Conversation Analysis and Research Through Design — A Fruitful Interaction?

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

ABSTRACT

Research through design papers often point out that there is a gap between theory and practice. This paper argues for a more nuanced view by introducing multimodal conversation analysis (CA). This is an approach to human interaction that allows generalization and theory-building, while being simultaneously rooted in individual cases. The paper shows how CA can be fruitfully combined with design practices in exploring different sound patterns for autonomous shuttle buses.

KEYWORDS

research through design, conversation analysis, sound, autonomous vehicles

1 INTRODUCTION

Many technologies focus on *mediating* interaction with other people or with the material world. Robots (and voice assistants) in contrast are themselves participants that *coordinate with humans* in fundamentally human social practices, e.g. when performing greetings or autonomously moving among crowds of people. The ability for autonomous action raises interesting questions regarding agency and intersubjectivity. To explore these questions, Research through Design (RtD) may benefit from engaging more with methods and methodology that have long been studying human sociality as a collaborative and joint process. In this paper, I will introduce *multimodal conversation analysis* (CA) and highlight how a combination with design practices can generate new knowledge.

RtD often highlights a gap between abstract theory and design practice [19, 20]. "Theory" in these examples indeed seems to refer to models that aim to abstract away from specific situated examples of interaction. Approaches rooted in cognitive science, for instance, typically focus on describing models of psychological and interactional processes. However, there are other approaches to human interaction and sociality, which focus on the specific details of individual cases and therefore may be better suited for RtD.

I have argued earlier [14] that multimodal conversation analysis can generate intermediate-level knowledge [9]. CA grounds all analysis in individual cases and stays true to the local and specific even when identifying more general patterns. Being both rooted in specific examples and generally compatible with more abstract descriptions of interactions such as Herb Clark's [3] theories of common ground, multimodal conversation analysis is an approach that can mediate some of the tensions between the abstract and the specific. By exploring different sound patterns for autonomous shuttle buses, I will illustrate how multimodal conversation analysis and RtD can be combined. The contributions of the paper are twofold: 1) it introduces an approach that can mediate between abstract theory and design practice and 2) it provides a brief example of how multimodal conversation analysis and RtD can enter a fruitful dialogue that generates new knowledge.

2 CONVERSATION ANALYSIS

While the term *conversation* analysis may suggest a focus on language, CA has a more general interest in action-in-interaction, and is sometimes labeled as interaction analysis [2, 10]. CA began with the study of spoken language in phone conversations in the 1960s (e.g. [17]), using the technological possibilities of the time. As video cameras and tools for analysis have advanced, CA has transitioned to more holistic studies of action. Language is seen as tightly intertwined with the body [7, 11, 13] and actions are regarded as *multimodal Gestalts* that may involve language, gesture, gaze, body postures, movement and embodied manipulations of objects [13]. In treating embodiment as a crucial element in social action, multimodal conversation analysis shares theoretical foundations with Paul Dourish's [4] and Kia Höök's [8] perspectives on the body as being deeply entangled with our sense-making practices.

CA's main objective is the study of social interaction and its organization in fine detail. Analysis involves detailed study and transcription of short video snippets that feature human interaction. Conversation analysis is itself rooted in ethnomethodology and is sometimes described as a micro-sociological method. Actions are seen as situated, contingent and dynamically adjusted, rather than indifferently following a pre-determined plan [18].

Focus lies on how participants demonstrably make sense of each other's actions - including those of machines and robotic artefacts. The central question for conversation analysts is "why that (behavior) now (at this point in time)?" [16]. CA focuses on the *sequential organization* of action, studying how participants through a current action display their understanding of a previous action and make particular next actions relevant. Such sequences can emerge both between a voice agent and a human through language, or fully non-verbally when jointly moving in traffic for example.

3 CA MEETS RTD

Conversation analysis and design can be productively combined to generate new knowledge. CA is particularly relevant for designing speech (see e.g. [5, 6, 12], but it can also inform the design of non-verbal actions [1]. Drawing on my work-in-progress on autonomous shuttle buses in everyday traffic, I will demonstrate how CA and RtD can jointly explore sound.

3.1 Identifying a Design Opportunity

Two autonomous shuttle buses are currently demonstrated in Linköping, a city in Sweden. This provides an opportunity to generate

new knowledge about how a future of traffic could look like, and how these vehicles could co-exist and coordinate with other traffic participants. Initial observations showed that the buses keep emergency braking when people, especially cyclists, get too close. Taking a CA perspective and comparing this to basic principles of human social interaction, it became evident that humans seem to expect that shuttle buses participate in practices of *being overtaken* – by slowing down or making space [15]. Braking is generally possible but often was triggered too late. Online updates of the trajectory in contrast are not within the shuttles current abilities, as they follow a pre-determined route. In some way, the buses thus need to indicate that people are getting too close well in time, so that a smooth traffic flow instead of emergency braking is facilitated.

3.2 Exploring Three Sound Designs

In this work I decided to focus on sound as a communicative resource, as it does not require that one is looking at the shuttle bus. I explored several sounds iteratively in a Wizard-of-Oz setting. A small Bluetooth speaker was taped to the front of one of the buses and sounds were played through a smartphone app. The goal of this was not to identify the best sound, but to learn something about how shuttle buses can be made at home on public roads.

