without being followed by a cameraman or having themselves to carry around recording devices that they would have to turn on and off when entering and exiting rooms. The cameras and microphones were all operating simultaneously and non-stop, except for a half-hour break to change the cameras' batteries and memory cards. The cameras recorded activity in Corridors A and B, as people walked through them and

---

1 The goal of the initial project, conducted from September 2011 to February 2015, was to study the mobile and contingent work interactions of the clinic's staff in the corridors (Swiss National Science Foundation [SNSF], grant no. 134875). It was a video-based field study that entailed the data collection methods specified in this article and relied heavily on conversation and multimodal analysis. The current study was part of a new research project conducted from September 2019 to August 2024, in which we used the data collected for the initial project, relying on the same analytical approaches, and focused on specific mobility practices and mobile activities, such as recruiting moves (SNSF, grant no. 185152).

2 The microphones captured the sound and transferred it to the transmitters, to which they were wire connected, and the transmitters then wirelessly transferred the resulting signal to the reception station. We selected these microphones and transmitters because they were user-friendly and professional, small in size and light in weight, offering high quality sound and reliability, and were compatible with the other parts of the recording set-up.

entered and exited the adjacent rooms, as well as activities in the entrance area of the day hospital and the urgent-care room (see Figure 1).<sup>3</sup>



Figure 1: Clinic premises and recording set-up. The triangles represent the video cameras, the dots the wireless microphones, and the striped rectangle the reception/mixing/editing station. The area covered by the video cameras is represented in gray. Corridor A is 27.40 meters long, Corridor B (the section between Corridors A and C) 4.16 meters long, and Corridor C 31.50 meters long. [8]

The research team thus collected 331 hours of video. The video files were synchronized with each other and the audio recordings in Final Cut Pro, and coded with StudioCode, as the researchers looked for instances ("occupations") when there was a person in the cameras' field of vision.<sup>4</sup> In compliance with the research protocol, the research team only considered the images of themselves, the clinic's staff and other hospital personnel who had agreed to the recordings in advance: Thirty-seven individuals in total. The research protocol provided for the reproduction, for the purpose of scientific publications, of still images of the participants taken from the video recordings. For the present study, we considered a collection of fifty-four passing-by interactions involving talk for which multicam clips were first edited and subsequently transcribed following JEFFERSON's conventions (2004) for conversation analysis (SACKS et al., 1974). [9]

3 Thanks to the present study, we realized the importance of marking the locations of the recording devices as precisely as possible on the map.

4 We chose Final Cut Pro because it was a user-friendly professional video-editing application that offered the tools needed for our project and was compatible with the other items of the recording set-up. StudioCode was another user-friendly professional application that also offered the required compatibility and provided the options of viewing the recordings, marking and categorizing events in them, and exporting the resulting coding data files for subsequent statistical analysis.

#### **4. Editing and Transcribing Sound Recordings of Passing-By Interactions**

The results of the initial research project showed that staff corridor occupations were very frequent, extremely brief and also mobile, since only a few of them involved stops (GONZÁLEZ-MARTÍNEZ, BANGERTER, LÊ VAN & NAVARRO, 2016). Talk was common but the majority of the time, even when talking, staff kept moving. Team members thus produced numerous "on the fly" work interactions which included inquiries, informing, requests, checks, and offers of help, as some of the most frequent verbal actions. The interactions were brief since the staff members were not necessarily side by side or face to face but moving around, often following a distinct trajectory and/or in different parts of the clinic. Talk sometimes began and continued while the interactants were far apart. Passing-by is the spatial and interactional configuration (GONZÁLEZ-MARTÍNEZ et al., 2017) that was of special interest for the present study. When producing a passing-by, participants formed close parallel trajectories, each one going in the opposite direction from the other. They talked before, while and/or after passing by each other, without stopping. [10]

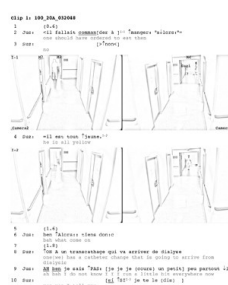
Yet, physical closeness is a fundamental feature of co-present interaction as it is classically defined. Thus, GOFFMAN in his foundational writings (e.g., 1959, 1963, 1971) conveyed the idea that co-present interaction involves individuals in immediate physical presence, who are close together, facially oriented to one another, forming eye-to-eye huddles, clusters or knots. For KENDON (1976), the transactional segment was the space extending in front of persons and which they could reach with their hands. A formation was a sustained spatial pattern that people followed when they grouped: "An F-formation arises whenever two or more people sustain a spatial and orientational relationship in which the space between them is one to which they have equal, direct and exclusive access" (p.291). Individuals set up an F-formation when they orient their bodies in a way that creates and maintains an intersection of their transactional segments, which implies close physical proximity. The spatial arrangements assumed by F-formations may differ, but all correspond to knots or huddles, in a common spatially restricted site. For MONDADA (2009), an interactional space is established as participants engage in mutual gaze and converging trajectories, and transition from walking to standing in a face-to-face position. [11]

The fact that passing-by participants talk while physically apart, following trajectories going in opposite directions, raises an analytical question as well as a methodological one. On the analytical side, the question is "How do speakers talk when not in close proximity but six, eight or twelve meters apart?" On the methodological side, if speakers are not spatially close, what is the relevant sound recording for the analyst to select when editing and transcribing the excerpt? Is it the sound produced by the speakers' talk and the other sounds in their immediate surroundings? Or is it the sounds heard by the recipient who is several meters away? How to handle the fact that the participants who are producing and hearing sounds are on the move? The routine analytical procedures for recording, editing and transcribing in conversation and multimodal

analysis are derived from the basic presumption of a common acoustic environment. For instance, researchers may place a microphone in between two interlocutors or ask one to wear a microphone, and consider the resulting sound to be the relevant sound for both participants. When A and B are talking in a close physical formation, the analyst supposes that the interlocutors can hear the same thing at the same time. If they do not, and the problem in hearing poses interactional troubles, it is repaired in observable manners. But what is to be considered the sound that is relevant for both participants when they talk while physically apart and also moving? How to take into account and visually represent the participants' distinct and/or intertwined acoustic environments? In the remainder of the article, we explore the pertinence of such a methodological puzzle by discussing six fragments drawn from three different clips of passing-by interactions in the central corridor of the clinic. [12]

