

# Limits and opportunities for mathematizing communicational conduct for social robotics in the real world? Toward enabling a robot to make use of the human's competences

Response to the *Question 2*: “Are there limits to mathematization?”. In Gesa Lindemann, (this volume), *Social interaction with robots—three questions*

Karola Pitsch<sup>1</sup>

Received: 28 April 2015 / Accepted: 8 October 2015 / Published online: 22 December 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** Given the widespread goal of endowing robotic systems with interactional capabilities that would allow users to deal with them intuitively by using means of natural communication, the text addresses the question to which extent it would be possible to mathematize (aspects of) social interaction. Using the example of a robotic museum guide in a real-world scenario, central challenges in dealing with the situatedness and contingency of human communicational conduct are shown using fine-grained video analysis combining the robot's internal perspective with the user's view. On a conceptual level, the text argues to consider human and robot as *one* ‘interactional system’ that jointly solves a practical (communicational) task. This opens up the perspective to integrate the human's interactional competences and adaptability in the design and modeling of interactional building blocks for HRI. If we provide the technical system with systematic resources to make use of the human's competences, the limits of mathematization might gain an interesting twist. Through careful design of the robot's conduct, a powerful resource exists for the robot to proactively influence the users' expectations about relevant subsequent actions, so that the robot could contribute to establishing the conditions which would be most beneficial to its own functioning.

**Keywords** Human-robot-interaction · Real world · Museum guide robot · Co-orientation · Mathematization · Interactional system · Social construction of technology

*“If some useful interchange between these modalities of work is to be realized, it is most likely to come not from transforming the object which you would like to learn, but from taking it seriously in its own terms.”* (Schegloff 1996: 29 on “computational approaches of discourse”).

## 1 Introduction

Research in Social Robotics strives to endow robotic systems with interactional capabilities that allow users to deal with them intuitively by using means of natural communication and social interaction. This goal is particularly challenging because of the discrepancy between the situatedness, contingency and indexicality of human social conduct and the formalized descriptions required to program technical systems (Suchman 1987). Rule-based approaches to discourse modeling stand in direct conceptual contrast to the openness and unpredictability of social interaction, and it is unclear on what grounds a technical system can select an appropriate and relevant subsequent action. Levinson (2006: 45/56) points out that there is “no such thing as a formal grammar of discourse” because interaction is “governed not by rule but by expectations” (see also Schegloff 1996; Button 1990; Luhmann 1984). This becomes particularly evident at times that require a high degree of interactional coordination between co-participants, such as the opening of an encounter and attempts to establish co-orientation (e.g., Pitsch et al. 2013, 2014). Thus, Lindemann (this vol.) asks critically whether it would be possible to mathematize joint attention,

---

✉ Karola Pitsch  
karola.pitsch@uni-due.de

<sup>1</sup> Institute of Communication Studies, Faculty of Humanities, University of Duisburg-Essen, Universitätsstr. 12, 45141 Essen, Germany

expectations or indexical expressions, or more generally: “Are there limits to mathematization?”

In what follows, we will discuss this question using the example of an autonomous robotic research prototype set up as guide in a real-world museum site (e.g., Pitsch and Wrede 2014). We will point to challenges in mathematizing social conduct on different levels, in particular those that become evident when combining the robot’s internal perspective with the participants’ view. Provided the conceptual and factual impossibility of equipping technical systems with full human-like social and interactional competences, we suggest taking the idea of “hybrid socio-technical systems” (Rammert and Schulz-Schaeffer 2002) further. Adopting an interactional perspective that understands human and robot as *one* interactional system (Luhmann 1984), we suggest that an important—yet mostly neglected—resource for the robotic system consists of the *human’s* interactional competences and adaptability. If we can provide the technical system with systematic resources to make use of them (Pitsch et al. 2013), the limits of formalization might gain an interesting twist.

