Postprint

Karola Pitsch (2020): Answering a robot's questions. Participation dynamics of adult-child-groups in encounters with a museum guide robot.
In: Réseaux, 220-221 (2-3), 113-150, https://doi.org/10.3917/res.220.0113.

ANSWERING A ROBOT'S QUESTIONS: PARTICIPATION DYNAMICS OF ADULT-CHILD-GROUPS IN ENCOUNTERS WITH A MUSEUM GUIDE ROBOT

KAROLA PITSCH[1]

Current trends in artificial intelligence and in the digitization of everyday life bring forward the idea that technological artefacts – such as humanoid robots, embodied conversational agents or voice user interfaces – might serve as autonomous assistants or companions in the home, in public and in work settings. From early on, it has been pointed out that their "success […] hinges not only on their utility but also on their ability to be responsive to and interact in a natural and intuitive manner" (Breazeal 2003: 167). As a basis, research in informatics/engineering endows robotic systems with advanced capabilities of monitoring their environment and for acting upon the users' communicational conduct. However, it remains a major challenge for technical systems to deal with the situated, contingent and in principle unpredictable nature of human communicational and interactional conduct (Suchman 1987, 2006, Button 1990, Schegloff 1996). In this article, I suggest that interdisciplinary cooperation between informatics/engineering and research on multimodal communication based on Ethnomethodological Conversation Analysis might help researchers to respond to this challenge and to create a novel research instrument for studying situated (inter-)action and human social practices (Pitsch *et al*. 2013, 2014).

To achieve this goal, a team of researchers are exploring how small groups of visitors – comprised of children and adults – engage in question-answer sequences[2] with an autonomous humanoid robot. As part of a larger research

---

[1] The author acknowledges the financial support of the Volkswagen Foundation for the Dilthey Fellowship 'Interaction and Space' and from the Cluster of Excellence Cognitive Interaction Technology (EXC 277, project IP-18 conducted in cooperation with Dr. Sebastian Wrede). The author is indebted to her colleagues Sebastian Wrede, Raphaela Gehle and Timo Dankert for collaborative work on the project and the study investigated in this paper, to Lukas Hindemith and Hannah Pelikan for making the log files accessible, and to Vivien Ebben for her support in anonymizing the video stills.

[2] Stivers & Robinson (2006) distinguish between 'answers', i.e. specific types of responses which prototypically deal with the conditional relevance established through a question, and 'responses', i.e. anything that occurs (or not) subsequent to the previous turn. Here, I use the concept of 'answer' rather than the broader

project on situated (inter-)action, multimodal coordination and human-robot encounters, a NAO robot (Aldebaran, v4) has been set up to act as guide in a German historical museum and was deployed during a science festival (Pitsch *et al*. 2016). The guide robot asks the visitors questions and attempts to detect their answers via speech recognition and to respond in an appropriate manner. In human-robot/agent interaction and speech-based dialogue systems, such question-answer sequences have usually been designed for a *single* user and have been explored with regard to their potential for structuring and 'controlling' the interaction (Breazeal *et al*. 2016). In contrast, our data of adult-child groups reveal specific forms of *interactional dynamics between the visitors* when attempting to provide an answer to the robot's question. It turns out that the ensuing forms of participation can be problematic for the robot and lead to situations in which visitors, despite being eager to answer, are not able to do so.

Using video recordings of the human-robot encounters and log files of the system's speech recognition, we address the following questions: (1) How do small groups comprised of adults and children attempt to answer a guide robot's questions in a real-world museum setting? (2) Which interactional dynamics emerge between the visitors? (3) Which challenges and implications for the design of an autonomous robot system arise from these insights into the visitors' forms of participation? Analysis will reveal the interactional complexity of seemingly simple social practices and sequential structures – here, question-answer sequences – once they have to be implemented for an autonomous technical system 'in the wild'. This interest in human-robot interaction (HRI), or social robotics, links to studies on talk and embodied interaction, to the dynamics of group activities and to human social practices when dealing with novel technologies in real-world environments.

In what follows, I introduce HRI as a tool for investigating situated (inter-) action and detail existing research on participation, question-answer sequences and children as co-participants in human-human interaction (HHI) and HRI. After presenting the technical setup, study and data, I analyse three cases revealing the issues of participation and contrast these interactional dynamics with the robot-internal view of these instances. I then discuss the results in the context of HRI as a tool for investigating situated (inter-)action and consider practical and conceptual issues for designing robotic technologies.

'response' as the human participants' try to answer within the conditions posed by the robot's questions – even though this is sometimes difficult.

HUMAN-ROBOT-INTERACTION AS A TOOL FOR INVESTIGATING SITUATED (INTER-)ACTION

In the endeavour to investigate situated (inter-)action, the diversification of settings with their particular conditions has proven to be fruitful both for discovering new phenomena and for discussing established concepts. For instance, the spatial dimensions of an elevator (Hirschauer 1999) or a butcher's market stall (Dausendschön-Gay & Krafft 2009), or technological mediation in virtual realities (Schnier *et al*. 2011) with their specific accessibility, set constraints on how communicational resources can be mobilized and forms of sociality established. The socio-technical setting of HRI constitutes a novel methodological and conceptual tool to investigate situated (inter-)action (see Pitsch *et al*. 2013, 2014). Opposed to ideas of technologically mediated settings (Arminen *et al*. 2016) in which the *conditions under which human co-participants interact with each other* are manipulated through the technology, in the socio-technical setting of HRI, the *features and conduct of a co-participant* (i.e. the robot) can be manipulated.

This approach takes further Suchman's (2006, 186) suggestion of an analogy between the research ideas of building 'intelligent' and 'interactive' artifacts: "Just as the project of building intelligent artifacts has been enlisted in the service of a theory of mind, the attempt to build interactive artifacts, taken seriously, could contribute to an account of situated human action and shared understanding." Suchman's considerations were initially based on participants' attempts to use a complex photocopier and pointed to the basic difference between *plan-based* approaches of modelling in human-computer interaction (HCI) and the *situated* nature of human communication which is (mainly but not only) guided by the moment-by-moment sequential organization of activity (Goffman 1983, Schegloff 1996). Considering humanoid robots and forms of artificial intelligence (AI) systems as 'co-participants' in a socio-technical setting transforms the research playground of what Suchman called interactive artifacts. The technical entity might dispose of an embodiment and interactional conduct inspired by humans, the underlying modelling principles might go beyond simple plan-based approaches (see also Rammert 2002, 2008), and the system might learn in and through social interaction. If researchers interested in situated (inter-)action begin to explore these features as resources for manipulating aspects of the technical co-participant, they have a novel methodological tool at their fingertips.

How is this tool different from previous approaches in linguistics and the social sciences dealing with novel technologies? Linguistic research has an important tradition in computer-mediated communication (CMC) which has explored, often in a comparative mode, the specific conditions under which human participants communicate using different communication technologies (e.g. Herring 1996, Arminen *et al*. 2016). Robotic systems or other artificial agents have not been considered in this domain, either conceptually nor empirically (although Arminen *et al*. 2016 mentions this emerging field). The role of co-participants in socio-technical systems is incompatible with current approaches to 'mediated' communication between human participants. In a different vein, linguistic research has been occupied with human-machine interaction. Research has focused on how people talk to technology (e.g. Weingarten 1989; Fischer 2006). Yet, the multimodal nature of conduct and in particular the interactional dimension – i.e. the participants' attempts to coordinate their actions – has rarely been considered.