## 5. Acoustic Features of Three Passing-By Interactions

Clip 1 corresponds to an interaction between Suzi (Suz in the transcript), a nurse, and Justa (Jus), an aide, who pass by each other in Corridor B.<sup>5</sup> Suzi comes out of the urgent-care room (UCR) and walks to the day hospital room (DHR). Justa walks in the opposite direction, from the DHR to the UCR. At the beginning of the clip, Justa talks (2, 6) to people who are inside the DHR. Suzi also talks (3, 4) to someone who is inside the UCR. Then Suzi and Justa talk to each other before, during and after passing by each other (8-10) in Corridor B. Suzi informs Justa that a new patient is about to arrive at the clinic to have a catheter changed (8). In response, Justa questions the news, advancing her own epistemic state ("ah bah I do not know," 9). In overlap, Suzi emphatically reaffirms the factual accuracy of the informing, presenting herself as its guarantor ("yes yes I tell you," 10) and Justa accounts for her reluctance to engage further on the issue, claiming that she is busy (9).



Clip 1: Justa and Suzi's passing-by in Corridor B. Please click [here](#) to download the PDF file. [13]

5 We have replaced personal information, such as names of recorded individuals or people referred to in the recordings with fictitious ones. In the transcripts, underneath the original French talk, we provide an English translation (in gray) that is in line with its organization yet as natural as possible. We have supplemented the transcripts with images extracted from the recordings and added superscript annotations at the points in the transcript where the participants perform the actions shown in the images. In this analytical section, the numbers in parentheses refer to the lines of the transcript and the corresponding images; "(1, 1-1)," for instance, equates to "(Line 1, Image 1)." In the [Appendix](#), we provide the meaning of the symbols used in the transcripts based on JEFFERSON (2004).



In Clip 1, the participants walk by three static microphones suspended from light fixtures: Microphone 4, above the UCR entrance; Microphone 7, above the stock room (SR) door in the middle of Corridor B; and Microphone 8, above the DHR entrance. To edit and transcribe the clip, we needed to select a soundtrack. We could have chosen the track of Microphone 7, which is more or less in the middle of the scene. Nevertheless, the sound captured by the microphone closer to the speaker is better for hearing what she is saying: For instance, Microphone 8 when Justa is talking (2, 6) in the DHR and Microphone 4 when Suzi is talking (3, 4) in the UCR. For this reason, we edited and transcribed Clip 1, as with the two other clips examined in this section, by choosing the soundtrack of the microphone closest to the speaker for each turn-at-talk. If there is no talk, the microphone corresponds to the position of the person who has talked last. [14]

The result makes it easier for the analyst to hear what the speaker is saying, but does not correspond to the sound that could be heard at the spot where a distant listener might be positioned. For instance, Clip 1 reproduces the sound produced by Suzi talking from the UCR and not the sound heard in the DHR by Justa. However, these sounds have different features, which can have implications in terms of transcribing and analyzing the talk. We will demonstrate this point by comparing the sound recorded by several microphones for different fragments of each clip via visual representations. [15]

By representing some of the acoustic features of the audio clips, the figures are intended to furnish readers with some insight into the phenomenon under examination.<sup>6</sup> In the figures, we provide the following information in successive rows, from top to bottom: 1. clip and fragment, 2. microphone, 3. conversational transcript of the sound fragment based on the selected microphone, 4. Praat<sup>7</sup> visual representation of the sound fragment including orthographic labels of verbal production (spatially centered in the respective tier) based on the composite transcript, aligned with waveform, spectrogram, intensity trace (continuous) and pitch trace (dotted). The waveform and the pitch and intensity traces are not representations of the sound of individual voices but correspond to the ensemble of sound captured by the microphone at any given time. The analysis relies on them only when the fragment captures a single speaker talking in a quiet environment. We are nevertheless providing these representations for all the examined fragments for reasons of consistency and because readers may find them useful. [16]

Fragment 100.1 is drawn from Clip 1, Lines 2 and 3. Justa is talking to someone in the DHR and Suzi simultaneously talks to someone in the UCR (I-1). In Figure 2, we provide vis-à-vis visual representations of the sound of Microphone 4 (M4), the closest to Suzi, and Microphone 8 (M8), the closest to Justa.

6 In preparing the figures, we followed the recommendations of WALKER (2017, pp.365-385) except for issues like "scaling to the speaker's pitch rate" (pp.381-382), which were beyond the scope of our analytical interest and/or technical skills, distant from detailed phonetic study of the clips under consideration.

7 Praat is the scientific software that we used to read, analyze (spectral, pitch, formant and intensity analysis) and annotate the speech, and to create the graphics for our article.

