Designing Robot Sound-In-Interaction: The Case of Autonomous Public Transport Shuttle Buses

Hannah R. M. Pelikan
hannah.pelikan@liu.se
Linköping University
Linköping, Sweden

Malte F. Jung
mjf28@cornell.edu
Cornell University
Ithaca, NY, USA

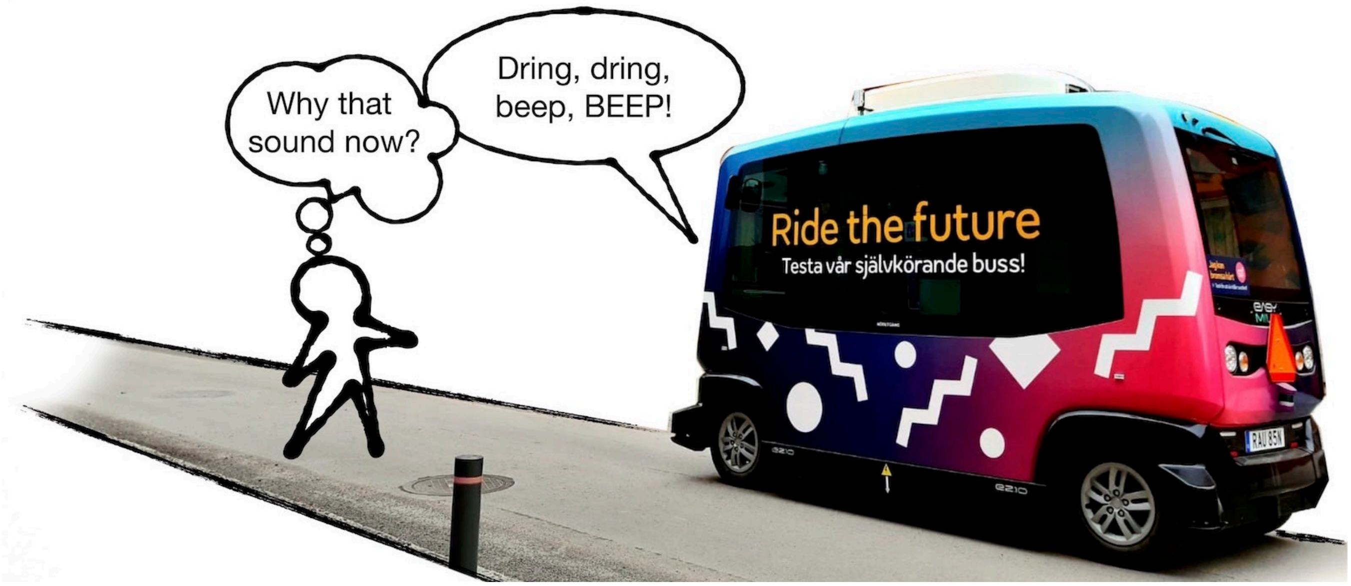

Figure 1: Bus playing sounds at a pedestrian, who asks "why that now?"[88] in line with the EMCA next-turn proof-procedure.

ABSTRACT

Horns and sirens are important tools for communicating on the road, which are still understudied in autonomous vehicles. While HRI has explored different ways in which robots could sound, we focus on the range of actions that a single sound can accomplish in interaction. In a Research through Design study involving autonomous shuttle buses in public transport, we explored sound design with the help of voice-overs to video recordings of the buses on the road and Wizard-of-Oz tests in live traffic. The buses are slowed down by (unnecessary) braking in response to people getting close. We found that prolonged jingles draw attention to the bus and invite interaction, while repeated short beeps and bell sounds can instruct the movement of others away from the bus. We highlight the importance of designing sound in sequential interaction and describe a new method for embedding video interaction analysis in the design process.

HRI '23, March 13–16, 2023, Stockholm, Sweden

ACM ISBN 978-1-4503-9964-7/23/03.
https://doi.org/10.1145/3568162.3576979

CCS CONCEPTS

• **Human-centered computing** → **Interaction design process and methods**; **Auditory feedback**; *Field studies.*

KEYWORDS

sound design, ethnomethodology, conversation analysis

ACM Reference Format:
Hannah R. M. Pelikan and Malte F. Jung. 2023. Designing Robot Sound-In-Interaction: The Case of Autonomous Public Transport Shuttle Buses. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3568162.3576979

1 INTRODUCTION

Sound becomes an important resource for communication when opportunities for verbal expression are limited. In traffic, horns have traditionally been used for this purpose: Tram and train drivers toot to warn others to get off the tracks, boat captains use fog horns to instruct other ships to get out of the way. Car drivers honk to warn about upcoming situations or to complain about others' maneuvers [52]. In all these cases, the choice of different sounds is heavily restricted, since usually there is only one horn. Yet, with only a single sound, people are able to navigate an astonishing array of interactions, including warning others, negotiating difficult traffic situations, as well as greeting or saying goodbye. This paper explores how this can be applied to robots, looking at the case of autonomous shuttle buses in public transport.

Robot sound has received increased attention over the last decade, with researchers exploring a range of different ways in which robots could sound [99]. Taking inspiration from movies [51], a variety of possible sounds have been explored, including motor hums [68, 95], musical sonifications [26, 82, 86], and beep sequences [25, 79, 101]. The majority of these studies focuses on validating the design of specific sounds, ensuring that users interpret them consistently. However, building on the insight that understanding is negotiated in interaction [11, 85], an utterance or sound in isolation can only ever have *meaning potentials* [56]. A sound's specific meaning then emerges in concrete, situated interactions [74, 91]. Building on this work, we are interested in how robot sound gets interpreted in live interaction and what actions robots can achieve through the way sound is timed in interaction. We present an approach to designing sound in concrete interactional sequences rather than in isolation.

Looking at design research more generally, recent work has problematized how knowledge is produced in HRI. Lupetti et al. [57] point out that typical HRI design processes result in standalone design instances that are often difficult for other designers to build on. A lack of intermediate-level knowledge [42] makes translation to other contexts difficult. Zamfirescu-Pereira et al. [100] similarly argue that research should engage in design exploration rather than in controlled experimentation that validates one specific design. Research through Design approaches that aim to facilitate generalization beyond individual designs are only recently gaining popularity in HRI [27, 41, 54, 58, 59, 82]. Our work aims to make a methodological contribution to this emerging body of work: We demonstrate how Ethnomethodology and Conversation Analysis [8, 12, 30] (EMCA), an approach to studying interaction, previously proposed to be suitable for moving beyond design critique towards design practice [7], can be embedded into HRI design processes. EMCA has been applied in the evaluation of interaction with robots [32, 74, 75], and it has informed robot design through literature [71, 75] and concurrent ethnographic studies [48]. However, we provide the first attempt at systematically integrating EMCA video analysis into ongoing design iterations.

In this paper we designed sound for an autonomous shuttle bus based on video-recordings of live traffic and tested them with a Wizard-of-Oz [76] setup during rides on public roads. Taking an EMCA approach, we study traffic as inherently social and focus on observable actions (rather than internal states or intentions). Concentrating on what can be observed, the approach is particularly relevant for studying traffic safety: When coordinating on the road, people do not typically discuss their impressions and preferences but mostly act in response to others' visible and recognizable behavior.

The contribution of this paper is threefold: first, we contribute a novel design approach that tightly intertwines EMCA video analysis with interaction design. Second, we contribute intermediate-level knowledge [42, 57] in the form of transcribed video recordings of our designs for public transport buses, showing the specificity of the designs while opening possibilities for generalization. Third, we share lessons learned from testing robot sound in the wild, on public roads.

2 MOBILITY, COORDINATION AND SOUND

Moving in spaces where others are present is a highly social endeavor. People do not move like a bullet fired towards a destination, clashing when their trajectories interfere but instead they carefully coordinate their mobility. This section reviews prior work on traffic as social interaction, highlights how autonomous vehicles have been designed to communicate on the road, and how sound may be used as such a resource in robots.

2.1 Traffic is Social

Ethnomethodologists have studied the "common sense" involved in moving in space for decades, showing that people dynamically adapt their speed and trajectory to be recognized as walking and cycling alone or together [62, 84], and demonstrating that moving together with others is a skillful accomplishment [36, 63]. Car drivers negotiate movement with other road users, for instance by offering space for others to pass through [37], or letting others overtake by moving to the side of a lane and slowing down [16]. As Goffman [33] highlighted early on, people act in recognizable ways in public spaces, and deviations from the social order require explanations. Opportunities for providing verbal accounts are typically limited on the road, and instead drivers communicate through movement [80], indicators [3], headlights [37], and horns [52]. While work in HRI has acknowledged the importance for robots to move in socially recognizable ways [93], autonomous vehicles still struggle to participate in the social coordination in traffic [5, 6, 73, 90, 96].