Linking Ethnomethodology (EM) and Conversation Analysis (CA) with an interest in novel technologies (Suchman 1987, 2006 ; Heath *et al*. 2000, Heath & Luff 2018, Dourish 2001), the approach of Workplace Studies has a strong tradition of informing the design of technologies and researching their exploration in interactional situations. Early studies pointed out the potential of informing the design of text-based dialogue systems by transferring ideas about sequential structures from HHI to technological modelling (e.g. Frohlich & Luff 1990). While these attempts were severely criticized for building a 'simulacrum of conversation' (Button 1990), they inspired a fruitful conceptual discussion which specified the difference between rules and regularities in EM/CA. After years of silence, the emergence of robotic technologies has raised new questions (Yamazaki *et al*. 2013, Pitsch *et al*. 2016, Fischer *et al*. 2019). How can robot conduct be designed so that users can intuitively engage with it? How can insights from human multimodal communication inspire the design of the technology without the pitfall of the simulacrum of conversation? Which concepts might be fruitful for thinking about human-agent configurations?

Sociological studies of technology have focused on the conceptual level addressing questions about the status of artificial agents, their agency and how to conceptualize their exchanges with humans. This has led to questioning the notion of 'interaction' (Braun-Thürmann 2002, Krummheuer 2011) and to developing novel concepts of 'socio-technical systems' with distributed and/or graded forms of agency (Rammert 2008). This work provides a central framework for my investigation of human-robot encounters: to consider the participants in HRI as forming a 'socio-technical configuration' (Rammert 2008) without predefining their specific roles; to assume a basic (not yet detailed) difference between HHI and HRI; and to use insights about human interactional practices in social situations as inspiration for the design of robot conduct without attempting to build a 'simulacrum of conversation'.

Most recently, novel perspectives have emerged from interdisciplinary cooperation between CA, Interactional Linguistics, and informatics. On the one hand, Fischer (2016) suggests considering "robots as confederates" which provide the benefit of being "reliable, controllable, interactional confederates" (3) and to "behave exactly in the way programmed and identically for each participant" (4). While this account pinpoints an important issue, it is, however, highly idealized and – at least in the current state – appears difficult to apply in practice: autonomous robotic systems still tend to lack precise controllability due to perceptual insecurities or learning algorithms; hard-coded interactional structures are limited with regard to their interactional 'fitting'; and Wizard-of-Oz approaches are critical with regard to their direct applicability to autonomous systems. On the other hand, Yamazaki *et al*. (2007, 2010, 2016) and Kuno *et al*. (2007) propose an "experimental sociological approach" which consists of an interdisciplinary incremental process of (1) analysis of human interaction, (2) development of a robotic system, (3a) analysis of how humans and robots interact, and (3b) evaluation of the robot's effectiveness to refine the development of the robot system. My understanding of HRI as a tool to investigate situated action builds on this latter research design, refining and expanding ideas and concepts (e.g. Pitsch 2019). In a nutshell, I suggest that HRI can serve as a methodological and conceptual tool for investigating situated (inter-)action on at least five levels: (1) exploring basic building blocks of situated (inter-)action involving the orchestration of communicational resources as complex 'multimodal gestalts' under the condition of interactional contingency and the co-participants' interpretation; (2) theories and concepts of interaction; (3) socio-technical constellations; (4) exploring human sociality and moral orders as they transpire in novel technologically rich situations; (5) exploring the integration of novel technologies in the ecology of our daily lives and resulting societal questions. In this article, I consider issues (1) and (4).

PARTICIPATION IN HHI AND HRI: QUESTION-ANSWER SEQUENCES & ADULT-CHILD CONSTELLATIONS

PARTICIPATION: SHAPING CONDUCT & INTERACTIONAL DYNAMICS

In human interaction, how the parties to a social encounter participate in an ongoing or projected activity is a highly dynamic issue and mutually negotiated moment by moment (Goodwin & Goodwin 2004). Refining the notion of 'speaker' and 'hearer', Goffman (1981) has introduced the idea of a 'participation framework' which differentiates between ratified and non-ratified participants, and introduces further interactional roles such as bystander, overhearer or eavesdropper. Once robot systems are deployed in real-world situations such as a museum, they are exposed to encounters with a variety of human co-participants who negotiate their forms of participation in situ. The twofold question arises as to how a technical system can be prepared for such situations and how the human participants attempt to participate in such encounters.

In Social Robotics and the development of dialogue systems, encounters have mostly been assumed to be 1:1 situations with a single user. Recently, a small set of studies exploring situations with a guide robot, a receptionist or an assistive system have begun to address the issue of multiple participants. One line of research explores how a robotic system – through its own interactional conduct – could shape how users behave (e.g. Pitsch *et al*. 2013, 2014). A robot might be able to shape spatial configurations during the opening of an encounter (Yousuf *et al*. 2012), when orienting visitors to a specific exhibit (Kuzuoka *et al.* 2010, Yousuf *et al*. 2013; Shiomi *et al.* 2007; Pitsch *et al*. 2016), to organize multiparty turn-taking (Bohus & Horvitz 2010, Bohus *et al*. 2010-a, 2010-b, 2011), or to make the ordering of service transparent when several clients ask for help (Kobayashi *et al*. 2012, Yamazaki *et al*. 2016). These studies explore the transfer of concepts from human interaction to interactional strategies for a robot shaping participation in multiparty situations. Yet, in some of the HRI experiments testing these models the human participants are given fixed positions or predefined roles. While this is helpful for evaluating the proposed model, it does not allow unexpected conduct and potential dynamics between the visitors to emerge.

Another line of research uses an exploratory approach in a real-world scenario to shed light on the participatory dynamics between visitors during a human-robot encounter. Gehle *et al*. (2015) show that visitors begin to interact with each other once a museum guide robot produces unexpected (i.e. situationally awkward) conduct: they turn to each other, avoid looking at the robot, and in so doing closely coordinate their actions. While the authors suggest using such systematic description as a basis for developing a 'trouble detector' for the technical system, these analyses invite further investigation of the interactional dynamics between visitors. Here, I am interested in the interactional dynamics of participation which emerge when small groups of adults and children encounter a museum guide robot and attempt to answer the robot's questions to them.

QUESTION-ANSWER SEQUENCES: STRUCTURAL TOOLS & INTERACTIONAL DYNAMICS

The sequential and interactional environment of question-answer (QA) sequences provides for specific forms of participation. 'Asking questions' is not only concerned with managing epistemics and access to knowledge, but also constitutes a "powerful tool to control the interaction" in that it sets constraints for next actions on the level of both topic and agenda (Hayano 2013: 395, 402; see also Stivers & Hayashi 2010). Making use of QA sequences in the dialogue design of technical systems helps to pre-structure the exchange with the user and to better deal with the otherwise contingent and unpredictable nature of human communication. Depending on the structural 'fitting' of the answer and its contribution to the progressivity of the ongoing action, Stivers & Robinson (2006) distinguish between 'answers', i.e. specific types of responses which prototypically deal with the conditional relevance established through a question, and 'responses', i.e. anything that occurs (or not) subsequent to the previous turn. Fischer *et al*. (2019) demonstrate that this difference is particularly relevant in encounters with technical systems, like Amazon's Alexa, as these often fail to produce an adequate second turn. While this is similarly expectable for today's robotic systems, I nevertheless use the term 'answer' here, because in the QA sequences under investigation, the robot is in the position to ask the question and the human co-participants are oriented to providing the answer, even though this sometimes turns out to be difficult.