## 2 Goal: intuitive human–machine interface or reproducing human communication?

Thinking about the possibilities and limitations of mathematizing (human) communicational conduct is closely tied to the goals of Social Robotics. One strand of research seeks to *reproduce natural human communication* explicated in formulations such as “we consider that establishing models is a path to make such a robot fully behave in a natural way as humans do” (e.g., Kanda and Ishiguro 2012:102). Another strand considers HRI as a *particular type of human–machine interface* that should allow the user to deal with a technical system in most intuitive ways by using means of human natural communication (e.g., Breazeal 2003). This way, the design of the interface is—as suggests Suchman (1987:22) on human–machine interaction more generally—“less a project of simulating human communication than of engineering alternatives to interaction’s situated properties.” These two approaches entail different requirements for the formalization of communicative principles and conduct: In the first case, researchers would need to build models able to address the inter-individual variability of multimodal conduct, local-indexical sense-making practices and the unpredictability of emergent interactional processes. This is a goal so ambitious that this author would be too humble to strive for. In contrast, the second approach would enable us to conceptually take into consideration the different (evolving) competences and status of machines and humans and their particular (changing) relationship to each other. This

would allow us to open the perspective toward solutions *functional for human–robot interaction (HRI)* and include—as an important resource—the *human’s* competences and adaptability in the modeling.

## 3 Mathematization: transforming communication for real-world HRI

Formalization and mathematization of real-world phenomena—such as communicational conduct—are based on the assumption of *idealized* objects (Lenhard and Otte 2005) and thus constitute a *transformation* that changes the phenomenon itself (Lynch 1988; Schegloff 1996). While it is impossible to escape the challenge of unpredictability and contingency when dealing with real-world phenomena, a particular phenomenon can be modeled in different ways. The limits of mathematization are thus not predefined *per se*, but depend on the frame we choose (Lenhard and Otte 2005). In this regard, the current conceptualizations in Social Robotics/HRI range from highly restricted one-way communication over laboratory experiments with highly idealized conditions of the physical environment and pre-trained users (e.g., Sugiyama et al. 2012 for a “model of natural deictic interaction”) up to approaches dealing with the complexity of real-world settings (Shiomi et al. 2008; Yamazaki et al. 2009; Pitsch and Wrede 2014).

While highly idealized laboratory conditions provide better grounds to model more sophisticated interactional conduct, we believe that it is necessary to assume early on the challenge of exploring autonomous systems in real-world settings (see Lindemann and Matzusaki 2014). Such an approach enables us to gain a better understanding of the full complexity of the phenomenon and the specific conditions of the human–robot interface (as opposed to attempting to reproduce human communication). In doing so, we begin with inspiration from human communication (see also Yamazaki et al. 2007), but have to reduce its multimodal complexity to the most salient features (Pitsch et al. 2014), in addition to making other types of adjustments. Transformation of the phenomenon “human communication” is thus a *conditio sine qua non*, but does not pretend *per se* to discard the idea of interactivity (see Schegloff 1996:29).

## 4 Example of real-world human–robot interaction: a robotic museum guide

We will explore the question of limits and opportunities in mathematizing communicational conduct using the example of a robotic museum guide deployed at the

Bielefeld Historical Museum in September 2014 (see also Gehle et al. 2015). A humanoid NAO robot was positioned on a table (1.20 × 2 m, 0.7 m high) and set up to autonomously engage in a focused encounter with visitors and to give explanations about several exhibits by using talk, head and arm gestures and walking across the table (see images of the setting below). The system's functions relied on the perceptual results from the robot's internal VGA camera(s) and an external microphone positioned on the table.

#### 4.1 Shaping expectations

When human and robot enter in contact with each other, they establish the conditions for their interaction. Users are faced with the task of discovering what the system can do and what it might be responsive to. This is a privileged moment in which the system can—through its own conduct—pro-actively shape the users' perception of its capabilities, their expectations about roles, ways of participating and relevant subsequent actions (Pitsch et al. 2012, 2013, in press).

In our case, the robot is designed to greet “hello; i am nao” accompanied by a head nod. It then offers to provide information and asks “would you be interested,” which, again, is accompanied by a small head nod at the end of the utterance. Video recordings of such situations show that the visitors build hypotheses about relevant subsequent actions and the robot's interactional capabilities based on the communicational resources used by the robot in the opening phase. This becomes particularly visible when the robot does not provide the subsequent action in the time-frame expected by the visitors, and they begin to explore different ways of making the robot continue. In session 4-004 of our corpus (which will be used here as a case example), the visitors try out different ways—[head nod + “yes”] (V2), repeated head nods (V1), pronounced and loud “yes” (V2), “yes” (V3)—to answer the robot's question and take up the multimodal resources which the robot has introduced itself in its initial utterances.