2.2 Interfaces for Autonomous Vehicles

To help mobile robots and autonomous vehicles navigate public spaces, research has explored a range of modalities. While sound has been explored typically in combination with visual feedback [49, 61], a majority of studies focuses on visual displays such as animated lights [13, 17] or using a robot head in the position of the driver for anthropomorphic feedback [66]. While such external human-machine interfaces typically consist in novel additions to vehicles, a small body of research has highlighted that autonomous vehicles already *implicitly* communicate their states and intents through their movement and explicit signals may only seldomly be necessary [19, 69, 80]. Synthetic motor sounds, legally compulsory in the European Union since 2019 [23], are one example of such implicit interfaces [68]. To date, the majority of studies of interfaces for autonomous vehicles are carried out in controlled settings, through video prototypes [18], virtual reality [13, 43] or experiments in closed-off parking lots [35, 61]. A small body of work has explored autonomous driving in the wild with a hidden human driver in the "ghost-driver" paradigm [55, 83]. Following this paradigm, Moore et al. [68] demonstrated that synthetic motor sound can augment a car's slowing movement, highlighting that the vehicle will yield. To our best knowledge however, our paper is the first report on sound design iterations on a fully autonomous vehicle in live traffic.

2.3 Sound in HRI

HRI work has demonstrated that even beyond autonomous vehicles, motor sounds and musical sonifications influence how robots are perceived [25, 70, 82, 95, 101]. They can communicate intent and support the localization of mobile robots [9, 47, 68]. Sound as a broader category (see [99] for an overview) has received increased attention in HRI during the last years. The large majority of research has focused on validating how well specific sounds can communicate specific emotions [10, 26, 45, 51, 79, 86]. A small body of work has pointed out that validating sounds may only be useful to a certain degree, since humans interpret robot sound differently depending on the specific interactional context that they occur in [74, 78]. While most papers deal with evaluating specifically designed sounds, recent work has started to formulate more general principles for sound design based on the work of professional sound designers [50, 81]. These works provide guidelines on *how* robots should sound, but the question *when* sound is actually relevant in interaction remains underexplored. As Beaudouin-Lafon [2, p.21] put it for human-computer interaction (HCI): "HCI is not the science of user interfaces, just as astronomy is not the science of telescopes. HCI needs interfaces to create interaction, and we should focus on describing, evaluating and generating interaction, not interfaces." Similarly, we argue that HRI sound design should focus not on designing (standalone) sounds, but instead on describing, evaluating and generating interaction through these sounds.

Building on the insight that traffic is social, we explore how sound can facilitate this interaction. While adjusting sounds to the specific character of the robot is important, we argue that a crucial step should come before this: designing the interaction, focusing on what actions sound can and should accomplish.

3 SETTING

We present a case study based on a project called Ride the Future[1], where autonomous shuttle buses are tested as a future public transportation solution in the Swedish city of Linköping. The project is driven by the local public transport provider, several research institutes, as well as the municipality. The buses serve several stops on a university campus and in a close-by neighborhood. Rides are for free, and while there is no fixed schedule, passengers can check the location of the buses on a live map. We followed the project since the first bus started rolling in January 2020.

3.1 Autonomous Shuttle Buses

Being built for public use, the electric shuttle buses need to be affordable for local transport providers and are quite different from autonomous cars. They resemble a tram on invisible tracks and drive on a programmed route, which they cannot divert from. When facing an obstacle, the buses reliably slow down, and eventually stop but they do not change their trajectory. The shuttles largely maneuver without human input but always have a safety driver on board who may switch to manual control when the bus gets stuck.

The project started with an EasyMile EZ10 shuttle (see Figure 1) and a Navya Autonomous Shuttle DL4, later a second EZ10 was added. We initially studied buses from both manufacturers and found that they face similar challenges on the road [73]. However, since the EZ10 shuttles have fewer built-in sounds, we carried out all sound prototyping on them. Our designs are overlaid onto the EZ10's existing soundscape, which includes beeping while opening and closing doors, as well as a bell sound when leaving stops and triggering emergency braking. Through our recordings and interviews with the safety drivers we found that this sound is much louder on the inside of the shuttle than on the outside. During the course of the project, EasyMile added a manually triggered horn sound on request of the safety drivers.

[1]https://ridethefuture.se/in-english/

3.2 Safety Drivers

The legally required safety drivers are specifically trained professional drivers with an alternating schedule on manual and autonomous public transport vehicles. The six different drivers who we worked with all have several years of experience in driving buses or trams. We observed and video recorded them during their work, asking questions about specific events on the road and how they would use the horn on a bus or tram. One safety driver volunteered to work more closely with us in participatory sound design sessions. We presented project insights to all safety drivers.

3.3 Ethics

The safety drivers signed informed consent and video usage forms before we took rides with them. We did not collect any personal data from them since we were only interested in their professional roles. We always informed them of planned recordings a day in advance and practiced ongoing consent, asking whether they were still okay to be recorded on each specific recording instance.

For the other road users, obtaining written consent was not feasible. As recording in public is legal for research purposes in Sweden, we decided to tape large warning signs with camera symbols onto the bus whenever we recorded. When videotaping on the road, we put on yellow vests that said "research in progress" in large letters. Occasionally, people on the road would turn into passengers and hop onto the bus. Since they were not the focus of our work we informed them about the video recordings as soon as the doors of the bus opened, and asked whether it was okay to keep the cameras running. If they denied to be recorded we immediately switched off the recording equipment and erased any recordings that they were on. If they agreed to be recorded we provided them with more information about the study and our contact information, highlighting that they may contact us if they wanted their recordings removed.

4 APPROACH

In this work we integrate Ethnomethodology and Conversation Analysis with a Research through Design [31, 59, 102] approach to interaction design, contributing to what has been described as "technomethodology" [7, 14, 21]. While EMCA so far has hardly been intertwined with HRI design, it has been established in other fields, such as computer-supported cooperative work (CSCW) [14, 64, 77], and HCI [1, 15, 97], and has particular relevance for conversational user interfaces and dialogue systems [24, 72].

Ethnomethodology [28, 40] initially developed as a contrary approach to mainstream sociology, and is concerned with studying the methods by which people accomplish socially recognizable actions and activities in everyday interaction. Ethnomethodologists investigate the (often tacit) commonsense knowledge that members share about how to interact in specific settings and situations. Conversation Analysis [44, 85, 89] can be considered a subfield of ethnomethodology that has been further influenced

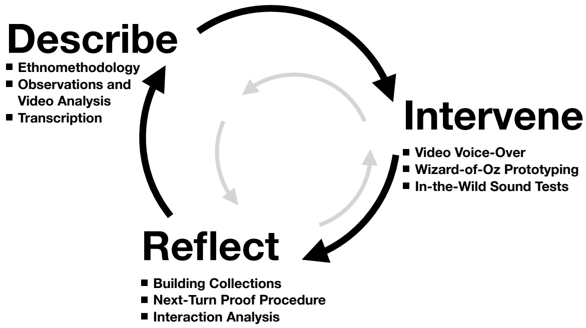

Figure 2: Design framework.

by fields such as anthropology and language philosophy. It focuses particularly on how participants accomplish interaction on a moment-by-moment basis, usually using video recordings as study material. While EMCA was initially concerned with human-human interaction, Lucy Suchman [91, 92] pioneered its introduction to system design in the 1980s in her studies of Xerox copying machines. Ethnomethodologists contributed studies of how people use technology, often at workplaces [38], and theoretically informed participatory design approaches [34], but found it difficult to move from design critique towards actively designing [14].

HRI work has taken inspiration from EMCA literature [32, 71, 98] and taken a conversation analytic stance in analyzing interaction with robots, however without systematically feeding these insights back into ongoing design iterations [32, 74]. Krummheuer et al. [48] have perhaps come closest to an iterative design cycle in a study in which the general design of a robot was supported by concurrent ethnomethodologic observations. Our work can be seen to build on and integrate previous efforts by design researchers working on system design at Stanford in the 1980s. Collaborating with Suchman, they drew on conversation analysis to study design processes, video recording each design iteration, analyzing it and informing the next iteration by this analysis [65, 94]. At this time, HCI researchers also proposed to use video recordings as design material for prototpying [39, 60], which has inspired our work.