The QA sequences in dialogue and robotic systems are mostly designed for 1:1 situations. Approaches to the issue of multiple users in QA sequences are scarce, yet show the importance of this topic. Yamazaki *et al.* (2013) explored the role of the robot's gaze in shaping visitor conduct during QA sequences with a museum guide robot. They found for multiparty QA sequences that English participants produced more reactions through both gaze shift and nodding when the robot gazed more at them. Kobayashi *et al*. (2010) show that when being asked a question by a (human/robot) guide, visitors tend to look down as a strategy to avoid being selected to answer the question and thus attempt to bypass the participation role of an addressee. While these first studies point to the interactional effects of a robot's gaze conduct in multiparty QA sequences, as far as I am aware, the dynamics between the co-participants (adult-child-robot) has not been investigated so far (see Pitsch *et al*. 2017). We also do not know how fine details of emerging and shifting participation frameworks might affect the robot's role in such situations.

In human communication, a range of interactional dynamics in QA sequences have been described which are interesting for this study. Mondada (2018) observed on guided tours that the embodied practices and mobile actions through which a visitor spontaneously asks questions are treated as either as private or public actions. Stivers (2012) and Stivers *et al*. (2018) investigated medical encounters in which the physician interacts with a child and the accompanying adult, showing that when children are asked questions by a physician, they are less responsive and slower than adults (failure to respond about one third for children vs 6% for adults; the children's time delay is less than 500 ms). Stivers and her colleagues found that children's performance is associated with the communicational practices mobilized by the physician. Yes/no questions and the doctor's gaze addressed to the child appear to help the child answer. In particular, by the age of 8, children are not yet reflexive about the particular social norms involved in QA sequences. With regard to robot-adult-child constellations, these observations raise the following questions: Are children slower to produce an answer? Is the robot's question design consequential for this delay? What role do social norms play in such encounters?

CHILDREN AS PARTICIPANTS IN HUMAN-ROBOT ENCOUNTERS

Studies exploring children as co-participants in human-robot encounters are still scarce. Importantly, existing studies have focused on autistic therapy sessions (e.g. Michalowski *et al*. 2009, Robins *et al*. 2009, Scassellati *et al*. 2012) or begun to explore robots as learning companions (e.g. Belpaeme *et al*. 2012) or story narrators (Tamura *et al*. 2017). Exploratory studies investigated how children[3] perceive the robot as a living creature or a material thing in and through their interaction (Alac 2011, 2016) or how the robot's status changes during play, from an inanimate object to an animate object to a polyfunctional object with which the child can interact in different (real and symbolic) worlds (Pitsch & Koch 2010). Research on adult-child participation

[3] Future work needs to further specify the notion of 'child' as age and developmental stage has a tremendous impact on forms of participation and conduct.

around technology-rich museum exhibits which invite active manipulation of parts of the exhibit suggests that important dynamics revolve around 'doing' and 'displaying doing' (Meisner *et al.* 2007). Children appear to like assuming a specific participation status of 'intent participation', in which they are initially active observers and, according to their own timing, successively shift to becoming a full collaborator in the activity. This raises a question: which specific interactional dynamics might emerge between adults and children around participation in encounters with a museum guide robot?

STUDY & DATA: ROBOT AS REAL-WORLD MUSEUM GUIDE

As part of a larger academic research project on situated (inter-)action, multimodal coordination and human-robot encounters (see footnote 1), a humanoid robot (Aldebaran's NAO) was set up to act as guide in a real-world museum in Germany as the basis of a set of empirical studies. The museum scenario was chosen by the researchers as it offers relatively stable conditions and scalable degrees of interactional complexity. The collaborating museums (history, arts) were interested in experimenting with novel forms of providing information to visitors and the presumable divergence between traditional exhibits (historical objects, paintings) and a robotic technology.

DESIGN OF ROBOT CONDUCT: QUESTION-ANSWER SEQUENCES

The robot was set up to explain a set of exhibits to visitors. The design included a set of specific moments during which different research questions on aspects of interactional coordination and multimodality could be investigated. A set of QA sequences was implemented in the robot's explanation following the structure of question (QU), answer (AN) and acknowledgement (AC) and designed to have varying degrees of complexity. QA sequence A was based on a yes/no question, which was initiated after the robot greeted the visitors (Fig. 1). QA sequence B was based on an open wh-question, which was initiated during the explanation of a tombstone. Its design consisted of two subsequent questions and was preceded by a deictic (potentially repairable) action (Fig. 2). In both cases, the robot was set up to acoustically monitor its environment for the visitors' answers, on the basis of which it proceeded in its conduct using a simple finite-state approach. As the robot's conduct was not specifically designed for a particular user/visitor group, the question design was not specifically targeted at children.

For this study, the robot's conduct was designed to use – in addition to talk – arm movements, head orientation and movement in space as elements of a complex multimodal gestalt. However, during the QA sequences considered here, the robot's head was oriented straight ahead facing the visitors; no further arm or body movement was involved. Additionally, in QA sequence B (wh- question), a deictic sequence – during which the robot points to the exhibit – is inserted after the first realization of the question and its resumption ("it's the man on the right-hand side") during which the robot, again, is immobile with its head oriented directly to the visitors. This way, the multimodal side of the robot's conduct is stable and comparable between the two QA sequences.