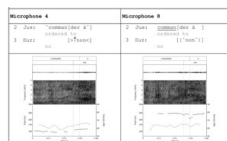


Figure 2: Clip 1, Fragment 100.1, sound of Microphones 4 and 8. Please click [here](#) for an enlarged version of Figure 2. [17]

M4 and M8 both capture what Justa is saying (2), but, as analysts, we get a different sense of it depending on the microphone. Justa's talk sounds softer in the recording captured by M4, the closest to Suzi, than with M8, the closest to the speaker. In the M4 recording, there is no trace of the initial emphasis that is noticeable in the M8 recording. Adopting the sound of M8 to represent Justa's utterance in the composite transcript is key for conveying not only how it is produced but also the fact that it can be heard by Suzi on the opposite side of the clinic. When Suzi starts her journey towards the DHR, she can tell, based on the sound of Justa's voice, not only that the aide is there but also that she is engaged in some type of discussion. However, the composite transcript, based on M8, does not correspond to how Suzi may have heard the fragment, perhaps softer than it was actually produced. On the other hand, the audibility and hearability of Suzi's "no" token produced in overlap (3) is very good in the recording made with M4, the closest to the speaker. In contrast, it is barely audible in the soundtrack of M8, the closest to Justa. However, marking it is necessary to portray what Suzi is doing at the beginning of the clip. She is talking to someone who is inside the UCR as she leaves it. She is apparently telling this person that she disagrees with him/her about something. Her next turn-at-talk in Line 4 ("he is all yellow") is connected to this utterance and addressed to the same interlocutor, not to Justa. [18]

Fragment 100.2 is also drawn from Clip 1, Line 4. Suzi is leaving the UCR, talking to a person who stays inside (I-2). In Figure 3, we provide contrasting visual representations of the sound of Microphone 4 (M4), the closest to Suzi, and Microphone 8 (M8), the closest to Justa, who is on the opposite side of the clinic.

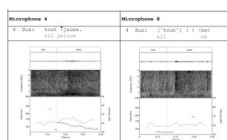


Figure 3: Clip 1, Fragment 100.2, sound of Microphones 4 and 8. Please click [here](#) for an enlarged version of Figure 3. [19]

Suzi's "he is all yellow" is clearly audible and understandable when using M4, the closest to her. Moreover, as shown in Figure 3, there is a sharp intonation at the beginning of "yellow" and it has a final intonation contour, which can be heard when using the track from M4. In contrast, it is not possible to hear the full utterance and make sense of it from the sound captured by M8, the closest to Justa. The sound of Suzi's voice is muffled, and the loudest recorded sound is a clicking noise from something in the surroundings about halfway through the

fragment (dark vertical band on the spectrogram). The difference is interactionally relevant. Suzi is already in the exit corridor of the UCR, heading towards the DHR. She cannot see Justa but she can already tell, based on Line 2, that the aide is there discussing some affair. With her utterance in Line 4, Suzi finishes talking to the person inside the UCR and makes herself available for interaction and talk with another speaker. In contrast, Justa, who is inside the DHR, talking in an animated tone of voice with other people, cannot really tell, from aural or visual means, that Suzi is heading towards her. She will discover this only when passing by the other side of the curtain of the DHR, still engaged in talk with the people inside (6). [20]

Fragment 100.3 is the third and last one drawn from Clip 1. Suzi and Justa have already engaged in talk and passed each other by. Suzi is about to reach the entrance of the DHR and Justa is walking in the opposite direction, towards the UCR (1-3). In Lines 9-10, Justa complements her doubt-implicative response to Suzi's previous informing with an account. In partial overlap to it, Suzi backs the informing up. In Figure 4, we compare the sound of Microphone 7 (M7), the closest to Justa, in the middle of Corridor B, and Microphone 8 (M8), the closest to Suzi, above the DHR entrance.

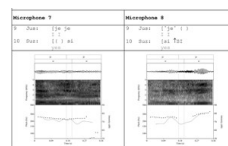
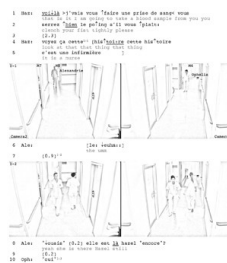


Figure 4: Clip 1, Fragment 100.3, sound of Microphones 7 and 8. Please click [here](#) for an enlarged version of Figure 4. [21]

Combining the sound of M7 and M8 is necessary to mark the exact onset of the overlap, this being a matter of precisely timed coordination among speakers (JEFFERSON, 1986; SCHEGLOFF, 2000). In the composite transcript of Clip 1, we marked that Justa completes the first unit of her turn—the challenging response to the informing—and that right after it, Suzi starts reaffirming the informing. She twice utters the token "si," which in French is used to counter a negative utterance. In overlap, Justa initiates a new unit—the account—but twice utters "I" as if orienting to the fact that Suzi is competing with her for the floor. Using only one microphone would provide a different understanding of the overlap. With M7, the closest to Justa, one may imagine that Suzi says something as Justa produces the first "I," but whether this is the case and what it may be is indiscernible. As Suzi can be heard saying "s" only once, and in a flat tone of voice, the fact that she is emphatically reaffirming the informing is not conveyed. In contrast, M8 conveys that Justa produces a very soft "I" during the overlap, followed potentially by something that is indiscernible, when in fact she is adamantly keeping the floor. In comparison to M7, M8 shows that Suzi is determined to take the floor, stressing the first "si" and producing the second one with a high pitch and loud tone of voice to back the informing up. [22]

In Clip 2, Alexandrie (Ale), the rotating nurse, walks from the DHR to a consultation and treatment room, called the specialist's room (SPE). Another nurse, Hazel (Haz), is there talking to a patient (1-5). Simultaneously, Ophelia (Oph), an aide, walks in the opposite direction, from the UCR towards the reception (I-1). As the two coworkers pass by each other (I-2), Alexandrie starts saying something to Ophelia (6) but breaks off and ends up, after the actual passing-by, asking her to confirm Hazel's whereabouts (8): She points towards the SPE and refers to it with the deictic *là* [there]. Ophelia nods, turns her head towards Alexandrie and provides the required confirmation (10). At this point, both coworkers have left Corridor B and are about to disappear from each other's view (I-3). As this is happening, Hazel herself identifies her location (11) and Alexandrie poses directly to her the question that she may have momentarily considered asking Ophelia in Line 6. As with Clip 1, while editing and transcribing Clip 2, we always used the microphone closest to the speaker.