In this way, through careful design of the robot's conduct, a powerful resource exists to pro-actively influence the users' expectations for relevant subsequent actions. The robot could thus contribute to establishing the interactional conditions which would be most suitable for its own functioning. We suggest that such an interactional approach could help to reduce parts of the contingency and openness of communication without, however, eliminating them. Systematic empirical research will need to explore in which ways these issues might become more manageable in HRI and how far we can go with, e.g., combinations of rule-based and probabilistic modeling (see Lison 2015)

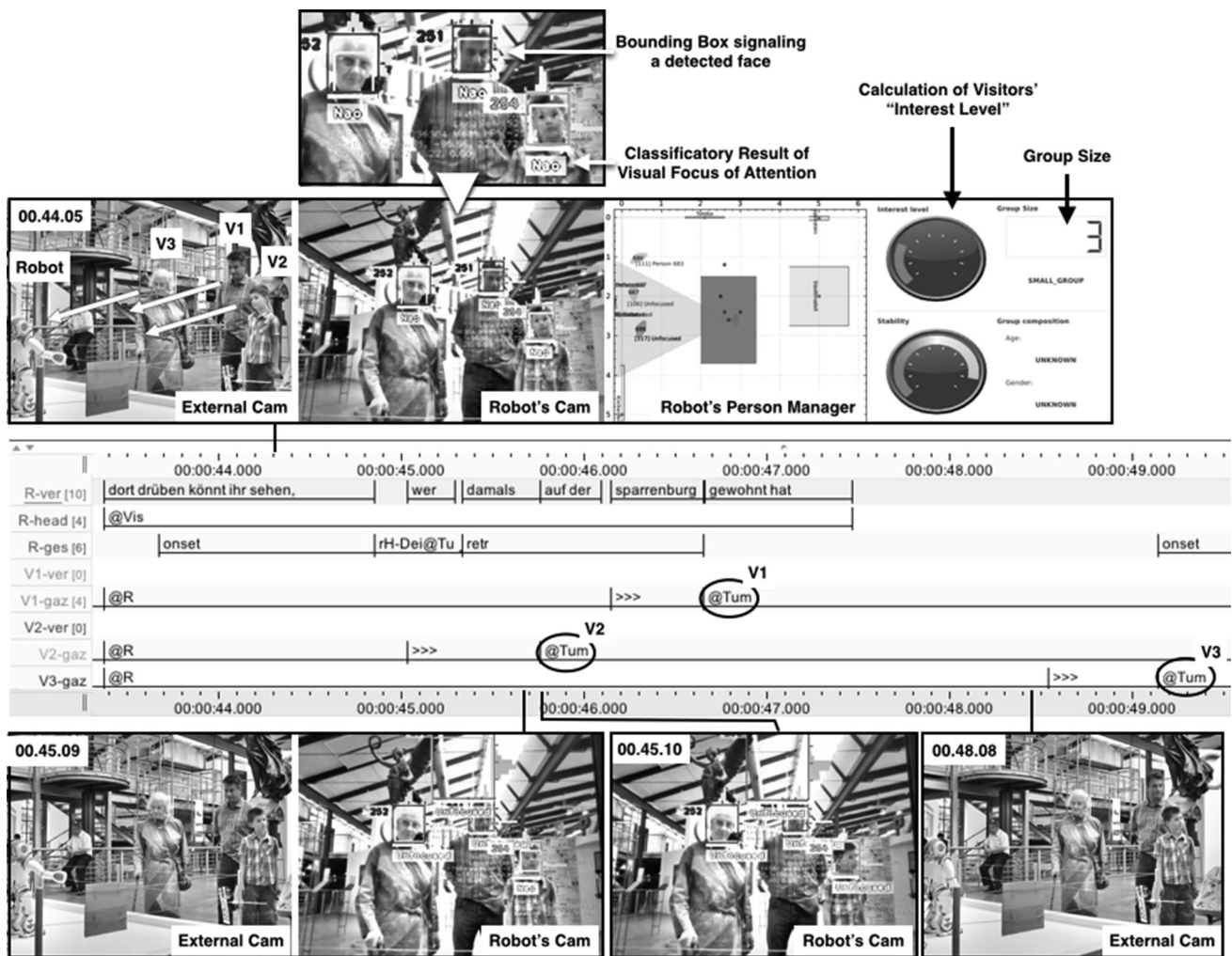
combined with local building blocks for dealing with misunderstanding.