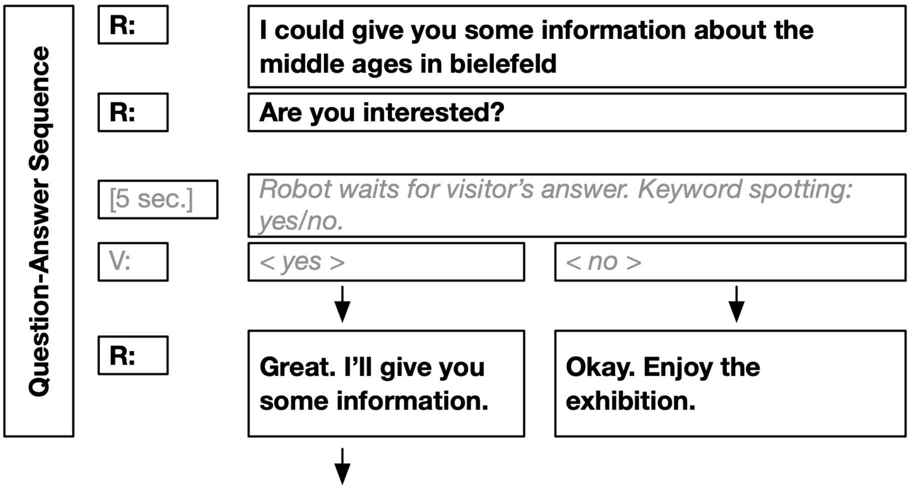


Figure 1. QA Sequence A (yes/no) "Are you interested?"
Source: author

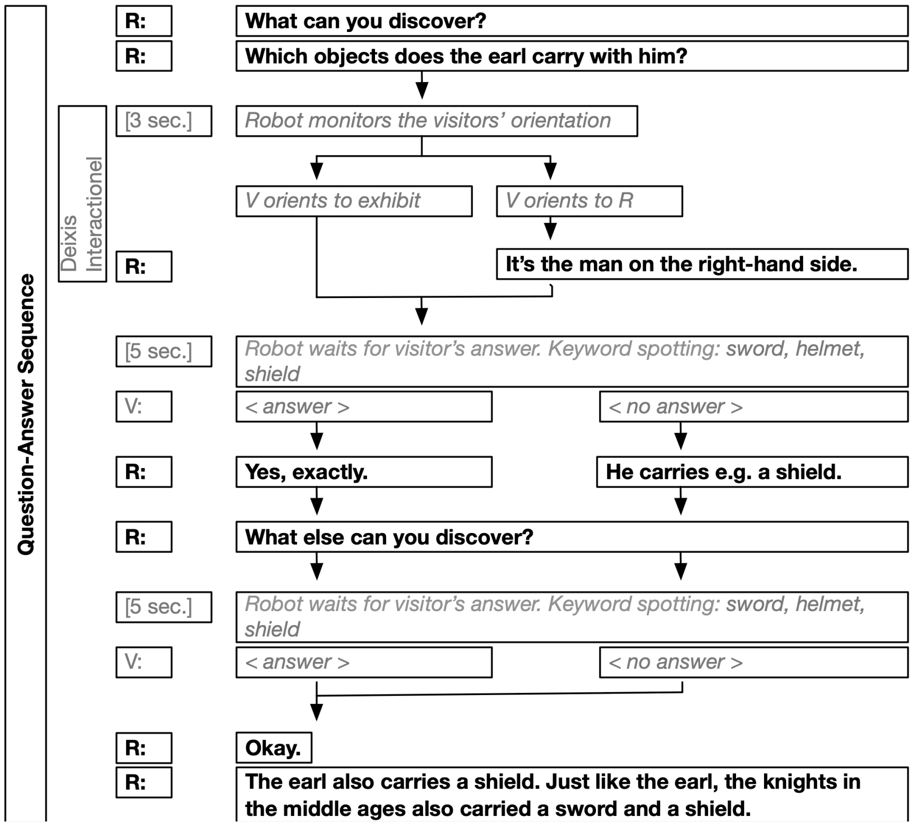


Figure 2. QA Sequence B (wh- question) "What can you discover?"
Source: author

ROBOT SYSTEM

As the robotic platform should provide intuitive access for lay users and be robust enough to be deployed in a real-world environment, a humanoid NAO robot (Aldebaran, V4, 58 cm high) was set up to run autonomously. It was based on a set of modularized system components organized by a dedicated middleware architecture. For the functions investigated here, the following components are important (see also Pitsch *et al*. 2016):

Dialogue conduct: The robot's multimodal utterances consisted of preconfigured, synchronized speech-gesture (head orientation) behaviours, which occurred in a fixed order during the robot's explanation. A dedicated dialogue component was used in combination with a finite-state machine to respond to the visitors' conduct at defined moments.

Speech recognition: To better cope with the noisy environment of a museum, the robot's onboard microphone was replaced with an additional microphone (Acoustic Magic Voice Tracker) placed under the robot's table. The speech recognition engine used was Pocket Sphinx. While the robot waited for the user's answer, a keyword spotting algorithm monitored the incoming speech for a small set of expectable words, such as 'yes', 'no' (and variants) in case A, and 'sword', 'helmet' and 'shield' in case B.

STUDY & DATA

To explore the features implemented in the robotic system and how users would attempt to deal with it, a field trial was conducted at a German historical museum. The robot was set up to act as a museum guide and was positioned in an open space alongside a set of exhibits from the medieval collection to which the robot referred during its explanation (Pitsch *et al*. 2016). To compensate for its small size (58 cm), it was placed on a table (1.20 x 2 m, 0.7 m high). It was set up to get in contact with visitors, give explanations about three different objects, walk across the table, involve visitors in small QA sequences, and close the encounter. Trials lasted for about 4 minutes. In total, 72 runs of HRI with mostly 2 to 5 participants each were recorded with 4 external HD video cameras, as were the robot's perception and log files of the system's calculations for some interactions. The fragments investigated here were selected to pinpoint specific issues around participation dynamics in QA sequences and as a basis for subsequent analysis of the larger corpus.

FACILITATING PARTICIPATION: ESTABLISHING A CHILD AS AN ACTIVE PRIMARY CO-PARTICIPANT FOR THE ROBOT

In the first analytical step, I explore the interactional dynamics which emerge when the robot addresses its first question to the visitors. After greeting the visitors by saying "Hello I am Nao" and suggesting "I could give you some information about the middle ages in Bielefeld" the robot asks: "Are you interested?" At this early stage of the encounter, it is not clear yet how best to deal with the system, expectations are being built and ways of participation in the group are emerging (see Pitsch 2015, 2016; Gehle *et al*. 2017). Analysis will show how children are interactively established as ‘primary co-participants’ for the robot and what interactional work this involves.

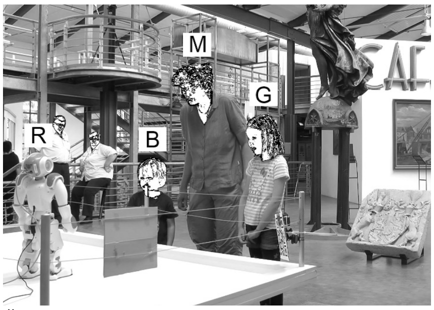
In the first fragment (4_007), we encounter three participants, a young boy (B, < 6 years old), a girl (G, 6 to 10 years old) and their mother (M) who approach the robot (R) and position themselves in front of its table. B and G are standing in the first row and M slightly behind in between them (#00.12.82). The unfolding exchange reveals that the adult plays a central role in analysing the structural provisions established by the robot, helping the child(ren) to perform the expected actions at the appropriate time. Thereby she enables them to assume the role of an active co-participant.

(a) *Answering the robot’s structural provisions and establishing the encounter as a social group activity:* Once the robot addresses the visitors with a greeting “HELlo; i am NAo” (01), M reciprocates with “HELlo; i am the <name>” (02-03). Observing this exchange, the young boy B turns to M with a bright smile and thus establishes the encounter with the robot as a social activity for the group (03, #00.13.58).