Clip 2: Alexandrie and Ophelia's passing-by in Corridor B. Please click [here](#) to download the PDF file. [23]

Fragment 533.1 of Clip 2 corresponds to the very first line of the transcript: Hazel is inside the SPE talking to a patient, Alexandrie is still in the DHR and Ophelia in the UCR; none of them are visible to the others or on camera. Because Alexandrie will look for Hazel and ask for Ophelia's confirmation about her whereabouts, it is interesting to compare what Alexandrie and Ophelia can hear, from their respective locations. Can Ophelia hear that Hazel is in the SPE? What about Alexandrie? In Figure 5, we compare the sound of Microphone 4 (M4), the closest to Ophelia, Microphone 8 (M8), the closest to Jessica, and Microphone 5 (M5), the closest to Hazel.

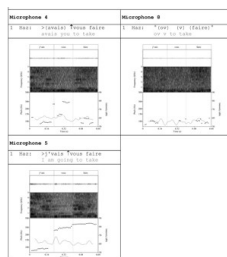


Figure 5: Clip 2, Fragment 533.1, sound of Microphones 4, 8 and 5. Please click [here](#) for an enlarged version of Figure 5. [24]

At the beginning of Clip 2, from their respective locations, both Alexandrie and Ophelia can hear someone talking in a nearby room, but the fact that the sound is Hazel's voice, what she is saying and her precise location comes across very differently in the recordings of M4, M8 and M5. Unsurprisingly, the clearest recording comes from M5, which is positioned in front of the SPE, where Hazel is attending to the patient. She is recognizable as the speaker and what she is saying can easily be heard and understood. Next is the track of M4, the closest to Ophelia, which is placed above the UCR entrance in the same corridor as the SPE. Inside the UCR, Ophelia may have heard the sound of Hazel's voice and been able to identify the speaker, but she would have had a hard time locating her in the clinic and telling exactly what she was saying. Microphone 8, which is above the DHR's entrance and is the closest to Alexandrie, captures Hazel's talk only very faintly. From inside the DHR, it is very unlikely that Alexandrie would have heard Hazel's voice and been able to identify her and her location. Clearly discerning that Hazel is speaking from inside the SPE was possible only for the participants near this section of the clinic. This comes first for Ophelia, as she exits the UCR and then walks through Corridor B at the same time that Hazel talks again in Lines 4-5 (I-1). This is relevant for understanding the subsequent interaction between Alexandrie and Ophelia. Alexandrie does not go out of the DHR and head to the SPE in search of Hazel based on information about her location that would be available perceptually as she starts her journey. At this time, she seems rather to rely on previous knowledge about Hazel's doings in the clinic, which she may assume Ophelia has as well. Indeed, later on, she asks Ophelia whether Hazel is "still there" (8), implying that Hazel has been in the SPE for some time and that she, as well as Ophelia, knows this. At the time of this request for confirmation, Alexandrie may also be relying on the fact that Ophelia has just passed close to the SPE and that Hazel's voice can by then be heard by both coworkers as they pass each other by in Corridor B. [25]

Fragment 533.2 of Clip 2 corresponds to the moment when Ophelia responds to Alexandrie's confirmation request about Hazel's whereabouts (10). Hazel is inside the SPE, out of her coworkers' view. Jessica has just turned into Corridor A, Ophelia has done the same into Corridor C, and both are about to disappear from each other's view. In Figure 6, we compare the sound of Microphone 4 (M4), the closest to Alexandrie, Microphone 8 (M8), the closest to Ophelia, and Microphone 5 (M5), the closest to Hazel.

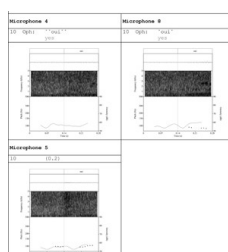
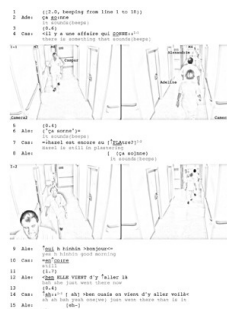


Figure 6: Clip 2, Fragment 533.2, sound of Microphones 4, 8 and 5 Please click [here](#) for an enlarged version of Figure 6. [26]

In Line 10, Ophelia utters an affirmative response to Alexandrie's confirmation request. Although produced in a soft tone of voice, the response is clearly heard when using M8, the closest to Ophelia. It sounds very soft with M4, the closest to Alexandrie, and cannot be heard at all with M5, the closest to Hazel. Before responding to Alexandrie, Ophelia turns her head and looks at her, which may have been useful for eliminating any possible ambiguity in the indexical "/à" featuring in the request. When responding softly to Alexandrie, Ophelia may have oriented to what she has just seen: Alexandrie is already far away from her and walking towards the SPE, to which she is pointing (l-3). The fact that Hazel cannot hear Ophelia's response and that the "yes" token is barely audible to Alexandrie is interactionally relevant. On the one hand, Hazel seems to hear Alexandrie inquiring about her whereabouts (8), waits for a few tenths of a second and then, concluding that there has not been an answer, responds herself (11). On the other hand, Alexandrie does not orient to Ophelia's response but engages in talk with Hazel that is responsive to Hazel's own turn in Line 11. [27]

In Clip 3, Caspar (Cas), the head nurse, is followed by Adeline (Ade), the hospital's chief nurse, who is visiting the clinic. They come from the reception, turn into Corridor B and walk over to the UCR. In Corridor B, they pass by Alexandrie (Ale), a rotating nurse, who talks to Caspar and greets Adeline as she walks in the opposite direction from the UCR to the DHR. At the beginning of the clip, Adeline produces a noticing (2) about a beeping sound coming from the DHR. Beeping machines may indicate either the conclusion of a treatment procedure or improper use, and the clinic's nurses are expected to closely monitor them to resolve any problem immediately. In Line 4, Caspar acknowledges and revoices the noticing using a "there is x" statement that also functions as a nudge to take action (GONZÁLEZ-MARTÍNEZ, 2023) oriented to any nearby nurse and especially, as soon as she shows up, to the approaching Alexandrie (l-1). The nurse also acknowledges the beeping (6, 8) and engages in talk with Caspar about the fact that Hazel, the nurse in charge of the DHR, is in the plaster room (7-15, l-2). Caspar conveys that she has been there for too long (7, 10), thus not attending to the DHR, where the machine is beeping, and Alexandrie counters this (12) but does not manage to persuade Caspar, who counters back (14, l-3). At the end of the clip, Alexandrie goes into the DHR, talks to a patient about the beeping (15), and stops it.