#### 4.2 Establishing co-orientation: challenges for mathematization and the interactional system ‘human and robot’

To provide information about some exhibit, the robot is faced with the task of orienting visitors to a particular object. This not only constitutes an individual deictic act, but also requires—at least basic—forms of interactional coordination (Pitsch and Wrede 2014). In our case (session 4-004), the robot is set up to invite the visitors to orient to the life-size image of a tomb slab by saying “over there you can see who used to live at the Sparrenburg [i.e., the name of local medieval castle]” and extend its right arm to perform a pointing gesture with its head turned to the visitors (Fig. 1, #00.44.05). From the three visitors in our fragment, who are initially facing the robot (#00.44.05), two (V2, V1) follow the robot's deictic reference and successively turn their head in the indicated direction (#00.45.09). Only visitor V3 keeps looking at the robot during the utterances and during the following 1.5 s (#00.48.08). This situation offers insights into a set of issues on mathematizing interactional phenomena.

##### 4.2.1 Uncertainty of the robot's perception

The robot's perspective in this situation is based on the input of its internal VGA camera and the calculation resulting from modules for detecting/tracking users and categorizing their visual focus of attention (Sheikhi and Obodez 2012). At the beginning of the robot's utterance (#00.44.05), three visitors (displayed as bounding boxes around their heads, group size = 3) are detected and classified as oriented “to Nao” and correctly located in the robot's spatial model. When V2 shifts his orientation to the exhibit—from #00.45.09 to #00.45.10—this is directly perceivable by the robot and correctly interpreted—from “to Nao” to “unfocused.” While these results are highly promising on the technical level of perception, also the challenges set by the real-world setting become visible at the same time: V1 and V3 are also oriented to the robot, but they are not classified as such by the system, and a structure in the ceiling is momentarily categorized as a human face. Even with the ongoing improvements in the detection algorithms and filtering processes, a conceptual challenge remains: Interactional modeling needs to take into account different levels of (un)certainty in the system's perception. While there are mathematical methods for ‘smoothing’ such data streams, it is not clear to which extent they would be compatible with the moment-by-moment contingencies



**Fig. 1** Session 4-004, Transcript part 01

of social interaction or whether (from a human's perspective) interactionally relevant details might have been cancelled out this way.

#### 4.2.2 Reducing the complexity of interactional conduct

By the end of the robot's utterance, two visitors have followed the robot's invitation to inspect the relevant exhibit while V3 remains oriented to the robot (#00.48.08). How should the system interpret this situation and which next action should it undertake with what expected consequences?—On the one hand, modeling decisions are required for dealing with the diverging states of participation of multiple visitors. On the other hand, formalizations need to account for the visitors' assumed diverging states of participation. These would need to be based on perceivable interactional cues (such as head orientation) and result in quantifiable measures, probably similar to the current analogy of the 'speed indicator' used in the current

system to describe a visitor's 'Interest Level' (#00.44.05). How to best reduce the complexity of visitor conduct and interactional history in such ways and as a basis for deciding locally on the robot's subsequent action constitutes a central challenge.