3.2.1 Keeping the distance. Starting from the observation that people get too close to the bus, a first idea was to keep people at a safe distance. Inspired by the sounds of bees and other insects, I explored how humming sounds could achieve this. Such sounds seemed to make a lot of sense when played through headphones in a simple simulation by a sound design colleague. However, it turned out during the field tests that these sounds are typically ignored by people. The safety driver on the bus seemed not surprised by this and indicated that such low frequency sounds are omnipresent in public space, coming from e.g. building sites, cooling devices or generators and people may therefore not pay attention to them. A high-pitched mosquito-like sound did draw some gazes – though mostly towards the wheels of the bus. From an autoethnographic view, the sounds felt very unnatural when played in the real world setting. The safety driver also expressed concerns that these sounds may keep people from taking (free) rides on the buses.

Taking a CA view, people in the vicinity of the bus clearly did not make sense of the sounds as an invitation to jointly coordinate the shared use of the road. Instead, they seemed to treat the sounds as natural part of the movement of the bus, at most indicating some problems with the wheels (in the case of the mosquito sound).

3.2.2 Building sequences. Drawing on the conversation analytic concept of sequence, sounds that could be easily repeated were explored in more detail in a second iteration. A sound needs to be interpretable in its local context, as part of a multimodal Gestalt, which includes e.g. the shuttle buses' movement. A sound that is played while the bus is halted, followed by the bus slowly starting to roll forward from a stop will be interpreted as an announcement that "the ride is starting". A sound that is played during the ride, repeatedly with rising intonation may be heard in a very different way. CA focuses on how actions are made sense of in the current situation and how they shape what immediately happens afterwards. If a sound is not reacted upon, it is not supporting coordination.

In reverse, one may repeat sounds until they produce a visible reaction.

I selected a range of bell chimes and clown horn sounds for this purpose and played them well in advance when spotting someone in the shuttle buses' invisible tracks. While typically not eliciting the desired reaction when only played once, repeating the sounds three or more times when someone was crossing the buses' trajectory typically resulted in people looking at the bus and adjusting their route away from the bus. From a conversation analytic perspective, participants thereby displayed their analysis of the sound as "you are in my way" or "please move". Repetition of the sounds serves to build sequences, which helps to highlight that they are not random noises but that they have a communicative purpose. In social interaction, humans typically repeat their speech or their actions when assuming that they are not heard or noticed by others.

Interestingly, this was not the only way these sounds were interpreted. In one case, a student decided to take a ride on the bus after a bell sound had been played while he was walking in front of the bus. As he stepped onto the shuttle bus he said to the safety driver and the researcher that he felt like the bus had invited him to take a ride by playing sounds at him. In this case, the bell sound was heard and accounted for as a summons to take a ride. This highlights that by being more ambiguous than speech or specific symbols, sound offers a multitude of situated interpretations, which can lead to interesting interactions.

3.2.3 Continuous jingles. Finally, a more prolonged sound was explored, as sound sequences may be constantly triggered when maneuvering among crowds. The refrain of the song "The Wheels on the Bus" was selected for this purpose. The upbeat electronic jingle may have reminded locals of the ice cream trucks, which are popular in Swedish neighborhoods The jingle elicited unexpected responses. For instance, a young girl made her dad stop, got off her bike and danced while the shuttle bus was passing her. Several people stopped the bus to ask what it was doing or hopped on the bus, saying that they had heard the bus approach from a distance. Interestingly, the jingle easily achieved what the humming sounds did not - as people heard the shuttle bus arriving, they typically adjusted their trajectories as soon as they acknowledged the bus.

Taking a CA perspective can help to understand such jingles in terms of *accountability*. People (and robots) are accountable for their actions, e.g. they typically provide an account when they are not complying with interactional expectations. On campus, the shuttle buses drive in what is loosely designated as a bike lane. They are thereby entering a space that is not neutral but which exclusively belonged to cyclists before. Announcing their presence loudly and clearly draws attention. Cyclists and pedestrians seem to answer the question "why that jingle now?" by treating it as an invitation to engage with the shuttle buses and as an account why they are there. This may be a way to actively mitigate the intrusion into a space that used to belong to cyclists. Indicating their presence loudly and clearly instead of acting as a silent intruder may help to raise acceptance for such new public vehicles.

4 DISCUSSION

Starting from the task of preventing emergency brakes through subtle soundscapes, the iterative process deployed in the reported

research eventually led to a deeper understanding of the embedding of the shuttle buses into the local context. Intertwining design practices with conversation analytic methodology can generate new knowledge that is rooted in a detailed understanding of human sociality and interaction. It highlights that roads are not empty spaces in which pedestrians and cyclists are occasionally present, but that traffic is highly coordinated social interaction, in which autonomous vehicles, such as shuttle buses for public transport, need to find their place and role as constantly accountable agents.