Clip 3: Caspar, Adeline and Alexandrie's passing-by in Corridor B. Please click [here](#) to download the PDF file. [28]

As with Clips 1 and 2, as we edited and transcribed Clip 3, we always used the microphone closest to the speaker. But in this case too, it is possible to hear things slightly differently depending on the microphone, which is consequential for understanding the interaction going on. Fragment 1205.1 corresponds to Line 4 of Clip 3, when Caspar produces the "there is x" statement as he is located between Microphones 8 (M8), at his back, and Microphone 7 (M7), at his right side. In Figure 7, we compare the sound between these two microphones and Microphone 4 (M4), the closest to Alexandrie.

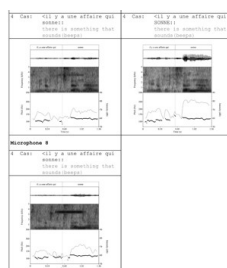


Figure 7: Clip 3, Fragment 1205.1, sound of Microphones 4, 7 and 8. Please click [here](#) for an enlarged version of Figure 7. [29]

With the three microphones, we, the analysts, can hear that Caspar puts some emphasis on the beginning of "*sonne*" [sounds, beeps]. Nevertheless, the increase in volume is most noticeable with M8 and M7, the closest to him. Moreover, it is only with M7 that we notice that Caspar is really talking much louder than at the beginning of his utterance when he had just turned into Corridor B. The observation is interactionally relevant because it is just before "*sonne*" that Alexandrie shows up and Caspar looks at her. From acknowledging and revoicing Adeline's noticing to her attention and that of any nearby nurses, Caspar moves to nudging Alexandrie in particular, who in response shows that she is aware of the beeping (6, 8). If we focus on the machine's sound, we, the analysts, can also note that M8 captures the beeping very distinctly; it is represented by a long horizontal band in the center of the spectrogram, towards the upper part. With M4, the beeping is still audible, but not as loud. In contrast, the beeping cannot be heard at all with M7. [30]

In the spatially distributed workplace of the clinic, a skill that the staff needs to have is the ability to hear spatialized and directional sounds and to identify not only the sounds' places of origin but what is producing them and, in the case of a moving source, where it is going. The fact that the audibility of machine signals is spatially differentiated—they can be heard better from some locations than others—is common knowledge among the hospital staff. They orient towards this with practices like keeping the doors open to allow sound to travel easily as well as with specific interactional activities like the ones at work in Clip 3: Sharing sonic observations relevant to the clinic's functioning, revoicing them to the attention of third parties, nudging these third parties to take action, receipting the nudge and acting on it, as well as interpreting the sound for lay fellows. In this sense, Clip 3 serves as evidence that audibility and intelligibility are not only a problem for the analyst but also a practical concern for the clinical personnel, who address them

through specific practical activities and even turn them into "talkables" in the studied professional setting. [31]

## 6. Discussion and Conclusion

Through the present study, we have arrived at three main findings. First, the analysts' take on the audio recordings of the studied interactions differs substantially from the members' aural experience in situ and in real time. The participants hear other persons' speech, first-time-through, one time and the only time, from their unique and contingent spatial and practical perspectives. Conversely, the analyst's hearing and understanding is a scientific artifact. The analyst is able to listen to the recordings repeatedly, to compare several soundtracks of the same fragment of talk as captured by various microphones and to set multiple instances of talk by the very same speakers side by side. One recording elaborates the others, and is elaborated by them in turn, producing the acoustic features and contents of the talk, which end up being heard and understood "all along" as presented in composite transcripts and analysis. [32]

Second, interactionally relevant features of speech, such as volume, emphasis and pitch, compositional elements, aspects of sequential organization and the identity and location of the speaker are sometimes conveyed differently by the recordings of a single talk fragment captured by static microphones placed in different locations. When participants are on the move, none of the recordings exactly represents the participants' own situated hearing "experience" since their acoustic environment changes as they move through the space. [33]

Third and last, audibility and intelligibility are consequentially spatially differentiated: Sounds are heard and understood differently depending on the location of the receiver. When participants are spatially apart, their acoustic environments differ. This is relevant for the overhearing analyst as well as for interactional purposes and it occasionally also becomes a members' topic. [34]

Through this article, we contribute to the project of turning sensorial resources for interaction and analysis into objects of study. We caution against the idea of a "realist tape": A non-problematic slice of reality "that can be recaptured through replaying" (ASHMORE et al., 2004, p.355; see also ASHMORE & REED, 2000). Famously, SACKS (1984, p.26) argued that the recording was to be regarded as a "good enough" record of what happened": While other things might have happened too, "at least what was on the tape had happened." Built on this basis, the approaches of conversation and multimodal analysis to which we adhere, have been extraordinarily fruitful in detailing the situated achievement of talk and social action in interaction. At a time when research is moving towards increasingly complex recording set-ups and objects of study—mobile, spatially distributed, multisensorial interactions—we consider that it is worth rethinking "For whom exactly what is on the tape had happened?" In 2012, LUFF and HEATH forewarned that