#### 4.2.3 Perceptual delay and diverging representations

In our case, the robot is set up to interpret V3's focus of attention as an indicator of trouble with regard to her following the robot's reference to the exhibit and thus offers a second reference to the exhibit ("over there on the big picture"). However, the exact timing around this decision proves difficult and a perceptual delay of about (in terms of current autonomous systems: only) 0.5 s leads to diverging representations of the situation between human and robot, best visible in #00.48.12 (Fig. 2). In fact, V3—similarly to V1 and V2—begins to turn to the exhibit after #00.48.08 which is perceivable to the robot only after #00.49.01, i.e.,

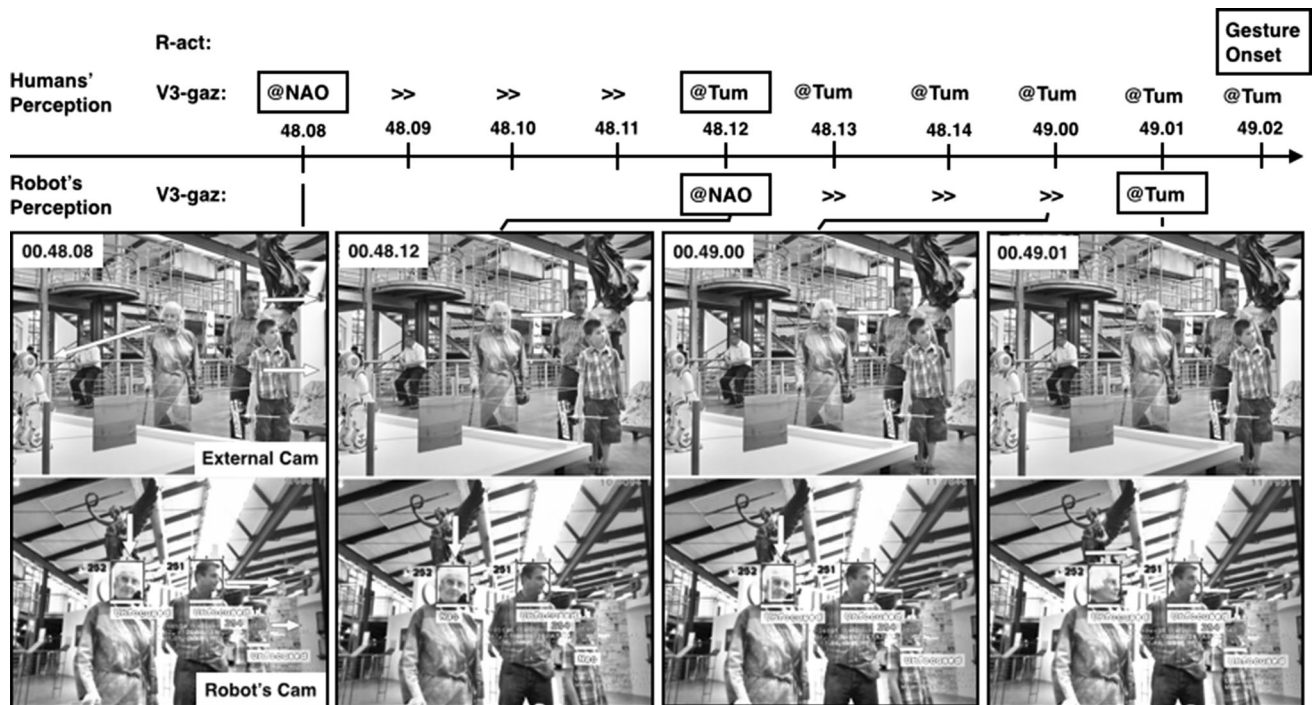


Fig. 2 Session 4-004, Transcript part 02

at the moment when it is just starting the deictic gesture of the second orientational hint. Thus, in sequential structural terms, the robot’s next action comes out ‘misplaced’, i.e., directly after also V3 is oriented to the exhibit.

4.2.4 Confusion with regard to sequential structures

While V1 does not react to the second reference (“over there on the big picture”), V2 looks back to the robot for about 4 s (Fig. 3, #00.50.06) and then re-orientes to the exhibit (#00.53.10). In contrast, V3 appears visibly confused orienting back and forth between robot and exhibit during the robot’s utterance (#00.50.06—#00.51.03—#00.51.13—#00.52.12), turns round to inspect the room (#00.53.10) and finally gazes back to the exhibit shielding her eyes with a hand visibly indicating a ‘search activity’ (#00.58.05). Thus, she treats the robot’s second reference as a repair of her last action (i.e., of her orientation to the exhibit)—an interpretation which *adequately* follows the sequential structure as it has emerged, but which is—due to the time lag—*different* from the one aimed for by the robot.