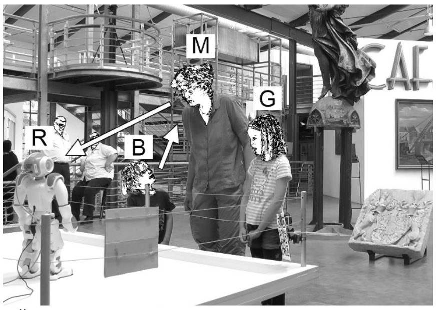
```
01     R:        HAL|lo; |ich bin |NAo;
                 HELlo;  |i   am  |NAo;
       R-act:       |NOD |
       M:                |HAllO;  |
                         |HELLO;|

02     M:        HALlo;  #
                 HELlo;
                         #00.12.82

03     M:        ich bin  #die <name;>
                 i   am    the <name; >
                           #00.13.58
```



#00.12.82



#00.13.58

(b) *Inviting the young child to assume an active role:* When the robot asks whether the visitors would like to have some information about the middle ages in the town in which the museum is located (04-06), a short moment of silence occurs during which none of the participants answers (06). During this, M taps at B's back (06, #00.20.14) which is followed by B responding "yes" (07), and G quickly adding a pronounced "YES" (08). This way, forms of participation are negotiated between the participants during which the adult identifies the relevant next actions and invites a particular participant (here: B) to contribute. The adult thus helps the child(ren) to assume an active role in the encounter with the robot.

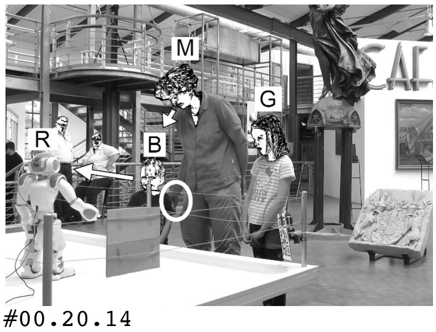
```
04    R:      |ich könnte   euch etwas     über das
              |i   could    you  something about the
      R-act:  |bras-droit-levé ...

05    R:      mittelalter in bielefeld erzählen;|
              middle ages in bielefeld tell;    |
      R-act:  ...                               |

06 QU R:      |habt ihr  |LUST dazu,  |#(1.2)            |
              |are  you   INTerested,|
      R-act:  |bras-droit-descend     |
      R-hea:             |NOD         | NOD              |
      M-act:                          |#taps at B's back |
                                       #00.20.14

07 AN B:      ja;
              yes;

08 AN G:      =JA;
              =YES;
```



#00.20.14

(c) *Backing up the children's answers:* After the children's answers, another moment of silence occurs (09). M again identifies the missing next action and, this time, herself provides a pronounced and loud "YES" backing up the children's answers (10). The robot responds to this with an acknowledgement and the continuation of its explanation (11).

```
09             (1.0)

10 AN M:       JA;
               YES;

11 AC R:       PRIma; dann mache ich das;
               GREat; then i will do that;
```

In sum, the adult plays an important role, firstly, in establishing the encounter with the robot as a social activity between herself and the child(ren) and secondly, in analysing the structural provisions established by the robot's conduct and helping the child(ren) to produce the expected actions at the appropriate time. Note that the adult only organizes access to the situation in terms of timing expected actions, but does not explain its nature or content. This way, the adult works to establish the boy as an active co-participant in the encounter with the robot which is observed and adopted by the girl. The children are thus established as primary co-participants with the adult assuming the role of 'participation facilitator' providing for the children's access to the situation and, in cases of doubt, backing up their actions.

A second fragment (4_008) allows us to expand on these issues. In this case, a young girl (G) and a young boy (B) both between 2 and 4 years old, their mother (M) and father (F) position themselves vis-à-vis the robot (R). The adults kneel down matching the physical height of their children (00.24.20).
(a) *The child's intuitive approach and its interactional confirmation:* Once the robot addresses the visitors with a greeting (01), the girl reciprocates with "HELlo" (03) and thus directly takes up the modus operandi suggested by the robot's actions. Her conduct is shadowed by F (04: nod + "HELlo") and is thus interactionally ratified as appropriate.

```
01    R:       HAL|lo; |ich bin |NAo;
               HEllo;  |I   am  |NAo;
      R-act:      |NOD |
```

```
02   (0.3)

03    G:       |HALlo;   |
               |HELlo;|
      G-act:   |nod      |

04    F:       |(0.3)|HALlo;
                     |HELlo;
      F-act:   |NOD  |
```

(b) *Making the relevant next action transparent:* When the robot asks whether the visitors would like some information about the middle ages (05-07), the other adult M silently nods to the robot, reciprocating the robot's previous 'nod' (l. 08, #00.24.20). Then, she turns her head to G uttering 'just say yes-' (l. 08, #00.26.02). F also turns to G with a softly uttered 'yes' and pronounced nod (l. 08, #00.26.02). The adults thus either explicitly tell the young girl how to answer the robot's question or model doing so. These suggestions are taken up by the child answering a soft 'yes' (09).

```
05    R:       |ich könnte   euch etwas         über das
               |i   could    you  something     about the
      R-act:   |arm-right-lifted ...

06    R:       mittelalter in bielefeld erzählen;|
               Middle ages in bielefeld tell;|
      R-act:   ...                               |

07 QU R:       |habt ihr  |LUST dazu, |(0.8)|
               |are  you   INTERested,
      R-act:   |arm-right-descinding    |
      R-hea:              |NOD        |NOD  |


08 AN M:       | (0.3)|<<p> sag mal   | ja-
               |      |<<p> just say  | yes-
      M-act:   |#NOD  |
      F:                              | <<p> ja; |
                                      | <<p> yes;|
      F-act:          |NOD            | NOD      |
               |#00.24.20             |#00.26.02

09 AN G:     ja
             yes
```

#00.24.20

#00.26.02

(c) *Working on the details of the relevant next action and backing up the child's answer:* The girl's softly uttered 'yes' (09) is followed by a short silence (10). After about 1 second, F whispers to her to speak "louder" (11) and M also whispers something (inaudible on the recording) to her. The girl responds with a louder and more pronounced 'YES-' (11, #00.27.23), so we see her working on the details of realizing the suggested next action. G's action is then shadowed and backed up by F who also produces a 'yes' directed at the robot. Finally, the robot responds with an acknowledgement and continues its explanation (12).

```
10    (1.0)

11 AN M:                     |<<p> (    )|
      F:     <<p> lauter- > |                  |ja-
             <<p> louder- > |                  |yes-
      G:                                 |#JA-   |
                                         | YES-  |
                                         |#00.27.23

12 AC R:     PRIma; dann mache ich das;
             GREAT; then i will do that;
```

#00.27.23

In sum, this fragment reveals a more complex and expanded version of the adults' role in helping the child assume the role of an active co-participant in the encounter with the robot. In particular, here the girl is not only told that she needs to produce an action at a specific time (as in extract 1), but the adults make transparent what type of action is expected, give her some details of how to do it, model, and finally back up the child's actions. Thus providing the seemingly simple answer 'yes' to the robot's question becomes a complex interactional endeavour. Conceptually we can distinguish between the 'answer-as-interactional-process' and the 'answer-as-result'.

NEGOTIATING PARTICIPATION: WORKING TOWARDS PRODUCING AN APPROPRIATE ANSWER FOR THE ROBOT

In a second analytical step, I explore an interactionally more complex situation of QA sequences. In QA sequence B, the robot asks wh- questions: "What can you discover? Which objects does the earl carry with him?" This leads to a negotiation between the participants about who is responsible for answering the robot and the child's careful attempts to first test a potential answer with the adult. The unfortunate result is that the robot continues its explanation without the child's utterances being either addressed or taken into account by the robotic system.