"[u]sing multiple high-definition cameras and combining visual (and audio) data from different sources provide greater access to the setting, to the details of an activity, and can allow researchers to reveal the interrelationships between distributed activities. However, when analyzing these, the researcher still has to consider the ways in which these multiple views resonate with the perspectives of the participants being recorded, how what is visible in the data is visible to them" (p.275). [35]

We have come to reflect on the aural side of the problem of multiple views of a "single" scene by examining a common but understudied interactional and spatial formation: Participants talking while apart, before and after passing each other by. Examining mobile and spatially distributed interactions furthers our understanding of social arrangements in space and also of the methodological and analytical resources needed to come to grips with them. We could even argue that the aural problem is trickier than the visual one, since it is possible to show multicam "split screen" clips but several audio tracks playing simultaneously do not merge into a consistent rendering of a scene's soundscape. How and to what extent do the recordings "preserve" (MONDADA, 2006, p.54) the distinct lived experience of each participant of the scene? How are the participants themselves providing for the audibility, "recognizability, [and] intelligibility of the phenomena at hand"? (p.52). How do data collection activities provide insights "into the organizational features of the recorded practices themselves"? (ibid.). [36]

MODAFF and MODAFF (2000) identified differences between the recordings made at both ends of the line in a telephone conversation but concluded that they did not "appear to have interactional consequences for the co-participants" (p.111). We have shown differences that may matter when producing and analyzing spatially distributed mobile interactions: "[A]ttending to the details of the spacing of actions are as much part of the members' analyses as attending to the details of timing of actions" (LAURIER et al., 2020, p.354). In passing-by interactions, "just where" someone talks is interconnected to "just when" they do, as well as to what they say and accomplish through it; all the aspects deserve the same attention. Considering the central role of the audibility and intelligibility of talk in conversation and multimodal analysis, spatializing the sound fragments under examination would be a relevant move in the current turn towards multisensoriality. [37]

## **Acknowledgments**

This article was prepared within the framework of the research projects "Mobile and Contingent Work Interactions in the Hospital Care Unit" (SNSF, grant no. 134875) and "Requesting in Hospital Nurses' Unscheduled Interprofessional Interaction" (SNSF, grant no. 185152). The first author presented a preliminary version at the Big Video Sprint 2017 at Aalborg University on November 22, 2017; she thanks the participants for their comments. Both authors are grateful to the anonymous reviewer and the editor for their insightful comments and to Elisabeth LYMAN for her editing work.

## Appendix: Transcription Conventions

[ ]	overlapping talk
=	continuous talk
(0.2)	silence in tenths of a second
.	final intonation
?	rising intonation
:	prolongation of the preceding sound
speci-	cut-off
<u>you</u>	emphasis
°yes°	softer talk
°°yes°°	markedly softer talk
DANIEL	louder talk
↑	rise in pitch
↓	fall in pitch
>yes<	talk is compressed
<because	hurried start
·h	inhalation
h	exhalation
((beeping))	transcriber's description of events
( )	unachievable or likely hearing

Table 1: Transcription conventions based on JEFFERSON (2004)

## References

Ashmore, Malcolm & Reed, Darren (2000). Innocence and nostalgia in conversation analysis: The dynamic relations of tape and transcript. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(3), Art. 3, <https://doi.org/10.17169/fqs-1.3.1020> [Accessed: July 29, 2024].

Ashmore, Malcolm; MacMillan, Katie & Brown, Steven D. (2004). It's a scream: Professional hearing and tape fetishism. *Journal of Pragmatics*, 36(2), 349-374.

Bjelić, Dušan (2019). "Hearability" versus "hearer-ship": Comparing Garfinkel's and Schegloff's accounts of the summoning phone. *Human Studies*, 42(4), 695-716.

Coulter, Jeff & Parsons (Ed.) (1990). The praxiology of perception: Visual orientations and practical action. *Inquiry*, 33(3), 251-272.

De Stefani, Elwys & Mondada, Lorenza (2014). Reorganizing mobile formations: When "guided" participants initiate reorientations in guided tours. *Space and Culture*, 17(2), 157-175.

Drew, Paul (1997). "Open" class repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics*, 28(1), 69-101.

Gibson, Will & vom Lehn, Dirk (2021). Introduction: The senses in social interaction. *Symbolic Interaction*, 44(1), 3-9, <https://onlinelibrary.wiley.com/doi/full/10.1002/symb.539> [Accessed: July 29, 2024].

Glenn, Phillip J.; LeBaron, Curtis D. & Mandelbaum, Jenny (2003). *Studies in language and social interaction*. Mahwah, NJ: Lawrence Erlbaum.

Goffman, Erving (1959). *The presentation of self in everyday life*. New York, NY: Doubleday Anchor Books.

Goffman, Erving (1963). *Behavior in public places: Notes on the social organization of gatherings*. New York, NY: The Free Press.

Goffman, Erving (1971). *Relations in public: Microstudies of the public order*. New York, NY: Basic Books.

Goico, Sara; Gan, Yumei; Katila, Julia & Goodwin, Marjorie Harness (2021). Capturing multisensoriality: Introduction to a special issue on sensoriality in video-based fieldwork. *Social Interaction: Video-Based Studies of Human Sociality*, 4(3), 1-13, <https://tidsskrift.dk/socialinteraction/article/view/128144> [Accessed: July 29, 2024].

González-Martínez, Esther (2023). "Faire du coude" avec un énoncé "il y a x": Inciter à l'action et recruter une personne responsable d'agir en milieu hospitalier ["Nudging" using the phrase "there is x": Prompting action and recruiting a person responsible for action in a hospital setting]. *Langage et Société*, 179(2), 111-139.

González-Martínez, Esther; Bangerter, Adrian & Lê Van, Kim (2017). Passing-by "Ça va?" checks in clinic corridors. *Semiotica*, 215, 1-42.