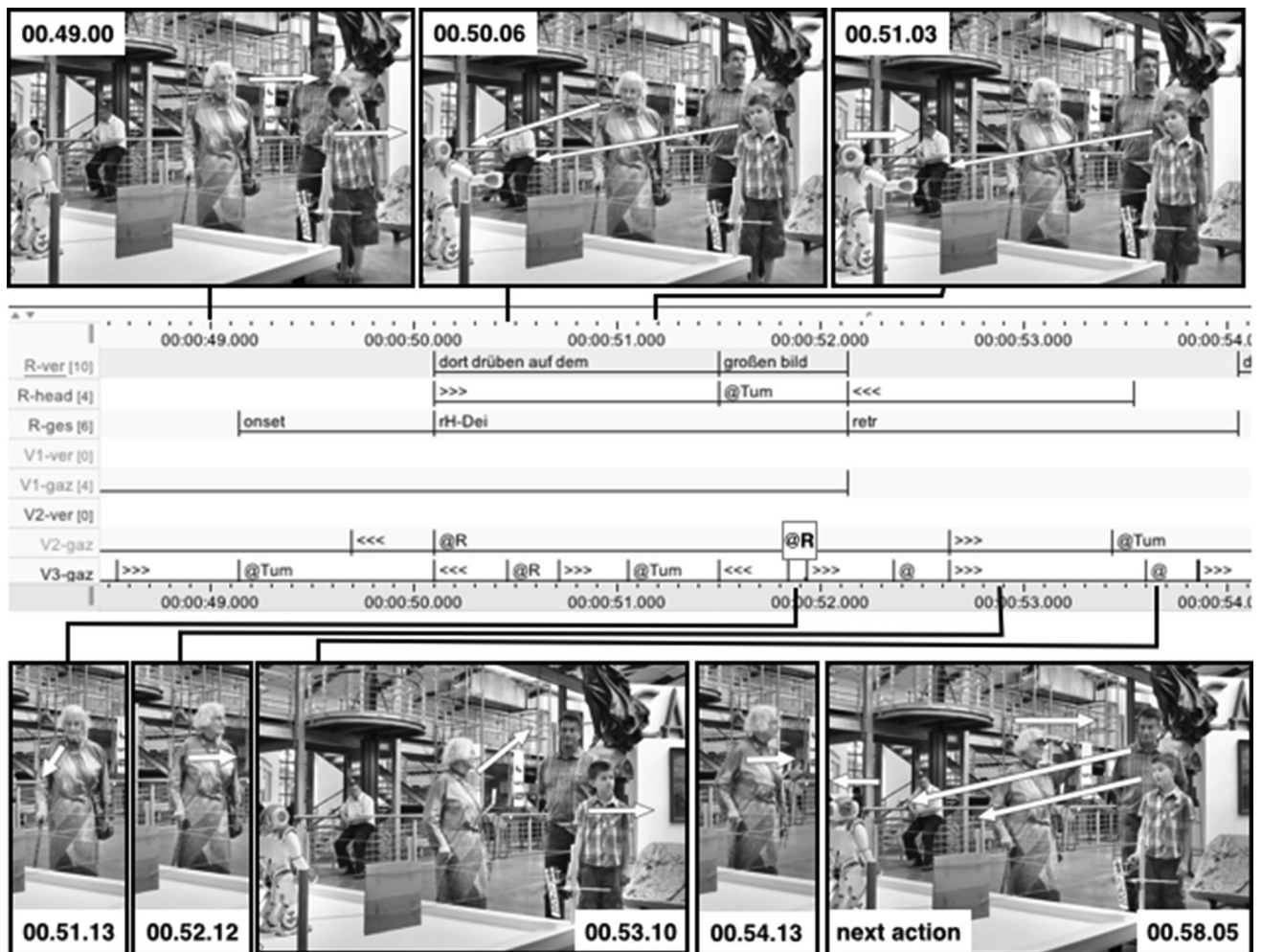
4.2.5 Robot’s resources between ‘interaction’ and ‘functioning’

The robot’s second reference was designed as an upgrade, i.e., verbally more explicit (“over there on the big

picture”) and bodily including also a head turn (in addition to the deictic gesture—#00.51.03) toward the exhibit. This entails that the robot’s cameras—located at the front of its head—cannot monitor the visitors’ conduct at this point, and as a consequence, the robot is unable to detect V3’s confusion. Formalizing interactional phenomena for HRI thus must also address the challenge of how to manage the robot’s resources in a way as to produce—at the same time and with the same resources—interactionally relevant conduct and provide the basis for its own functioning.