Figure 2 represents an overview of our design process. In an iterative cycle, we first DESCRIBE interaction on the road and with our prototypes through observations, video recordings and transcription following EMCA practice [4, 44]. We then INTERVENE, using our video recordings as design material in voice-overs and testing sound prototypes in live traffic, again recording these sessions. Finally, we REFLECT on what actions the tested sounds achieved by building collections of similar cases and relating our findings to previous EMCA literature. Since this is a dynamic process, all stages may involve moving back and forth between description and intervention, intervention and reflection, or reflection and description. Following the Research through Design paradigm that is opportunistic in its nature, we do not want to claim that the events that we observe in this study happen statistically more frequently, but rather that they are significant moments, which are relevant for design. We present transcripts of events that can be seen as exemplary for what typically happens on the road. The next sections describe our approach in more detail.

4.1 Observations and Video Analysis

This fist stage is informed by an ethnomethodologic interest in *accountability*. This concept highlights that people are accountable for making their actions recognizable, for instance by doing something that can be recognized as doing walking rather than loitering, or doing driving rather than moving uncontrollably. If people fail to act in recognizable ways, others may request explanations (*accounts*) for why they did not act in the expected way. From this perspective we derived several guiding questions: Is the bus moving in recognizable ways? Are there moments when the bus moves contrary to human expectations? Are there moments in which the bus fails to explain its movement?

To answer these questions, the first author started observing and video recording, both on the road and as a passenger on the buses. We discovered early on that the buses brake repeatedly, hindering their progression from one stop to the next and ultimately delaying travel for the passengers. These stops typically occur when people get close to the bus, at moments when human drivers would not brake, at least not as much. (Auto)ethnographic notes and interviews with the safety drivers further revealed that these unnecessary brakes could potentially be dangerous, and the buses were modified in several ways to mitigate the impact of the braking on safety drivers and passengers. We went on to study these brakes in detail, by cropping videos snippets from our corpus that depicted the moments before a bus came to slow down or brake, and transcribing these videos (see section 4.7 and [73]).

4.2 Video Voice-Over

Taking inspiration from video prototyping [39, 60] and combining it with a sound design technique called vocal sketching [22], we turned the previously analyzed videos into our design material. We were interested in how the bus could provide situated explanations, then and there, of how to behave around it. Sound as a dynamically adjustable resource appeared particularly interesting, since a static visual display on the bus reading "I can brake hard" in Swedish clearly was not enough to prevent braking.

Playing our previously cropped videos (typically about 30 seconds) on loop and improvising with different sounds as a "voice-over" to the running videos enabled us to ground our design ideas in concrete cases from the beginning. We first used our voice and later pre-recorded sound snippets to explore when and how the bus could sound. These improvisation sessions allowed exploration of a broad range of sounds with immediate feedback on how they would sound in interaction and in the general soundscape of the road. The videos were especially important in working with timing of the sounds, as they highlighted how fast other traffic users move and how a sound may be relevant in one moment but become incomprehensible or mean something different in the next moment.

4.3 In-the-Wild Sound Testing

Following our focus on concrete interaction, we tested sounds in live traffic early on. Standing on the road, we played sounds from a phone, checking how they would sound and to what volume they needed to be set. This turned out to be very important, teaching us that while sounds may appear as too loud and intrusive in a lab or office setting, they may be barely noticeable on a busy road. For tests in live traffic we taped a waterproof Bluetooth speaker to the

Sound type	Sound description	Iteration
humming sound	low pitch hum	1
	high pitch hum	2
	motor sound	3
jingle	wheels on the bus	2,3
	reverse wheels on the bus	2
human-like	"ahem"	2
horns	high-pitch horn	1,3
	rising pitch sax roll	3
	beep button	4
bells	high-pitch bell	2,3
	repeated bell ring	3
	hand bell	4

Table 1: Overview of the sounds tested during each iteration.

front of the shuttle. While it may be beneficial to place speakers all around the bus, including the back, we wanted to keep the setup as simple as possible and followed the design of synthetic motor sound systems, which are usually placed in the front of cars. During the rides, we played sounds from a phone paired with the speaker.

4.4 Wizard-of-Oz Prototyping

We tested sounds during several Wizard-of-Oz [76] sessions, in which either the first author or a safety driver would act as wizard, triggering sounds. From outside the bus it was possible to recognize a passenger on the bus, but people could not immediately see that this was the wizard. The wizard would play sounds in moments when they felt relevant or necessary. Through joint discussion it emerged early on that sounds should only be played when someone was clearly in the same lane or otherwise close to the bus. The goal was to explore a broad range of sounds repeatedly during the rides. When the researcher acted as wizard, they would discuss and take into account what the safety driver said. When the safety driver acted as wizard, the researcher would watch and sometimes verbalize observations in a similar style as the safety driver had previously done (e.g. I can see that the cyclist is smiling). The researcher also sometimes asked the safety driver why they had played a sound at a specific moment and encouraged to reflect upon the design process in a think-aloud manner.

4.5 Sound Design

We explored a range of different sound types. Table 1 gives an overview of the 12 sounds tested on the road. Initially, the idea was to keep people away from the bus through humming sounds that would get louder as people got closer. We discovered that low pitch hums were difficult to hear on the road and high pitch hums were disliked by the safety driver. As an alternative, we explored jingles, the refrain of "the wheels on the bus go round and round" and its reversed melody. Following a different idea, we tested a range of short horn and bell sounds, which were easily repeatable. We also tested a human vocalization in the form of an "ahem" sound, inspired by the sound people may make when trying to pass on a blocked escalator, but did not find it particularly effective. In the fourth iteration, we switched from static sound files to sounds that could be dynamically controlled in the form of a beep button and a hand bell app. Some of the sounds are downloaded from the Soundsnap[2] database, others are modifications of these sounds created in Audacity[3]. The jingles are digital tunes created in Audacity. Please see supplementary material for details, including video recordings of all sounds on the road.

4.6 Data Collection

Over the course of 2.5 years, the first author collected 18 hours of video material from up to four camera perspectives. This includes campus traffic without (20 minutes) and with buses, recorded both from the road (40 minutes) and during rides with GoPros mounted to the bus (9 hours, 7 different occasions). During the latter, both road and inside of the bus were captured, documenting how people move around the shuttles and studying how safety drivers react to events on the road. The first author also video-recorded 4 iterations of in-the-wild sound tests (8 hours, during 26 rounds). In addition, she collected ethnographic notes on 18 occasions after walks on the road and attendance of the buses' inauguration events. Our work was complemented through several interviews: unstructured interviews with safety drivers during the rides, as well as semi-structured interviews with a sales director at EasyMile, a safety driver, and the project coordinator at the local transport provider. Ongoing work was presented at three workshops organized by the Ride the Future project, giving all stakeholders the opportunity to comment on and take inspiration from our work.

4.7 Transcription and Data Analysis

Following the conversation analytic approach, we treat video recordings as data and extracted video snippets from our corpus in which the bus came close to other road users, inductively building up collections of similar events. The first author transcribed clips that appeared interesting, to facilitate discussion and deeper analysis of how people respond to the bus on a moment-by-moment basis. This requires detailed transcription of verbal [46] and embodied [67] conduct, following EMCA transcription conventions.

Moving towards in-depth analysis, we focused on how people on the road display their understanding of the buses' sounds. From a conversation analytic perspective, the meaning of a current turn [85] in interaction can never be fixed in advance but others display their interpretation, what action they recognize it to be, in their next turn. This means, that by looking at how humans respond to a robot sound, we can access their understanding of it. This principle of looking at what happens next is summarized in the *next-turn proof procedure* [44, 85], which is often summarized in the analytic question "why that now?" [88] (see Figure 1).

5 FINDINGS

Taking an action-based approach, we organize our observations by how humans visibly interpret the sound in interaction. The focus is not on what the bus communicates about its inner status or intent but rather how humans react and adjust their actions in response.

[2] https://www.soundsnap.com
[3] https://www.audacityteam.org

5.1 Anticipation and Reaction

During our observations we noticed that the unplanned halts in between designated stops were typically caused by cyclists who were coming too close, entering the shuttle's safety bubble and thereby triggering braking or even harsh emergency stops. The video recordings revealed that safety drivers prepare for these stops well in advance.

```
01 DRI  ((gazes out of front window))
02 DRI  ska vi se hur han reagerar på de här
        let’s see how he reacts to these here
03      (.)
04 DRI  om [han]
        whether he
05 RES     [m  ] m
06 DRI  flyttar på sig
        moves
07 DRI  k(h)ehehe .h
08      (2.9)
09      +(0.2)
   dri  +moves left foot, puts weight on it, stretches right knee-->
10 DRI    h+är kommer bromsen
          here comes the brake
   dri  -->+
11      (1.5) ((decelerating sound from bus engine))
12      ((cyclists passing left side of bus, looking inside bus))
13 CYC  (oh* my god really xxxx)
   bus     *halted-->
14 DRI                °(han slår dit/det)°
                        he hits there/it
15      (0.2)
16 DRI  (h)m (h)m ((smiles))
17      (1.3)*
   bus    -->*continues on its route again-->>
18       (4.6) ((accelerating sound from bus engine))
19 DRI  det vart ingen tvärnit °i alla fall° ((smiles))
        it was no slamming of the brakes at least
20      (0.2)
21 RES  mhm
```

Figure 3: EM-2020-04-22-round3. Anticipating a stop. DRI= safety driver, RES= researcher, BUS= EZ10 shuttle, CYC= cyclist. See supplementary material for transcription symbols, image, and video clip.