In this third fragment (4_002), a girl (G, 6 to 10 years old) and her father (F) are active co-participants in the encounter with the robot. After about 2 minutes, the robot has asked the visitors to approach an exhibit (a tombstone). The robot has walked towards the side of its table, invited the visitors to orient to the tombstone and begun to ask a question about its features.

(a) *Inviting the child to produce an answer*: After the robot has asked the question “What can you discover? Which objects does the earl carry with him?” (01-02), a moment of silence occurs during which none of the participants answers (03, #02.46.05). During this time, F reorients to G and taps her back (#02.47.07-a/b) inviting her to produce an answer.

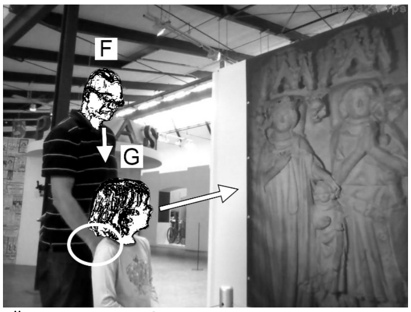
```
01 QU R:       was könnt ihr entdecken-
               what can  you discover-

02 QU R:       welche gegenstände hat der graf dabei,
               which objects does the earl carry with him,

03              |#(1.1)              |
      F-act:    | taps at G’s back   |
                |#02.46.05           |#02.47.07-a/b
```

#02.46.05

#02.47.07-a



#02.47.07-b

(b) *Reorientation to the exhibit*: To produce an answer, the participants need to closely inspect a particular exhibit, the tombstone. The robot thus attempts to establish conditions under which the participants are most likely to understand this and is programmed to add the increment “it’s the man on the right-hand side” (04). The child repositions her body to focus on the exhibit (05, #02.49.08). This action appears – on the interactional surface – to be acknowledged by the robot saying ‘yes; exactly’ (06).

```
04 Deix R:      also, der mann auf der rechten seite,
                so, it’s the man on the right-hand side,

05              |(2.0)            |
        G-act:  |turns to exhibit |
                                  |#02.49.08
06      R:      ja; genau;
                yes; exactly;
```

#02.49.08

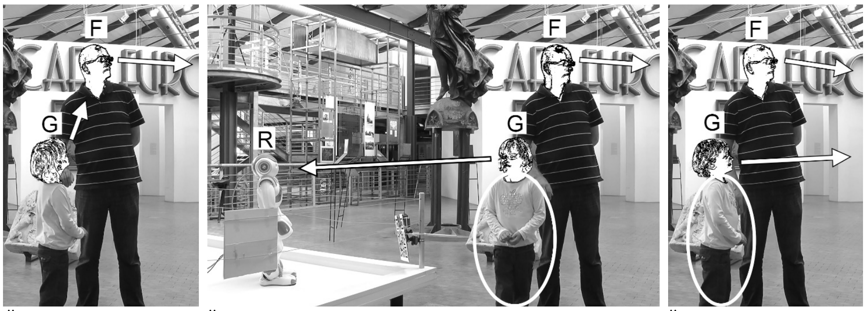
(c) *Robot rephrases the question & the participants interactionally work towards an answer*: The robot rephrases its initial question by asking “can you discover anything else?” (07). The child responds by preparing to produce an answer. She turns to F whispering a few very quiet (in the recording not understandable) words to him (09, #02.55.05), then turns back to the robot (09, #02.57.09). F reacts with a display of non-understanding ‘what,’ (10). The child then turns back to face the exhibit and explains in a whisper: “on the right-hand side” (10). This way, the participants can be seen to be actively working towards producing an answer to the robot’s questions.

```
07 QU   R:      |könnt ihr noch etwas |entdecken,
                |can   you discover anything else,
        G:      |<<p> (???)           |

08      (0.3)

09      G:      |<<p> (???)> |
                |<<p> (???)> |
        G-act:  |turns to F  |turns to R
                |#02.55.05   |#02.57.09

10     F:       was,  |
                what, |
       G:             |<<p> auf der | rechten seite >  |
                      |<<p> on the  | right-hand side >|
       G-act:         | se tourne vers l'objet          |
                                    |#02.59.07
```



#02.55.05 #02.57.09 #02.59.07

(d) *Robot continues its explanation without taking the child's work into account:* While G and F are still working towards producing an answer, the robot can be heard to answer itself and to proceed with its explanation: "the earl is also carrying a shield…" (11-13).

```
11 AN  R:       =der graf trägt auch ein schild-
                =the earl is also carrying a shield-

12     R:       so wie der graf,    trugen    im mittelalter
                just like the earl, carried in the middle age

13     R:       auch die ritter ein schwert und ein schild;
                also the knights a  sword   and a   shield;
```

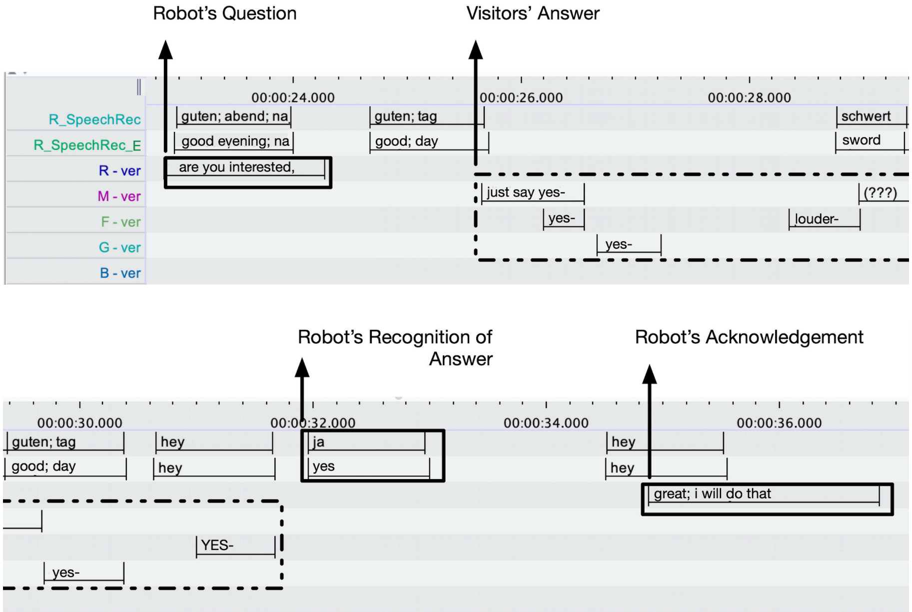
This way, despite the child's careful and motivated preparations to produce an answer, her attempts are not taken into consideration by the system. Taking up the notions introduced in the previous section, the participants were highly involved in the answer-as-interactional-process without arriving at the stage of an answer-as-result during the timeframe suggested by the robot's actions.