4.2.6 Human’s competence as a central resource in the interactional system ‘human(s) and robot’

When the robot announces the next action—i.e., to go to the exhibit indicated—V1 and V2 promptly acknowledge this invitation (#00.58.08) and begin to reposition themselves. In contrast, V3—who is searching for the indicated exhibit—does not engage in the new activity. As the robot needs to turn its head for navigational purposes, it is, again, unable to recognize nor to provide a solution to this problem. In this case (as in many others instances in our corpora), it is the *human’s competence* which solves the problem and helps to re-establish functional sequential structures. Here, V1 incites V3 to refocus her attention, invites her to join the next action and makes transparent the next relevant action. In this way, all three visitors happen to



**Fig. 3** Session 4-004, Transcript part 03

gather in front of the exhibit when the robot also arrives ready to engage in the next explanation.

## 5 Conclusion

In this text, we have attempted (1) to point out a set of challenges that researchers are faced with once they engage in modeling interactional conduct for autonomous robot systems in real-world situations with untrained users. And (2) we have developed a vision and a conceptual basis of how the limitations of technical systems in dealing with the situatedness of human social interaction might be pushed a little further. To consider human and robot as *one* ‘interactional system’ (Luhmann 1984; Rammert and Schulz-Schaeffer 2002; Pitsch et al. 2013), in which the participants jointly solve the practical tasks, makes it possible to integrate the human’s competence in the development of building blocks for interactional conduct in HRI. Through

careful design of the robot’s conduct, a powerful resource exists to pro-actively influence the users’ expectations about relevant subsequent actions, so that the robot could contribute to establishing the conditions which would be most beneficial to its own functioning.

As a consequence, the question of whether a technical system is able to deal with situatedness, contingency, indexical expressions, etc., could be reformulated to ask in what ways the interactional system ‘human and robot’ can solve these practical tasks. In this way, mathematization would not need to provide self-contained models, but rather think of ways to include the human’s competences of sense-making and organizing interaction as well as of equipping robotic systems with strategies to make their own actions and states transparent to the user. As such, the limits of mathematization might present with a different twist.

**Acknowledgments** The empirical data used in this text have been created as a joint interdisciplinary effort by Timo Dankert