The extract in Figure 3 illustrates how a safety driver (DRI) anticipates such an upcoming stop: He gazes out of the front window and announces that there is an interesting situation ahead (lines 01-02), a typical moment when the wizard would have triggered a sound. In this extract, there are just the EZ10's original sounds. The safety driver adds that he and the researcher (RES) will soon see whether the bus will (continue to) move (l. 04-06) and the bus first keeps moving forward with the same speed. A moment later, the safety driver slightly changes his position (l. 09), shifting his weight on his back leg. He then announces that the brake is kicking in (l. 10), which can also be heard by the decelerating motor sound (l. 11). As the bus comes to a halt, the cyclists (CYC) pass on its side (l. 12-13). The bus starts accelerating again after they have moved away (l. 18) and the safety driver evaluates the stop as "no slamming of the brakes at least" (l. 19).

This extract demonstrates that while the safety driver anticipates a potentially problematic moment (and could prevent it on a manually operated bus), the autonomous bus is approaching the cyclists without announcing any potential trouble, leaving no opportunity for cyclists and passengers to prepare for the brake. Even though the buses emit sound through the rustling of the plastic parts and a synthetic motor sound system, problems with unnecessary or unexpected braking persisted throughout the project. The manufacturers slightly adjusted the braking behavior and safety belts were added to prevent harm to passengers, but this did not seem to change the problem: when surrounded by many people or when passing narrow passages, the shuttles will stop.

Initially, we explored synthetic humming and motor sounds as a way to prevent cyclists and pedestrians from coming closer to the bus. However, it was difficult to find suitable sounds with our voice-over approach. When testing some designs, the safety driver argued that they were confusing and hard to hear on the noisy road, where buildings and other vehicles also contribute low-pitch humming sounds. A high-pitched sound inspired by mosquitoes was strongly opposed by the safety driver who was worried that it would scare away passengers. Interestingly, from an autoethnographic perspective, while these sounds appeared suitable in an office, they took more courage to play live on the road.

5.2 Explaining Presence through Jingles

Learning that subtle humming was not suitable, we took inspiration from local ice cream trucks, which announce their arrival with a recognizable tune. Testing electronic jingles, we made rather unexpected observations.

```
01      +(0.6)#
   dri  +gazes at road-->
   img        #img1
02 WIZ  $$(0.1)+(0.3)+da d|e da da de+# de d|a da de+# da da do|
   wiz  $$((button pressed))
   dri     -->+.....+gazes at RES---+gazes at cyc--+gazes at road-->l.05
   cyc                 |steers away------|passes bus--------|
   img                             #img2          #img3b
03      da de da da de de da da de *da da do#
   cycL                            *moves to sidewalk-->
   img                                      #img3a
04      (1.4)*
   cycL   -->*
05 WIZ  $$(0.4)da de da+ da de de da %da# &+%de# da&+ da [do     ]
06 CYC                                                   [gulligt]
                                                          cute
   wiz  $$((button pressed))
   dri            -->+follows cyclist group with gaze-->>
   cycM                              %waves-% ((while driving onto sidewalk))
   cycR                                   &waves---&
   dri                                     +waves---+
   img                                 #img4  #img5ab
07      da de da [da# de de d   ]a da de da da do
08 DRI           [gulligt sa hon]
                  cute, she said
   img              #img6
```

Figure 4: EM-2020-05-29. Wheels on the bus jingle. DRI= safety driver, WIZ= researcher acting as Wizard-of-Oz, CYC= cyclists. See supplementary material for transcription symbols, images, and video clip.

The extract in Figure 4 shows how cyclists move away from the bus well in time while the refrain of the song "the wheels on the bus" is played: A cyclist (cyc) emerges in the "tracks" of the bus and the researcher who acts as the wizard WIZ in this case soon triggers the jingle (l. 02), which plays for 12 seconds. Soon after the jingle starts playing, the cyclist starts turning away from the bus and subsequently passes it without any problems (l. 02). The safety driver, who had been following the cyclist with his gaze is now looking back on the road ahead (l. 02), where a group of three cyclists is approaching. The jingle is still playing, as the bus continues to move forward. While still at several bike-lengths distance, the leftmost cyclist (cycL) starts to move towards an area designated as a sidewalk, away from the bus (l. 03). The jingle terminates but is soon triggered again by the researcher (l. 04-05). As the jingle goes on, the group moves closer to the bus. The middle cyclist (cycM) now also moves towards the sidewalk area, waving at the bus (l. 05). Shortly after, the rightmost cyclist (cycR) also starts waving (l. 05). The safety driver (DRI) smiles and raises his hand, responding to their greeting (l. 05). As the cyclists are passing the bus, one of them says *gulligt*, "cute" in Swedish (l. 06), which the safety driver repeats to the researcher, saying "she said cute" (l. 08).

While the extract may appear as unusual, it is exemplary of the responses to the jingle that we observed in several ways: First, we saw people move away from the bus in response to the jingle, before any potential brakes became relevant. Cyclists approaching the bus in narrow passages did not seem to get in the way as much, acknowledging and reacting to the bus well in advance. Second, since the jingle continues for several seconds, it can be responded to by several different groups that the bus passes during this time. This makes it particularly suitable in dense environments, where addressing individuals is not practically feasible. Third, the "happy" tune seems to evoke positive associations with the bus. We observed people gazing towards the bus, smiling as they passed. We captured a range of greeting/welcoming actions towards the bus, such as a child getting off their bike and dancing on the sidewalk, and a mother with child stopping the bus, curiously asking where it was going, which according to the safety driver never happened before. By announcing its presence loudly, the bus may evoke associations with vehicles that are known to be moving slowly, allowing cyclists and pedestrians to adjust to its limited coordination abilities. At the same time, the friendly jingle seems to invite interaction, possibly attracting potential passengers.

5.3 Instructing Movement with Bells and Horns

In contrast to the jingles, we explored a variety of horn and bell sounds. A major benefit of such "beep" and "ding" sounds is that they are short and can be flexibly repeated.

```
01 WIZ  dingling [dingling::        ding::]
02 RES           [det är kanske för snällt] ((ending previous discussion))
                  maybe it is too friendly
03 DRI  [#nu plingar vi]
         now we are plinging
04 WIZ  [#ding:  dingli]ng::  %ding +dingling:::*:
   mip                        %...............walks away with phone--> 1.08
   lip                              +..........takes step-->
   rip                                          *...-->
   img   #img1
05 WIZ  #di*ngling ling:*:+::
   rip  -->*turns to bus*shifts weight-->
   lip                -->+step-->
   img  #img2
06 WIZ ding*::::::*+
   rip  -->*step--*
   lip          -->+
07 WIZ   #*+ding ding*: dingling di#ng+ di*ng
   rip    *step------*step----------------*
   lip  -->+step----------------------+step-->
   img   #img3                    #img4
08 WIZ  dingling: dingling+%: dingling*:::::::*+%
   lip                -->+step---------------+
   mip                 -->%turns towards bus---%
   rip            ((2 invisible steps))*step---*
09 WIZ  *+dingding:%*: #di+ng dingling:*:%: lingding: dingling:: di%ngling::+
   rip  *step-------*small step--------* ((then remains still))
   lip   +step-----------+((2 invisible steps))---------------------------+
   mip             %step-----------------%step--------------------%
   img                 #img5
```

Figure 5: EM-2022-09-09. Repeated bell sounds. DRI= safety driver, WIZ= safety driver acting as Wizard-of-Oz, lip= left pedestrian, mip= middle pedestrian, rip= right pedestrian. See supplementary material for transcription symbols, images, and video clip.