In sum, the emerging forms of participation between the visitors increase in complexity. They shift between negotiating who is responsible for the next turn, orienting to the material resource (i.e. exhibit) where the answer can be found and collaboratively working towards producing an answer. These dynamics between the visitors contrast with the sequential structure provided by the robot's conduct. The sequential structures and the timing of actions within the socio-technical system 'adult-child-robot' begins to come out of alignment.

ANALYZING PARTICIPATION: THE ROBOT'S PERSPECTIVE

The third analytical step is to look inside the 'black box' of the technical system. On the basis of log files recorded from the autonomous system, I consider the robot's perspective during the QA sequences. Situations when the participants are working towards producing an answer to the robot's question are of particular interest, as these constitute privileged moments for the dynamics of participation. Are the negotiations around the 'answer-as-an-interactional-process' and the 'answer-as-result' clear to the technical system based on its speech recognition and dialogue functions?

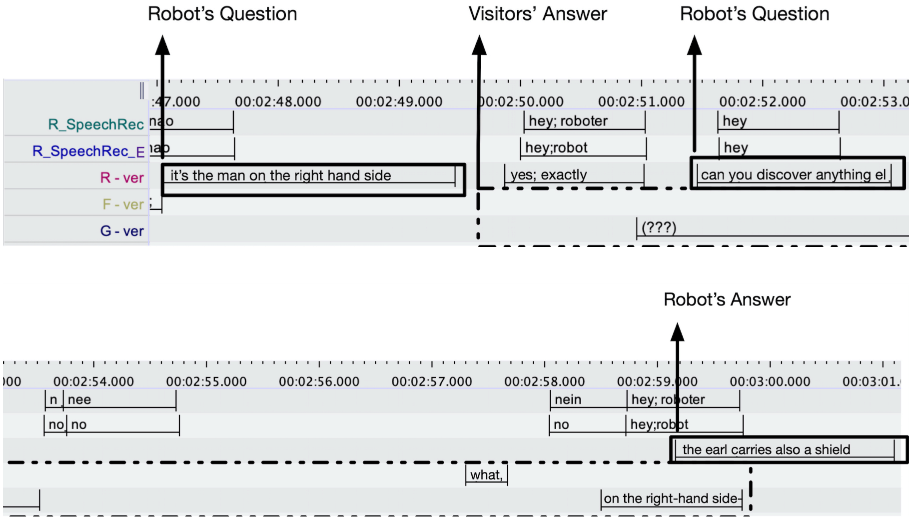
From this new perspective, let us reconsider the second fragment (4_008). The initial analysis revealed how both adults attempted to establish the child as an active co-participant for the robot by explaining both the type and content of the expected next action ('say yes') and details of how to do it ('louder') and by finally repeating the child's answer ('yes'). These activities can be found in the following timeline-based transcript from time code 25.600 to 31.800.



Source: author

Within the timeline-based annotation tool ELAN, the manual transcriptions of the participants' verbal activities are combined with the log files of the robot's internal perspective: the results of the speech recognition (Nao-speechRec). These indicate that the robot correctly recognizes a "ja/yes" (32.000) as the 'result' turn, however does not understand any previous other talk (24.00 to 32.000). Thus, the visitors' negotiations pass through silence for the robot. At the same time, the robot's speech recognition produces a range of 'false positives' (schwert/sword, guten tag/hello, hey).

In the interactionally complex situation of the third fragment (4_002), the log files reveal that the robot appears to realize that some sort of voice activity occurs after its initial question, yet its recognition as "hey nao" (02:46.800), "hey roboter/hey robot" (02:50.000), "hey" (02:51.700), "nee/nah" (02:53.800) and "nein/no" (02:58.000) are far from providing any sense of the participants' actual actions. Thus, the visitors' highly involved negotiations and the child's attempt to produce an answer which she initially suggests to her father are completely hidden to the robot.



Source: author

In sum, the moments of high interactional dynamics around establishing forms of participation in QA sequences and the co-participants' interactional work to produce an adequate answer are by and large unclear to the robot system. Instead, these privileged moments of heightened interactive involvement appear as mere 'silence' to the robot.

RESULTS & DISCUSSION

In this article, I explored how human-robot encounters work in a real-world site – a humanoid robot acting as museum guide – linking considerations from EM/CA and informatics to study situated (inter-)action and human social practices. I focused on QA sequences in which the robot addresses a question to visitors and – using speech recognition – attempts to understand user utterances as input to the system and to proceed accordingly. Like most current technical systems, the dialogue component is modelled on a 1:1 situation, yet in the real world, small groups of visitors comprised of children and adults attempt to engage with the robot. Analysis based on video recordings and system-internal log files of these encounters revealed the following observations:

(a) *Adults assume the role of a 'participation facilitator' and establish the child(ren) as primary co-participant(s) for the robot:* Visitors establish the encounter with the robot as a social activity within the group. Accompanying young children, adult(s) assume the role of analysing the structural provisions established by the robot's conduct and support the children in providing the expected next action at the appropriate time, if necessary explaining the specific type and nature of the required action, and in cases of doubt, shadowing and backing up the children's actions. This way, children are established as primary co-participants for the robot with, the adult assuming the role of a 'participation facilitator'. Note that the robot was not specifically designed for interacting with children; the 'average' visitor assumed in the design phase was an adult. The adults do interactional work to help the children participate actively in such encounters, presenting an analysis and interactional remedy of this situation. Beyond this focus on participation, it might be interesting to consider the pedagogical situation within the institutional context of the museum.

(b) *Answer-as-interactional-process is harder for robots to detect and interpret than answer-as-result:* Given the participants' interactional work involved in answering the robot's questions, we need to structurally distinguish between two types of answer: the 'answer-as-interactional-process' and the 'answer-as-result'. The former produces structural problems for the robot to interpret.

(c) *Producing an answer as interactional achievement:* Two structurally different situations of QA sequences were investigated: a simple yes/no question and a wh- question combined with the need for spatial reconfiguration to an exhibit displaying resources for finding the answer. Depending on the interactional complexity of the situation, the participants' interactional work to produce an answer differs. Under certain conditions, this work includes negotiating forms of participation, working on the appropriate orientation in space and initially addressing elements of the answer to a human co-participant before addressing the robot.

(d) *A robotic system designed for 1:1 dialogue may not be able to interpret multiparty interactions:* In the speech-based dialogue component of a robotic system for which QA sequences are designed assuming a 1:1 situation with the user, the structural and interactional difference between the 'answer-as-interactional-process' and the 'answer-as-result' is not clear. As this difference is not provided for in terms of structural expectations, the technical system can only monitor the situation for an expected keyword and has no framework for responding to any other activities. In sum, this leads to a situation in which a child is eager to answer the robot's question and the participants are jointly working towards enabling this, but the robot system can only interpret these developments as 'silence'. Thus the robot continues its explanation, and the child does not have an interactional slot in which to answer the robot.

These empirical observations raise conceptual and methodological issues with using HRI as a tool for investigating situated (inter-)action. This analysis has shown that QA sequences are a privileged moment for developing further insights into the moral order of human sociality.