(Informatics), Raphaela Gehle (Interactional Linguistics), Karola Pitsch (Interactional Linguistics) and Sebastian Wrede (Informatics). The author is indebted to the project team for continuous inspiring discussions. Maximilian Krug has helped with editing the video stills. This work was funded by both the Volkswagen Foundation (Dilthey Fellowship “Interaction and Space. From Conversation Analysis to Dynamic Interaction Models for Human–Robot Interaction,” K. Pitsch) and as part of the Cluster of Excellence Cognitive Interaction Technology “CITEC” (EXC 277), Bielefeld University (Project IP-18 “Interactional Coordination and Incrementality in HRI. A museum guide robot,” K. Pitsch and S. Wrede).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Breazeal C (2003) Toward sociable robots. *Robot Auton Syst* 42(3–4):167–175
- Button G (1990) Going up a blind alley. Conflating conversation analysis and computational modelling. In: Luff P, Frohlich D, Gilbert N (eds) *Computers and conversation*. Academic Press, London, pp 67–90
- Gehle R, Pitsch K, Dankert T, Wrede S (2015) Trouble-based group dynamics in real-world HRI—reactions on unexpected next moves of a museum guide robot. In: *Proceedings of Ro-Man 2015*, Kobe
- Kanda T, Ishiguro H (2012) *Human–robot interaction in social robotics*. CRC Press, Boca Raton
- Yamazaki K, Kawashima M, Kuno Y, Akiya N, Burdelski M, Yamazaki A, Kuzuoka H (2007) Prior-to-request and request behaviors within elderly day care: implications for developing service robots for use in multiparty settings. In: *Proceedings of ECSCW 2007*, Limerick, pp 61–78
- Lenhard J, Otte M (2005) Grenzen der Mathematisierung: Von der grundlegenden Bedeutung der Anwendungen. *Philosophia Naturalis* 42(1):15–47
- Levinson SC (2006) On the human “interaction engine”. In: Enfield NJ, Levinson SC (eds) *Roots of human sociality. Culture, cognition and interaction*. Berg, Oxford, pp 39–69
- Lindemann G, Matsuzaki H (2014) Constructing the robot’s position in time and space. The spatio-temporal preconditions of artificial social agency. *Sci Technol Innov Stud* 10(1):85–106
- Lison P (2015) A hybrid approach to dialogue management based on probabilistic rules. *Comput Speech Lang* 34(1):232–255
- Luhmann N (1984) *Soziale systeme*. Suhrkamp, Frankfurt am Main
- Lynch M (1988) The externalized retina: selection and mathematization in the visual documentation of objects in the life sciences. *Hum Stud* 11:201–234
- Pitsch K (in press) Ko-Konstruktion in der Mensch–Roboter-Interaktion. Kontingenz, Erwartungen & Routinen in der Eröffnung. In: Güllich E, Dausendschön-Gay U, Krafft U (eds) *Ko-Konstruktionen in der Interaktion. Die gemeinsame Arbeit an Äußerungen und anderen sozialen Prozessen*
- Pitsch K, Wrede S (2014) When a robot orients visitors to an exhibit. Referential practices and interactional dynamics in real world HRI. In: *Proceedings Ro-Man 2014*, Edinburgh, pp 36–42
- Pitsch K, Lohan KS, Rohlfing K, Saunders J, Nehaniv CL, Wrede B (2012) Better be reactive at the beginning. Implications of the first seconds of an encounter for the tutoring style in human–robot-interaction. In: *Proceedings of Ro-Man 2012*, Paris, pp 974–981
- Pitsch K, Vollmer AL, Mühlig M (2013) Robot feedback shapes the tutor’s presentation. How a robot’s online gaze strategies lead to micro-adaptation of the human’s conduct. *Interact Stud* 14(2):268–296
- Pitsch K, Vollmer AL, Rohlfing K, Fritsch F, Wrede B (2014) Tutoring in adult–child-interaction: on the loop of the tutor’s action modification and the recipient’s gaze. *Interact Stud* 15(1):55–98
- Rammert W, Schulz-Schaeffer I (2002) Technik und Handeln. Wenn soziales Handeln sich auf menschliches Verhalten und technische Abläufe verteilt. In: Rammert W, Schulz-Schaeffer I (eds) *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*. Frankfurt/Main, Campus, pp 11–64
- Schegloff EA (1996) Issues of relevance for discourse analysis: Contingency in action, interaction, and co-participant context. In: Hovy EH, Scott DR (eds) *Computational and conversational discourse: burning issues—an interdisciplinary account*. Springer, Berlin, pp 3–38
- Sheikhi S, Obodez JM (2012) Recognizing the visual focus of attention for human robot interaction. In: *Proceedings of HBU 2012*, pp 99–112
- Shiomi M, Sakamoto D, Kanda T, Ishi CT, Ishiguro H, Hagita N (2008) A semi-autonomous communication robot: A field trial at a train station. In: *Proceedings of HRI ‘08*, Amsterdam, pp 303–310
- Suchman L (1987) *Plans and situated actions. The problem of human machine communication*. Cambridge University Press, Cambridge
- Sugiyama O, Kanda T, Imai M, Ishiguro H, Hagita N (2012) A model of natural deictic interaction. In: Kanda T, Ishiguro H (eds) *Human–robot interaction in social robotics*. CRC Press, Boca Raton, pp 104–120
- Yamazaki K, Yamazaki A, Okada M, Kuno Y, Kobayashi Y, Hoshi Y, Pitsch K, Luff P, Heath C, Vom Lehn D (2009) Revealing gauquin: engaging visitors in robot guide’s explanation in an art museum. In: *Proceedings of CHI 2009*, Boston, pp 1437–1446