The extract in Figure 5 shows how the safety driver acts as wizard (WIZ), using a bell sound to instruct a group of three pedestrians to move out of the way: He starts playing "dingling" sounds as the bus begins to speed towards the group (l. 01), and simultaneously announces "now we are plinging (l. 03), thereby effectively ending a previous conversation with the researcher and marking the start of another trial. As the wizard repeats "ding dingling" (l. 04), a first reaction can be observed. The left pedestrian (lip), who is facing the bus, starts taking a small step. The right pedestrian (rip) whose back is facing the bus, starts turning. The middle pedestrian (mip) is walking away, possibly taking a phone call. The safety driver adds a "dingling ling" (l. 05), during which the right pedestrian fully turns towards the bus. As the sound lingers, the left pedestrian takes another, larger step, and the right pedestrian shifts his weight, taking a further step during another "ding" (l. 06) from the safety driver. The group has acknowledged the bus, but they need to move further for the bus to be able to pass. The safety driver initiates more sounds, this time in a fast, rhythmical sequence "ding ding dingling ding ding dingling dingling dingling" (l. 07-08), which can be heard as an upgraded version of the earlier bell sounds. The left and right pedestrian now take several steps, moving away from the bus. Towards the end of the sequence, even the middle pedestrian turns towards the bus again (l. 08). Both left and middle pedestrian are now right in front of the bus, and the safety driver initiates another upgraded round of "dingding ding dingling lingding dingling dingling" (l. 09), resulting in all three pedestrians taking further steps, until they have effectively moved out of the way.

The extract highlights how pedestrians react to repeated bell sounds in a finely coordinated manner, responding to the first "dingling" sequence by looking at the bus, then taking first moves as the sounds are repeated, and finally stepping away as the sounds are played with increased urgency. The sounds in this extract are generated with an app that mimics a hand bell by translating shaking movements of the phone into sound. It was most intuitive to use and preferred by the safety driver, as it gave immediate feedback without requiring gaze at the phone. This is example demonstrates the fine level of detail at which the safety driver tunes the sounds to the movement to the pedestrians, and how they in turn react to each bell ring as an indicator that they still have not done enough, incrementally moving as the sounds are repeated and intensified.

While some horns or bells certainly suit the "character" of the bus better than others, we observed that whether we used a high-pitch bell or horn, a rising saxophone riff or a hand bell, reactions were similar: Repeated when people were still visibly in the way of the bus, such sounds were responded to as instructions to further adjust one's trajectory. By repeating the sounds, we successfully instructed a moving pedestrian to wait at a crossing, cyclists to swerve, and stationary pedestrians to move out of the way.

6 DESIGN IMPLICATIONS

Our study can teach three main lessons about the design of sound for autonomous buses, and robots more generally. First, we highlight that sound should be designed in and for sequential contexts. Second, we discuss the type of insights that can be gained from designing such sound-in-interaction, and finally we discuss our approach of intertwining EMCA and interaction design.

6.1 Designing Sound in Sequential Contexts

Contributing to research on sound in HRI, our work demonstrates an approach to designing sound in concrete interactions. While current work focuses on validating sounds in isolation [10, 79, 82], we have demonstrated that it may be equally important to design for when a sound should be used in interaction and what meaning it gains in specific sequential contexts. Focusing on people's reactions in live traffic, we found that sounds of different pitch and timbre could accomplish the same action, effectively instructing others to adjust their movement to the bus. Similarly, depending on context, the same sound may accomplish different things, instructing one person to stop, and another one to move. Building on the conversation analytic concept of *sequence* [87], we would like to highlight that context does not only include the situation on the road, such as whether the sound is played at an intersection or when leaving a stop. Rather, the sequential context is defined by the movement of the bus and people on the road and therefore keeps changing, as people are moving or stopping. To design interaction rather than robot interfaces [2], we need to focus on designing sound-in-interaction, looking at the role of sound in concrete sequences.

6.2 Sound-In-Interaction

In section 5, we provided examples of how sound is interpreted on the road. Below we reflect on what these concrete examples can teach us about sound design more generally.

6.2.1 The Ambiguity of Humming Sounds. While synthetic motor sounds may make autonomous cars safer by announcing their presence [68], we found that they are not sufficient for instructing others to keep a distance from the shuttle buses. As we illustrated in the extract in Figure 3, unnecessary brakes are anticipated by safety drivers, who recognize that coordination becomes relevant. The braking of the bus can then be regarded as failure to act in recognizable ways [33], which would allow others to adjust their movement [16, 37]. The autonomous shuttle buses are more restricted in their movement than regular cars, and may require different means for asking for other traffic participants' support.

6.2.2 Jingles as Accounts. Playing upbeat jingles from the bus as in the extract in Figure 4 drew people's gaze towards the bus, and typically they also adjusted their trajectories. From the EMCA perspective, the jingles may be seen as accounts, or situated explanations [20, 29] for what the bus is doing: In the Swedish context they evoke associations with the ice cream trucks that drive around many neighborhoods. Reminding of another slow moving vehicle, the jingles may thereby serve as instruction for how to treat the relatively novel vehicle on the road. A different example for this would be ambulances and police cars, which use their continuous horn sounds to instruct everyone else to give way. Rather than distinguishing implicit from explicit interfaces [19, 69], one may ask what recognizable actions they can perform to instruct interaction with the new vehicle on the road.

6.2.3 Repetition Initiates Repair. Bell and horn sounds can turn into powerful instructions, as we demonstrated in the extract in Figure 5. Depending on the situated context, repetitions may flexibly instruct both moving out of the way or stopping and waiting until the bus has passed. In conversation analysis, repetition is known as a common strategy for *repair*. This term describes how people deal with problems in mutual understanding, also in interaction with machines [91]. By repeating a sound, the bus provides a new opportunity for the people around it to respond in a different way - until mutual understanding is reached. The repetition of a sound can thus be heard as highlighting that the current response is not sufficient and further adjustment is necessary.

6.3 Intertwining EMCA and HRI Design

We also make a methodological contribution by introducing earlier work in HCI and CSCW [7, 60, 65, 91, 94] to HRI. Our approach tightly integrates ethnomethodology and multimodal conversation analysis into the the HRI interaction design process (see Figure 2). We have developed novel design techniques that put focus on what actions sound can accomplish, by using video recordings of actual interaction with the robot as prototyping material. This has been crucial for us to explore what types of sound would be appropriate in the setting, and to test their timing and duration before deploying them on the road. Testing sounds in live traffic, with professional drivers, has taught us important practical lessons about the salience of sound in noisy settings, and differences in how they sound inside and outside the bus. While our transcripts describe specific situations, engaging in EMCA transcription enables designers to reflect on their designs. This opens opportunities for generalization, by putting focus on the types of actions that the sounds can accomplish in interaction. Our video recordings and transcripts can be seen as a contribution to intermediate-level knowledge [42, 57], in between concrete cases and theoretical abstractions. While grounded in specific interactions [53] they allow to describe general practices of how sound is used and interpreted on the road.

7 CONCLUSION

We reported on a two-year Research through Design study in which we explored how public transport autonomous shuttle buses could use sound to communicate with other road users. Focusing on what actions sound can accomplish during tests in live traffic, we demonstrated first that (motor) hums are not always sufficient for instructing movement around the bus. We then showed how jingles invite interaction by providing situated explanations of the buses' presence and how repeated short honks or beeps can instruct others' movement. The paper presents a novel method for the design of robot sound, which tightly intertwines ethnomethodology and conversation analysis with interaction design and highlights the importance of designing sound in interactional sequences.

ACKNOWLEDGMENTS

This work is funded by the Swedish Research Council (2016-00827) and the National Science Foundation (IIS-1942085). Several people deserve our thanks: the safety drivers and Ride the Future for generous access to the buses, cognitive science students Johanna Schützer, Hanna Henriksson, Tobias Augustsson, Anna Follin, Rosanna Lilliesköld, Annika Lundberg, and Tobias Ryrberg for help with data collection and analysis, Niklas Rönnberg for providing insights on sonification, attendees at the SiS, MOBSIN and AI in Interaction data sessions and the DMCA conference for fruitful discussion. We thank Mathias Broth, Leelo Keevallik, and Sam Thellman, as well as the anonymous reviewers for their invaluable feedback.