Figure 1: RtD meets CA

Figure 1 illustrates how CA and RtD can be combined. In a first step (ANALYSE), conversation analysis enables systematic identification of a research problem. At this point, transcripts of video snippets can facilitate detailed analysis and understanding of the specific design context. Transcripts of human interaction in prior work can further serve as a source of inspiration in the design process and help to understand how people typically interact in specific settings. In the second stage (DESIGN), design practices and CA analysis can enter an iterative cycle. While design methods are crucial in creating the actual prototypes, CA helps to reflect videotaped interactions in a systematic way, asking "why is this person reacting in this way right now?" (see section 3.1). Each iteration can serve to further explore and challenge interactional patterns. Mutual inspiration will likely be most fruitful (but also most challenging) at this stage. Finally, CA can take a central position in thoroughly analysing the design iterations (REFLECT). At this stage, assembling larger collections of similar video clips can help to describe emerging practices, which can elevate the analysis from description of interesting individual cases towards more general patterns. Data from different robots could be compared to further deepen the insights. CA can operate both with single cases and on large collections and can therefore be applied at different scales, depending on the goals of the research project.

5 CONCLUSION

In this paper I have attempted to show how multimodal conversation analysis and research through design can enter a fruitful dialogue. CA can help to identify design opportunities and to evaluate and reflect learnings. Most importantly, CA and design can enter an iterative cycle, in which CA offers a toolkit for immediately reflecting upon observed reactions. Basing analyses on single cases and simultaneously allowing description of more general patterns, CA is particularly suited for processes that generate intermediate-level knowledge, in between entirely abstract models and close descriptions of specific cases. Using the example of autonomous shuttle buses, I have briefly sketched how CA and RtD can be practically combined, and how this can generate new insights. Joining forces will be fruitful for designers and conversation analysts alike

and will likely deepen our understanding of robot agency, sociality and intersubjectivity.

ACKNOWLEDGMENTS

REFERENCES

[1] Lucy Anderson-Bashan, Benny Megidish, Hadas Erel, Iddo Wald, Guy Hoffman, Oren Zuckerman, and Andrey Grishko. 2018. The Greeting Machine: An Abstract Robotic Object for Opening Encounters. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 595–602. https://doi.org/10.1109/ROMAN.2018.8525516

[2] Mathias Broth and Leelo Keevallik (Eds.). 2020. *Multimodal Interaktionsanalys.* Studentlitteratur, Lund, Sweden. 464 pages.

[3] Herbert H Clark. 1996. *Using Language.* Cambridge University Press, Cambridge.

[4] Paul Dourish. 2001. *Where the Action is: The Foundations of Embodied Interaction.* MIT Press, Cambridge, MA, USA.

[5] 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

[6] 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

[7] Charles Goodwin. 1979. The interactive construction of a sentence in natural conversation. In *Everyday Language: Studies in Ethnomethodology*, George Psathas (Ed.). Irvington Publishers, New York, 97–121.

[8] Kristina Höök. 2018. *Designing with the body: Somaesthetic interaction design.* MIT Press, Cambridge, MA, USA.

[9] Kristina Höök and Jonas Löwgren. 2012. Strong Concepts: Intermediate-Level Knowledge in Interaction Design Research. *ACM Trans. Comput.-Hum. Interact.* 19, 3 (oct 2012). https://doi.org/10.1145/2362364.2362371

[10] Brigitte Jordan and Austin Henderson. 1995. Interaction Analysis: Foundations and Practice. *The Journal of the Learning Sciences* 4, 1 (sep 1995), 39–103. http://www.jstor.org/stable/1466849

[11] Leelo Keevallik. 2018. What Does Embodied Interaction Tell Us About Grammar? *Research on Language and Social Interaction* 51, 1 (jan 2018), 1–21. https://doi.org/10.1080/08351813.2018.1413887

[12] Dimosthenis Kontogiorgos and Hannah R. M. Pelikan. 2020. Towards Adaptive and Least-Collaborative-Effort Social 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, 311–313. https://doi.org/10.1145/3371382.3378249

[13] Lorenza Mondada. 2019. Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics* 145 (may 2019), 47–62. https://doi.org/10.1016/j.pragma.2019.01.016

[14] Hannah R.M. Pelikan. 2020. Intermediate-Level Knowledge: A Conversation Analysis Perspective. (2020). http://hridesign.eu/assets/pdf/Pelikan.pdf

[15] Hannah R. M. Pelikan. 2021. Why Autonomous Driving Is So Hard: The Social Dimension of Traffic. In *HRI '21 Companion* (Boulder, CO, USA, March 8–11, 2021). ACM, New York, NY, USA. https://doi.org/10.1145/3434074.3447133

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

[17] Emanuel A Schegloff. 1968. Sequencing in Conversational Openings. *American Anthropologist* 70, 6 (aug 1968), 1075–1095. http://www.jstor.org/stable/669510

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

[19] John Zimmerman and Jodi Forlizzi. 2014. *Research Through Design in HCI.* Springer New York, New York, NY, 167–189. https://doi.org/10.1007/978-1-4939-0378-8_8

[20] 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 (CHI '07).* Association for Computing Machinery, New York, NY, USA, 493–502. https://doi.org/10.1145/1240624.1240704