González-Martínez, Esther; Bangerter, Adrian; Lê Van, Kim & Navarro, Cécile (2016). Hospital staff corridor conversations: Work in passing. *Journal of Advanced Nursing*, 72(3), 521-532.

Goodwin, Charles (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32(10), 1489-1522.

Groeber, Simone & Pekarek Doehler, Simona (2012). Hearing impaired adolescents in a regular classroom: On the embodied accomplishment of participation and understanding. In Maria Egbert & Arnulf Deppermann (Eds.), *Hearing aids communication: Integrating social interaction, audiology and user centered design to improve communication with hearing loss and hearing technologies* (pp.76-89). Mannheim: Verlag für Gesprächsforschung, <http://www.verlag-gespraechsforschung.de/2012/egbert.html> [Accessed: July 29, 2024].

Haddington, Pentti; Eilittä, Tiina; Kamunen, Antti; Kohonen-Aho, Laura; Oittinen, Tuire; Rautiainen, Lira & Vatanen, Anna (Eds.) (2023). *Ethnomethodological conversation analysis in motion: Emerging methods and new technologies*. London: Routledge

Heinemann, Trine & Rauniomaa, Mirka (2016). Turning down sound to turn to talk: Muting and muffling auditory objects as a resource for displaying involvement. *Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion*, 17, 1-28, <http://www.gespraechsforschung-online.de/fileadmin/dateien/heft2016/ga-heinemann.pdf> [Accessed: July 29, 2024].

Hochuli, Kenan (2019). Turning the passer-by into a customer: Multi-party encounters at a market stall. *Research on Language and Social Interaction*, 52(4), 427-447.

Hoey, Elliot M. (2023). Anticipatory initiations: The use of a presumed reason-for-the-interaction in face-to-face openings. *Journal of Pragmatics*, 209, 41-55.

Jefferson, Gail (1986). Notes on "latency" in overlap onset. *Human Studies*, 9(2-3), 153-184, <https://liso-archives.liso.ucsb.edu/Jefferson/Latency.pdf> [Accessed: July 29, 2024].

Jefferson, Gail (2004). Glossary of transcript symbols with an introduction. In Gene H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp.13-31). Amsterdam: John Benjamins.

Jefferson, Gail (2010). Sometimes a frog in your throat is just a frog in your throat: Gutturals as (sometimes) laughter-implicative. *Journal of Pragmatics*, 42(6), 1476-1484.

Keevallik, Leelo & Ogden, Richard (2020). Sounds on the margins of language at the heart of interaction. *Research on Language and Social Interaction*, 53(1), 1-18, [https://eprints.whiterose.ac.uk/155325/8/Sounds\\_on\\_the\\_Margins\\_of\\_Language\\_at\\_the\\_Heart\\_of\\_Interaction.pdf](https://eprints.whiterose.ac.uk/155325/8/Sounds_on_the_Margins_of_Language_at_the_Heart_of_Interaction.pdf) [Accessed: July 29, 2024].

Kendon, Adam (1976). The F-formation system: The spatial organization of social encounters. *Man-Environment Systems*, 6(5), pp. 291-296.

- La Valle, Natalia (2014). Les corps, les mots, les choses: Particules discursives et "trajectoires parlées" dans l'ordonnement de la vie domestique [Bodies, words, things: Discourse particles and "spoken trajectories" in the arrangement of domestic life]. In Lorenza Mondada (Ed.), *Corps en interaction: Participation, spatialité, mobilité* [Bodies in interaction: Participation, spatiality, mobility] (pp.259-288). Lyon: ENS Editions.
- Laurier, Eric; Whyte, Angus & Buckner, Kathy (2001). An ethnography of a neighborhood café: Informality, table arrangements and background noise. *Journal of Mundane Behavior*, 2(2), 195-232.
- Laurier, Eric; Muñoz, Daniel; Miller, Rebekah & Brown, Barry (2020). A Bip, a beeeep, and a beep beep: How horns are sounded in Chennai traffic. *Research on Language and Social Interaction*, 53(3), 341-356.
- Lee, John R.E. (1987). Prologue: Talking organisation. In Graham Button & John R.E. Lee (Ed.), *Talk and social organisation* (pp.19-53). Clevedon: Multilingual Matters.
- Licoppe, Christian (2010). The "crisis of the summons": A transformation in the pragmatics of "notifications," from phone rings to instant messaging. *The Information Society*, 26(4), 288-302.
- Luff, Paul & Heath, Christian (2012). Some "technical challenges" of video analysis: Social actions, objects, material realities and the problems of perspective. *Qualitative Research*, 12(3), 255-279.
- Mcllvenny, Paul & Davidsen, Jacob (2017). A big video manifesto: Re-sensing video and audio. *Nordicom-Information*, 39(2), 15-21, [https://www.nordicom.gu.se/sites/default/files/kapitel-pdf/mcilvenny\\_davidsen.pdf](https://www.nordicom.gu.se/sites/default/files/kapitel-pdf/mcilvenny_davidsen.pdf) [Accessed: July 29, 2024].
- Merlino, Sara (2021). Making sounds visible in speech-language therapy for aphasia. *Social Interaction: Video-Based Studies of Human Sociality*, 4(3), 1-30, <https://tidsskrift.dk/socialinteraction/article/view/128151> [Accessed: July 29, 2024].
- Merlino, Sara; Mondada, Lorenza & Söderström, Ola (2022). Walking through the city soundscape: An audio-visual analysis of sensory experience for people with psychosis. *Visual Communication*, 22(1), 71-95, <https://journals.sagepub.com/doi/full/10.1177/14703572211052638> [Accessed: July 29, 2024].
- Modaff, John V. & Modaff, Daniel P. (2000). Technical notes on audio recordings. *Research on Language and Social Interaction*, 33(1), 101-118.
- Mondada, Lorenza (2006). Video recording as the reflexive preservation and configuration of phenomenal features for analysis. In [Hubert Knoblauch](#), Bernt Schnettler, Jürgen Raab & Hans-Georg Soeffner (Eds.), *Video analysis: Methodology and methods. Qualitative audiovisual data analysis in sociology* (51-67). Bern: Peter Lang.
- Mondada, Lorenza (2007). Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse Studies*, 9(2), 194-225.
- Mondada, Lorenza (2009). Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space. *Journal of Pragmatics*, 41(10), 1977-1997.
- Mondada, Lorenza (2021). Orchestrating multi-sensoriality in tasting sessions: Sensing bodies, normativity, and language. *Symbolic Interaction*, 44(1), 63-86.
- Ochs, Elinor (1979). Transcription as theory. In Elinor Ochs & Bambi B. Schieffelin (Eds.), *Developmental pragmatics* (pp.43-72). New York, NY: Academic Press, <https://www.sscnet.ucla.edu/anthro/faculty/ochs/articles/ochs1979.pdf> [Accessed: July 29, 2024].
- Psathas, George (1990). Introduction. Methodological issues and recent developments in the study of naturally occurring interaction. In George Psathas (Ed.), *Interaction competence* (pp.1-29). Washington D.C.: University Press of America.
- Psathas, George & Anderson, Timothy (1990). The "practices" of transcription in conversation analysis. *Semiotica*, 78(1-2), 75-99.
- Rauniomaa, Mirka & Heinemann, Trine (2014). Organising the soundscape: Participants' orientation to impending sound when turning on auditory objects in interaction. In Maurice Nevile, Pentti Haddington, Trine Heinemann & Mirka Rauniomaa (Eds.), *Interacting with objects: Language, materiality and social activity* (pp.145-168). Amsterdam: John Benjamins.
- Relieu, Marc (2023). Teaching and learning how to identify an audible order in traffic: Street-crossing instructional sequences for the visually impaired. In Sara Keel (Ed.), *Medical and healthcare interactions: Members' competence and socialization* (pp.152-175). London: Routledge.