REFERENCES

- [1] Paul M. Aoki, Matthew Romaine, Margaret H. Szymanski, James D. Thornton, Daniel Wilson, and Allison Woodruff. 2003. The Mad Hatter's Cocktail Party: A Social Mobile Audio Space Supporting Multiple Simultaneous Conversations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 425–432. https://doi.org/10.1145/642611.642686
- [2] Michel Beaudouin-Lafon. 2004. Designing Interaction, Not Interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '04)*. Association for Computing Machinery, New York, NY, USA, 15–22. https://doi.org/10.1145/989863.989865
- [3] Mathias Broth, Jakob Cromdal, and Lena Levin. 2018. Showing where you're going. Instructing the accountable use of the indicator in live traffic. *International Journal of Applied Linguistics* 28, 2 (2018), 248–264. https://doi.org/10.1111/ijal.12194
- [4] Mathias Broth and Leelo Keevallik (Eds.). 2020. *Multimodal Interaktionsanalys*. Studentlitteratur, Lund, Sweden. 464 pages.
- [5] Barry Brown. 2017. The Social Life of Autonomous Cars. *Computer* 50, 2 (2017), 92–96. https://doi.org/10.1109/MC.2017.59
- [6] Barry Brown and Eric Laurier. 2017. *The Trouble with Autopilots: Assisted and Autonomous Driving on the Social Road*. Association for Computing Machinery, New York, NY, USA, 416–429. https://doi.org/10.1145/3025453.3025462
- [7] Graham Button and Paul Dourish. 1996. Technomethodology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '96)*. ACM, Vancouver, British Columbia, Canada, 19–26. https://doi.org/10.1145/238386.238394
- [8] G. Button, M. Lynch, and W. Sharrock. 2022. *Ethnomethodology, Conversation Analysis and Constructive Analysis: On Formal Structures of Practical Action*. Taylor & Francis.
- [9] Elizabeth Cha, Naomi T Fitter, Yunkyung Kim, Terrence Fong, and Maja J Matarić. 2018. Effects of Robot Sound on Auditory Localization in Human-Robot Collaboration. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. Association for Computing Machinery, New York, NY, USA, 434–442. https://doi.org/10.1145/3171221.3171285
- [10] Lilian Chan, Brian J Zhang, and Naomi T Fitter. 2021. Designing and Validating Expressive Cozmo Behaviors for Accurately Conveying Emotions. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 1037–1044. https://doi.org/10.1109/RO-MAN50785.2021.9515425
- [11] Herbert H Clark. 1996. *Using Language*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511620539
- [12] Steven E Clayman and Douglas W Maynard. 1995. Ethnomethodology and conversation analysis. In *Situated Order: Studies in the Social Organization of Talk and Embodied Activities*, Paul ten Have and George Psathas (Eds.). University Press of America, Washington, D.C., 1–30.
- [13] Mark Colley, Elvedin Bajrovic, and Enrico Rukzio. 2022. Effects of Pedestrian Behavior, Time Pressure, and Repeated Exposure on Crossing Decisions in Front of Automated Vehicles Equipped with External Communication. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 367, 11 pages. https://doi.org/10.1145/3491102.3517571
- [14] Andy Crabtree. 2004. Taking technomethodology seriously: hybrid change in the ethnomethodology–design relationship. *European Journal of Information Systems* 13 (9 2004), 195–209. Issue 3. https://doi.org/10.1057/palgrave.ejis.3000500
- [15] Andrew Crabtree, Tom Rodden, Peter Tolmie, and Graham Button. 2009. Ethnography Considered Harmful. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 879–888. https://doi.org/10.1145/1518701.1518835
- [16] Arnulf Deppermann, Eric Laurier, Lorenza Mondada, Mathias Broth, Jakob Cromdal, Elwys De Stefani, Pentti Haddington, Lena Levin, Maurice Nevile, and Mirka Rauniomaa. 2018. Overtaking as an interactional achievement : video analyses of participants' practices in traffic. *Gesprächsforschung* 19 (2018), 1–131.
- [17] Debargha Dey, Azra Habibovic, Bastian Pfleging, Marieke Martens, and Jacques Terken. 2020. *Color and Animation Preferences for a Light Band EHMI in Interactions Between Automated Vehicles and Pedestrians*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi-org.e.bibl.liu.se/10.1145/3313831.3376325
- [18] Debargha Dey, Kai Holländer, Melanie Berger, Berry Eggen, Marieke Martens, Bastian Pfleging, and Jacques Terken. 2020. *Distance-Dependent EHMIs for the Interaction Between Automated Vehicles and Pedestrians*. Association for Computing Machinery, New York, NY, USA, 192–204. https://doi-org.e.bibl.liu.se/10.1145/3409120.3410642
- [19] Debargha Dey and Jacques Terken. 2017. Pedestrian Interaction with Vehicles: Roles of Explicit and Implicit Communication. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '17)*. Association for Computing Machinery, New York, NY, USA, 109–113. https://doi.org/10.1145/3122986.3123009
- [20] Paul Dourish. 2001. *Where the Action is: The Foundations of Embodied Interaction*. MIT Press, Cambridge, MA, USA.
- [21] Paul Dourish and Graham Button. 1998. On "Technomethodology": Foundational Relationships Between Ethnomethodology and System Design. *Human–Computer Interaction* 13, 4 (1998), 395–432. https://doi.org/10.1207/s15327051hci1304_2
- [22] Inger Ekman and Michal Rinott. 2010. Using Vocal Sketching for Designing Sonic Interactions. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems (DIS '10)*. Association for Computing Machinery, New York, NY, USA, 123–131. https://doi.org/10.1145/1858171.1858195
- [23] European Commission. 2019. Electric and hybrid cars: new rules on noise emitting to protect vulnerable road users. https://single-market-economy.ec.europa.eu/select-language?destination=/node/1136
- [24] Joel E. Fischer, Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. Progressivity for Voice Interface Design. In *Proceedings of the 1st International Conference on Conversational User Interfaces* (Dublin, Ireland) *(CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 26, 8 pages. https://doi.org/10.1145/3342775.3342788
- [25] Kerstin Fischer, Lars Christian Jensen, and Leon Bodenhagen. 2014. To Beep or Not to Beep Is Not the Whole Question. In *Social Robotics. ICSR 2014. Lecture Notes in Computer Science, vol 8755*, M. Beetz, B. Johnston, and MA Williams (Eds.). Springer, Cham, 156–165. https://doi.org/10.1007/978-3-319-11973-1_16
- [26] Emma Frid and Roberto Bresin. 2022. Perceptual Evaluation of Blended Sonification of Mechanical Robot Sounds Produced by Emotionally Expressive Gestures: Augmenting Consequential Sounds to Improve Non-verbal Robot Communication. *International Journal of Social Robotics* 14, 2 (2022), 357–372. https://doi.org/10.1007/s12369-021-00788-4
- [27] Mafalda Gamboa, Mohammad Obaid, and Sara Ljungblad. 2021. Ritual Drones: Designing and Studying Critical Flying Companions. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 562–564. https://doi.org/10.1145/3434074.3446363
- [28] Harold Garfinkel. 1967. *Studies in Ethnomethodology*. Prentice-Hall, Englewood Cliffs NJ.
- [29] Harold Garfinkel (Ed.). 1986. *Ethnomethodological studies of work*. Routledge & K. Paul, London.
- [30] Harold Garfinkel and Harvey Sacks. 1970. On formal structures of practical actions. In *Theoretical Sociology: Perspectives and Development*, John C. McKinney and Edward A. Tiryakian (Eds.). Appleton-Century-Crofts, New York, 337–366.
- [31] William Gaver. 2012. What Should We Expect from Research through Design?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 937–946. https://doi.org/10.1145/2207676.2208538
- [32] Raphaela Gehle, Karola Pitsch, Timo Dankert, and Sebastian Wrede. 2017. How to Open an Interaction Between Robot and Museum Visitor? Strategies to Establish a Focused Encounter in HRI. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. Association for Computing Machinery, New York, NY, USA, 187–195. https://doi.org/10.1145/2909824.3020219
- [33] Erving Goffman. 1963. *Behavior in Public Places*. The Free Press, New York.
- [34] Joan Greenbaum and Morten Kyng (Eds.). 1991. *Design at Work : Cooperative Design of Computer Systems*. CRC Press.
- [35] Azra Habibovic, Victor Malmsten Lundgren, Jonas Andersson, Maria Klingegård, Tobias Lagström, Anna Sirkka, Johan Fagerlönn, Claes Edgren, Rikard Fredriksson, Stas Krupenia, Dennis Saluäär, and Pontus Larsson. 2018. Communicating Intent of Automated Vehicles to Pedestrians. *Frontiers in Psychology* 9 (2018), 1336. https://doi.org/10.3389/fpsyg.2018.01336
- [36] Pentti Haddington, Lorenza Mondada, and Maurice Nevile (Eds.). 2013. *Interaction and Mobility*. DE GRUYTER. https://doi.org/10.1515/9783110291278
- [37] Pentti Haddington and Mirka Rauniomaa. 2014. Interaction Between Road Users: Offering Space in Traffic. *Space and Culture* 17, 2 (may 2014), 176–190. https://doi.org/10.1177/1206331213508498
- [38] Christian Heath and Paul Luff. 1992. Collaboration and controlCrisis management and multimedia technology in London Underground Line Control Rooms. *Computer Supported Cooperative Work (CSCW)* 1, 1-2 (mar 1992), 69–94. https://doi.org/10.1007/BF00752451
- [39] Austin Henderson. 1989. Video and Design. *SIGCHI Bull.* 21, 2 (oct 1989), 104–107. https://doi.org/10.1145/70609.70627
- [40] John Heritage. 1984. *Garfinkel and Ethnomethodology*. Polity Press, Cambridge [Cambridgeshire], New York, N.Y.
- [41] Marius Hoggenmüller, Wen-Ying Lee, Luke Hespanhol, Malte Jung, and Martin Tomitsch. 2021. Eliciting New Perspectives in RtD Studies through Annotated Portfolios: A Case Study of Robotic Artefacts. In *Designing Interactive Systems Conference 2021* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1875–1886. https://doi.org/10.1145/3461778.3462134

[42] Kristina Höök and Jonas Löwgren. 2012. Strong Concepts: Intermediate-Level Knowledge in Interaction Design Research. *ACM Transactions on Computer-Human Interaction* 19, 3 (oct 2012), 1–18. https://doi.org/10.1145/2362364.2362371

[43] Ming Hou, Karthik Mahadevan, Sowmya Somanath, Ehud Sharlin, and Lora Oehlberg. 2020. Autonomous Vehicle-Cyclist Interaction: Peril and Promise. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376884

[44] Ian Hutchby and Robin Wooffitt. 2008. *Conversation Analysis* (sse ed.). Polity Press, Cambridge, UK.

[45] Eun-Sook Jee, Yong-Jeon Jeong, Chong Hui Kim, and Hisato Kobayashi. 2010. Sound design for emotion and intention expression of socially interactive robots. *Intelligent Service Robotics* 3, 3 (2010), 199–206. https://doi.org/10.1007/s11370-010-0070-7

[46] Gail Jefferson. 2004. Glossary of transcript symbols with an introduction. In *Conversation Analysis: Studies from the first generation*, Gene H Lerner (Ed.). John Benjamins, Amsterdam, 13–31. https://doi.org/10.1075/pbns.125.02jef

[47] Gunnar Johannsen. 2001. Auditory Displays in Human–Machine Interfaces of Mobile Robots for Non-Speech Communication with Humans. *Journal of Intelligent and Robotic Systems* 32, 2 (2001), 161–169. https://doi.org/10.1023/A:1013953213049

[48] Antonia Lina Krummheuer, Matthias Rehm, and Kasper Rodil. 2020. Triadic Human-Robot Interaction. Distributed Agency and Memory in Robot Assisted Interactions. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*. ACM, New York, NY, USA, 317–319. https://doi.org/10.1145/3371382.3378269

[49] Tobias Lagström and Victor Malmsten Lundgren. 2015. *AVIP - Autonomous Vehicles' Interaction with Pedestrians: An Investigation of Pedestrian-Driver Communication and Development of a Vehicle External Interface.* Ph. D. Dissertation. Chalmers University of Technology.

[50] Adrian Benigno Latupeirissa, Emma Frid, and Roberto Bresin. 2019. Sonic characteristics of robots in films. In *Proceedings of the Sound and Music Computing Conference (SMC)*. 1–6. http://smc2019.uma.es/articles/P2/P2{_}07{_}SMC2019{_}paper.pdf

[51] Adrian B. Latupeirissa, Claudio Panariello, and Roberto Bresin. 2020. Exploring emotion perception in sonic HRI. In *Proceedings of the 17th Sound and Music Computing Conference*. Zenodo. https://doi.org/10.5281/zenodo.3898927

[52] Eric Laurier, Daniel Muñoz, Rebekah Miller, and Barry Brown. 2020. A Bip, a Beeeep, and a Beep Beep: How Horns Are Sounded in Chennai Traffic. *Research on Language and Social Interaction* 53, 3 (jul 2020), 341–356. https://doi.org/10.1080/08351813.2020.1785775

[53] Hee Rin Lee, EunJeong Cheon, Chaeyun Lim, and Kerstin Fischer. 2022. Configuring Humans: What Roles Humans Play in HRI Research. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, 478–492.

[54] Wen-Ying Lee and Malte Jung. 2020. Ludic-HRI: Designing Playful Experiences with Robots. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 582–584. https://doi.org/10.1145/3371382.3377429

[55] Jamy Li, Rebecca Currano, David Sirkin, David Goedicke, Hamish Tennent, Aaron Levine, Vanessa Evers, and Wendy Ju. 2020. On-Road and Online Studies to Investigate Beliefs and Behaviors of Netherlands, US and Mexico Pedestrians Encountering Hidden-Driver Vehicles. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*. Association for Computing Machinery, New York, NY, USA, 141–149. https://doi.org/10.1145/3319502.3374790

[56] Per Linell. 2009. *Rethinking Language, Mind, and World Dialogically*. Information Age Publishing, Charlotte, NC.

[57] Maria Luce Lupetti, Cristina Zaga, and Nazli Cila. 2021. Designerly Ways of Knowing in HRI: Broadening the Scope of Design-Oriented HRI Through the Concept of Intermediate-Level Knowledge. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 389–398. https://doi.org/10.1145/3434073.3444668

[58] Michal Luria, Marius Hoggenmüller, Wen-Ying Lee, Luke Hespanhol, Malte Jung, and Jodi Forlizzi. 2021. Research through Design Approaches in Human-Robot Interaction. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 685–687. https://doi.org/10.1145/3434074.3444868

[59] Michal Luria, John Zimmerman, and Jodi Forlizzi. 2019. Championing Research Through Design in HRI. https://doi.org/10.48550/ARXIV.1908.07572

[60] Wendy E Mackay, Anne V Ratzer, and Paul Janecek. 2000. Video Artifacts for Design: Bridging the Gap between Abstraction and Detail. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '00)*. Association for Computing Machinery, New York, NY, USA, 72–82. https://doi.org/10.1145/347642.347666

[61] Karthik Mahadevan, Sowmya Somanath, and Ehud Sharlin. 2018. Communicating Awareness and Intent in Autonomous Vehicle-Pedestrian Interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174003

[62] Paul McIlvenny. 2014. Vélomobile formations-in-action: Biking and talking together. *Space and Culture* 17, 2 (may 2014), 137–156. https://doi.org/10.1177/1206331213508494

[63] Paul McIlvenny, Mathias Broth, and Pentti Haddington. 2014. Moving Together. *Space and Culture* 17, 2 (may 2014), 104–106. https://doi.org/10.1177/1206331213508679

[64] Donald McMillan, Barry Brown, Ikkaku Kawaguchi, Razan Jaber, Jordi Solsona Belenguer, and Hideaki Kuzuoka. 2019. Designing with Gaze: Tama – a Gaze Activated Smart-Speaker. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (nov 2019), 1–26. https://doi.org/10.1145/3359278

[65] Scott L Minneman. 1991. *The social construction of a technical reality: Empirical studies of group engineering design practice.* Ph. D. Dissertation. Stanford University.