Sacks, Harvey (1984). Notes on methodology. In John Heritage & J. Maxwell Atkinson (Eds.), *Structures of social action: Studies in conversation analysis* (pp.21-27). Cambridge: Cambridge University Press.

Sacks, Harvey; Schegloff, Emanuel A. & Jefferson, Gail (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696-735, [https://liso-archives.liso.ucsb.edu/Jefferson/systematics\\_Schenkein.pdf](https://liso-archives.liso.ucsb.edu/Jefferson/systematics_Schenkein.pdf) [Accessed: July 29, 2024].

Schegloff, Emanuel A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(1), 1-63, <https://www.cambridge.org/core/journals/language-in-society/article/overlapping-talk-and-the-organization-of-turnstile-taking-for-conversation/D12AD4181BEEFB073869BE7C55B07660> [Accessed: July 29, 2024].

Schegloff, Emanuel A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.

Sharrock, Wes & Coulter, Jeff (1998). On what we can see. *Theory and Psychology*, 8(2), 147-164.

Smith, Michael Sean (2021). Achieving mutual accessibility through the coordination of multiple perspectives in open, unstructured landscapes. *Social Interaction: Video-Based Studies of Human Sociality*, 4(3), 1-24, <https://tidsskrift.dk/socialinteraction/article/view/128178> [Accessed: July 29, 2024].

Walker, Gareth (2017). Visual representations of acoustic data: A survey and suggestions. *Research on Language and Social Interaction*, 50(4), 363-387.

Weeks, Peter (2002). Performative error-correction in music: A problem for ethnomethodological description. *Human Studies*, 25(3), 359-385.

Yagi, Junichi (2021). Embodied micro-transitions: A single-case analysis of an amateur band studio session. *Social Interaction: Video-Based Studies of Human Sociality*, 4(4), 1-28, <https://tidsskrift.dk/socialinteraction/article/view/128655> [Accessed: July 29, 2024].

## Authors

*Esther GONZALEZ-MARTINEZ* is full professor of sociology at the University of Fribourg, Switzerland. In her research, she adopts an ethnomethodological perspective and relies on conversation and multimodal analysis, supplemented by ethnographic fieldwork, to study the situated organization of social interactions in institutional settings. In her first studies, she centered around police and judicial interactions in France and the United States. Over the past decade, she has conducted several research projects on nurses' unscheduled interactions with co-workers in Swiss hospitals, with particular focus on telephone conversations, corridor interactions, mobility practices and recruiting moves.

Contact:

Esther González-Martínez

Department of Social Sciences, University of Fribourg  
Boulevard de Pérolles 23  
1700 Fribourg, Switzerland

E-mail: [esther.gonzalezmartinez@unifr.ch](mailto:esther.gonzalezmartinez@unifr.ch)

URL: <https://projects.unifr.ch/grips/en/>

*Jakub MLYNÁŘ* is a scientific collaborator at HES-SO Valais-Wallis University of Applied Sciences Western Switzerland, working with the Human-Centred Computing Group at the Institute of Informatics. In his research, he focuses on the sociological aspects of digital technology and AI from the perspective of ethnomethodology and conversation analysis. He is currently working on a user study situated in the field of radiomics (using machine-learning algorithms to extract large-scale quantitative features from medical imaging) and a research project to investigate the constitutive practical knowledge of vehicle operators in street trials of driverless buses.

Contact:

Jakub Mlynář

HES-SO Valais-Wallis, Institute of Informatics  
Rue de Technopôle 3  
3960 Sierre, Switzerland

E-mail: [jakub.mlynar@hevs.ch](mailto:jakub.mlynar@hevs.ch)

## Citation

González-Martínez, Esther & Mlynář, Jakub (2024). Audibility and intelligibility of talk in passing-by interactions: A methodological puzzle [37 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 25(3), Art. 8, <https://doi.org/10.17169/fqs-25.2.4207>.