[66] Nicole Mirnig, Nicole Perterer, Gerald Stollnberger, and Manfred Tscheligi. 2017. Three Strategies for Autonomous Car-to-Pedestrian Communication: A Survival Guide. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. Association for Computing Machinery, New York, NY, USA, 209–210. https://doi.org/10.1145/3029798.3038402

[67] Lorenza Mondada. 2018. Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality. *Research on Language and Social Interaction* 51, 1 (2018), 85–106. https://doi.org/10.1080/08351813.2018.1413878

[68] Dylan Moore, Rebecca Currano, and David Sirkin. 2020. Sound Decisions: How Synthetic Motor Sounds Improve Autonomous Vehicle-Pedestrian Interactions. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '20)*. Association for Computing Machinery, New York, NY, USA, 94–103. https://doi.org/10.1145/3409120.3410667

[69] Dylan Moore, Rebecca Currano, G Ella Strack, and David Sirkin. 2019. The Case for Implicit External Human-Machine Interfaces for Autonomous Vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 295–307. https://doi.org/10.1145/3342197.3345320

[70] Dylan Moore, Hamish Tennent, Nikolas Martelaro, and Wendy Ju. 2017. Making Noise Intentional: A Study of Servo Sound Perception. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. Association for Computing Machinery, New York, NY, USA, 12–21. https://doi.org/10.1145/2909824.3020238

[71] Bilge Mutlu, Takayuki Kanda, Jodi Forlizzi, Jessica Hodgins, and Hiroshi Ishiguro. 2012. Conversational Gaze Mechanisms for Humanlike Robots. *ACM Trans. Interact. Intell. Syst.* 1, 2 (jan 2012), 12:1—-12:33. https://doi.org/10.1145/2070719.2070725

[72] Christiane Opfermann, Karola Pitsch, Ramin Yaghoubzadeh, and Stefan Kopp. 2017. The Communicative Activity of "Making Suggestions" as an Interactional Process: Towards a Dialog Model for HAI. In *Proceedings of the 5th International Conference on Human Agent Interaction* (Bielefeld, Germany) *(HAI '17)*. Association for Computing Machinery, New York, NY, USA, 161–170. https://doi.org/10.1145/3125739.3125752

[73] Hannah R.M. Pelikan. 2021. Why Autonomous Driving Is So Hard: The Social Dimension of Traffic. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction.* Linköping University, Faculty of Arts and Sciences, ACM, New York, NY, USA, 81–85. https://doi.org/10.1145/3434074.3447133

[74] Hannah R M Pelikan, Mathias Broth, and Leelo Keevallik. 2020. "Are You Sad, Cozmo?": How Humans Make Sense of a Home Robot's Emotion Displays. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*. Association for Computing Machinery, New York, NY, USA, 461–470. https://doi.org/10.1145/3319502.3374814

[75] Karola Pitsch, Hideaki Kuzuoka, Yuya Suzuki, Luise Sussenbach, Paul Luff, and Christian Heath. 2009. The first five seconds: Contingent stepwise entry into an interaction as a means to secure sustained engagement in HRI. In *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication.* IEEE, 985–991. https://doi.org/10.1109/ROMAN.2009.5326167

[76] Martin Porcheron, Joel E. Fischer, and Stuart Reeves. 2021. Pulling Back the Curtain on the Wizards of Oz. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 243 (jan 2021), 22 pages. https://doi.org/10.1145/3432942

[77] David Randall, Mark Rouncefield, and Peter Tolmie. 2021. Ethnography, CSCW and Ethnomethodology. *Computer Supported Cooperative Work (CSCW)* 30 (2021), 189–214. Issue 2. https://doi.org/10.1007/s10606-020-09388-8

[78] Robin Read and Tony Belpaeme. 2014. Situational Context Directs How People Affectively Interpret Robotic Non-Linguistic Utterances. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction (HRI '14)*.

Association for Computing Machinery, New York, NY, USA, 41–48. https://doi.org/10.1145/2559636.2559680

[79] Robin Read and Tony Belpaeme. 2016. People Interpret Robotic Non-linguistic Utterances Categorically. *International Journal of Social Robotics* 8, 1 (2016), 31–50. https://doi.org/10.1007/s12369-015-0304-0

[80] Malte Risto, Colleen Emmenegger, Erik Vinkhuyzen, Melissa Cefkin, and Jim Hollan. 2017. Human-Vehicle Interfaces: The Power of Vehicle Movement Gestures in Human Road User Coordination. *Proceedings of the 9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design: driving assessment 2017*, 186–192. https://doi.org/10.17077/drivingassessment.1633

[81] Frederic Anthony Robinson, Oliver Bown, and Mari Velonaki. 2022. Designing Sound for Social Robots: Candidate Design Principles. *International Journal of Social Robotics* 14 (2022), 1507–1525. Issue 6. https://doi.org/10.1007/s12369-022-00891-0

[82] Frederic Anthony Robinson, Mari Velonaki, and Oliver Bown. 2021. Smooth Operator: Tuning Robot Perception Through Artificial Movement Sound. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 53–62. https://doi.org/10.1145/3434073.3444658

[83] Dirk Rothenbücher, Jamy Li, David Sirkin, Brian Mok, and Wendy Ju. 2016. Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 795–802. https://doi.org/10.1109/ROMAN.2016.7745210

[84] Lincoln Ryave and James N. Schenkein. 1974. Notes on the art of walking. In *Ethnomethodology. Selected Readings*, Roy Turner (Ed.). Penguin Education, Middlesex, England, 265–274.

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

[86] Richard Savery, Ryan Rose, and Gil Weinberg. 2019. Establishing Human-Robot Trust through Music-Driven Robotic Emotion Prosody and Gesture. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE Press, 1–7. https://doi.org/10.1109/RO-MAN46459.2019.8956386

[87] Emanuel A Schegloff. 2007. *Sequence Organization in Interaction*. Cambridge University Press, Cambridge. 1–300 pages. https://doi.org/10.1017/CBO9780511791208

[88] Emanuel A Schegloff and Harvey Sacks. 1973. Opening up closings. *Semiotica* 8, 4 (1973), 289–327. https://doi.org/10.1515/semi.1973.8.4.289

[89] Jack Sidnell and Tanya Stivers. 2012. *The Handbook of Conversation Analysis*. Vol. 121. John Wiley & Sons.

[90] Erik Lee Stayton. 2020. *Humanizing Autonomy: Social Scientists' and Engineers' Futures for Robotic Cars (Doctoral dissertation)*. Ph. D. Dissertation. Massachusetts Institute of Technology. https://dspace.mit.edu/bitstream/handle/1721.1/129050/1227095115-MIT.pdf

[91] Lucy A Suchman. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, Cambridge.

[92] Lucy A Suchman. 2007. *Human-Machine Reconfigurations: Plans and Situated Actions, 2nd ed.* Cambridge University Press, New York, NY, US. 314 pages.

[93] Ryota Suzuki, Taichi Yamada, Masaya Arai, Yoshihisa Sato, Yoshinori Kobayashi, and Yoshinori Kuno. 2014. Multiple Robotic Wheelchair System Considering Group Communication. In *Advances in Visual Computing*, George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Ryan McMahan, Jason Jerald, Hui Zhang, Steven M Drucker, Chandra Kambhamettu, Maha El Choubassi, Zhigang Deng, and Mark Carlson (Eds.). Springer International Publishing, Cham, 805–814.

[94] John C. Tang. 1989. *Toward an understanding of the use of shared workspacess by design teams*. Ph. D. Dissertation. Stanford University.

[95] Hamish Tennent, Dylan Moore, Malte Jung, and Wendy Ju. 2017. Good Vibrations: How Consequential Sounds Affect Perception of Robotic Arms. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Lisbon, Portugal). IEEE Press, 928–935. https://doi.org/10.1109/ROMAN.2017.8172414

[96] Erik Vinkhuyzen and Melissa Cefkin. 2016. Developing Socially Acceptable Autonomous Vehicles. *Ethnographic Praxis in Industry Conference Proceedings* 2016, 1 (2016), 522–534. https://doi.org/10.1111/1559-8918.2016.01108

[97] Allison Woodruff and Paul M Aoki. 2004. Conversation analysis and the user experience. *Digital Creativity* 15 (12 2004), 232–238. Issue 4. https://doi.org/10.1080/1462626048520184 doi: 10.1080/1462626048520184.

[98] Akiko Yamazaki, Keiichi Yamazaki, Yoshinori Kuno, Matthew Burdelski, Michie Kawashima, and Hideaki Kuzuoka. 2008. Precision Timing in Human-Robot Interaction: Coordination of Head Movement and Utterance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. Association for Computing Machinery, New York, NY, USA, 131–140. https://doi.org/10.1145/1357054.1357077

[99] Selma Yilmazyildiz, Robin Read, Tony Belpeame, and Werner Verhelst. 2016. Review of Semantic-Free Utterances in Social Human–Robot Interaction. *International Journal of Human-Computer Interaction* 32, 1 (jan 2016), 63–85. https://doi.org/10.1080/10447318.2015.1093856

[100] J.D. Zamfirescu-Pereira, David Sirkin, David Goedicke, Ray LC, Natalie Friedman, Ilan Mandel, Nikolas Martelaro, and Wendy Ju. 2021. Fake It to Make It: Exploratory Prototyping in HRI. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 19–28. https://doi.org/10.1145/3434074.3446909

[101] Brian J Zhang, Nick Stargu, Samuel Brimhall, Lilian Chan, Jason Fick, and Naomi T Fitter. 2021. Bringing WALL-E out of the Silver Screen: Understanding How Transformative Robot Sound Affects Human Perception. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 3801–3807. https://doi.org/10.1109/ICRA48506.2021.9562082

[102] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through Design as a Method for Interaction Design Research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 493–502. https://doi.org/10.1145/1240624.1240704