(1) *Children's involvement in QA sequences:* the empirical observations reveal an analogy with the results presented in Stivers (2012) and Stivers *et al*. (2018) on children responding to questions in physician-child-adult situations. In the fragments investigated here, children appear to be slow to answer the robot's questions, and that it takes longer for them to deal with a wh- question than with a yes/no question. At the same time, the adults work as 'participation facilitators' to analyse the communicative situation and the structural provisions of expected next actions. The adults then respond in the light of their assessment of the children's ability to actively co-participate in this structure. This observation can be developed in a systematic analysis of both the adult's interpretation of structural complexity and their local procedures for supporting the child.

(2) *Participation in technology-rich environments:* the participants organize the encounter with the robot to establish the child(ren) as its primary co-participant(s). This procedure contrasts with the findings presented in Meisner *et al*. (2007) on technology-rich museum environments that children like to assume a status of 'intent participation' and then gradually shift via active observation to becoming a full participant in the activity. This difference appears to be well in line with the adults' important interactional work to help their children participate in an activity which has not been specifically designed for children.

The detailed empirical findings and their conceptualization may be used to develop the dialogue component of the technical system. For the empirical investigation presented here, the researchers adopted interactionally simple approach to QA sequences that was complex in its technical realization. While in principle aware of multiple participants in its surroundings, the robot assumed a 1:1 situation when carrying out speech-based dialogue. At the current stage of technological development and under the highly challenging circumstances of a real-world site, ‘perfect’ speech recognition results cannot be expected. Nevertheless, due to its exploratory nature and the deployment of the robot in a real-world setting, the study revealed an important set of user practices in the given situation. In subsequent steps, the following aspects should be considered:

(a) a *multimodal approach* to detect the users’ activity during QA sequences should not only be based on verbal activity but also on head and body orientation;

(b) to provide for the parallel detection of *multiple users* of the system;

(c) to be *sensitive* in the design of the QA sequence *to specific user groups*, such as adults or children, and to develop corresponding models for expectable user conduct.

CONCLUSION

In conclusion, empirical and conceptual observations have been used to show some ways in which human-robot interaction can be a methodological and conceptual tool for investigating situated (inter-)action. Fine-grained external observation based on video recordings has been linked with log files recorded from within the technical system to pinpoint discrepancies between authentic user conduct and the assumptions underlying the design of the dialogue component. The analysis presented here is an exploration to reveal the central issues. The subsequent steps are to describe and categorize these insights on the basis of a larger corpus so that the analysis can lead to a more fine-grained systematization. Following these lines, it should be possible to suggest both a conceptual basis for developing the dialogue component for multiparty situations and a fine-grained description of the interactional features that are part and parcel of human social practices.

REFERENCES

ALAC M. (2016), “Social robots: Things or agents?”, *AI & SOCIETY*, vol. 31, n° 4, p. 519-535.

ALAC M., MOVELLAN J, TANAKA F. (2011), "When a robot is social: Spatial arrangements and multimodal semiotic engagement in the practice of social robotic", *Social Studies of Science*, vol. 41, n° 6, p. 893-926.

ARMINEM I., LICOPPE C., SPAGNOLLI A. (2016), "Respecifying Mediated Interaction", *Research on Language and Social Interaction*, vol. 49, n° 4, p. 290-309.

BELPAEME T., BAXTER P.E., READ R., WOOD R., CUAYÁHUITL H., KIEFER B., RACIOPPA S., KRUIJFF-KORBOAYOVÁ I., ATHANASOPOULOS G., ENESCU V. (2012), "Multimodal child-robot interaction: Building social bonds", *Journal of Human-Robot Interaction*, vol. 1, n° 2, p. 33-53.

BOHUS D., HORVITZ E. (2010a), "Computational Models for Multiparty Turn Taking", https://www.microsoft.com/en-us/research/publication/computationalmodels-for-multiparty-turn-taking/, retrieved 27/07/2020.

BOHUS D., HORVITZ E. (2010b), "Facilitating multiparty dialog with gaze, gesture, and speec", *International Workshop on Machine Learning for Multimodal Interaction*, ICMI-MLMI 2010, Nov. 2010, Beijing, China, p. 1-8.

BOHUS D., HORVITZ E. (2011), "Decisions about turns in multiparty conversation: from perception to action", *Proceedings of the 13th international conference on multimodal interfaces,* Alicante, Spain, Nov. 2011, p. 153-160.

BRAUN-THÜRMANN H. (2002), *Künstliche Interaktion. Wie Technik zur Teilnehmerin sozialer Wirklichkeit wird*, Opladen, Westdeutscher Verlag.

BREAZEAL C. (2003), "Toward sociable robots", Robotics and Autonomous Systems, vol. 42, n° 3-4, p. 167-175.

BREAZEAL C., DAUTENBAHN K., KANDA T. (2016), "Social Robotics", *in* B. Siciliano, O. Khatib (Eds.), Springer Handbook of Robotics, p. 1935-1971.

BUTTON G. (1990), "Going Up a BlindAlley. Conflating ConversationAnalysis and Computational Modelling", *in* P. Luff, N. Gilbert,

D. M. Frohlich (Eds.), Computers and Conversation, San Diego, Academic Press, p. 67-90.

DAUSENDSCHÖN-GAY U., KRAFFT U. (2009), “Preparing Next Actions in Routine Activities”, *Discourse Processes*, vol. 46, n° 2, p. 247-268.

DOURISH P. (2001), *Where the Action Is. The Foundations of Embodied Interaction*, Cambridge, MA, MIT Press.

FISCHER K. (2006), *What computer talk is and isn’t: Human-Computer Conversation as Intercultural Communication*, Weinberg, AQ-Verlag.

FISCHER K. (2016), “Robots as Confederates: How robots can and should support research in the humanities”, *in* J. Seibt, M. Nørskov, S. S. Andersen (Eds.), *What Social Robots Can and Should Do*, Proceedings of Robophilosophy 2016 / TRANSOR 2016 (p. 60-66), IOS Press. Frontiers in Artificial Intelligence and Applications, vol. 290, https://doi.org/10.3233/978-1-61499-708-5-60, retrieved 27 04 2020.

FISCHER J. E., REEVES S., PORCHERON M., SIKVELAND R. O. (2019), “Progressivity for Voice Interface Design”, *1st International Conference on Conversational User Interfaces (CUI’19)*, New York, 10 p.

FROHLICH D., LUFF P. (1990), “Applying the technology of conversation to the technology for conversation”, *in* P. Luff, D. Nigel, D. M. Frohlich (Eds.), *Computers and conversation*, p. 187-220.
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”, *Ro-Man 2015*, Kobe, p. 407-412.

GEHLE R., PITSCH K., DANKERT T., WREDE S. (2017), “How to Open an Interaction Between Robot and Visitor? Strategies to Establish a Focused Encounter in HRI”, *HRI 2017*, Vienna, p. 187-195.

GOFFMAN E. (1981), “Footing”, in *Forms of Talk Philadelphia*, University of Pennsylvania Press, p. 124-159.

GOFFMAN E. (1983), The Interaction Order, *American Sociological Review*, vol. 48, n° 1, p. 1-17.

GOODWIN C., GOODWIN M. H. (2004), “Participation”, *in* A. Duranti (Ed.), *A Companion to Linguistic Anthropology*, p. 222-244.

HAYANO K. (2013), “Question Design in Conversation”, *in* J. Sidnell, T. Stivers (Eds.), *The handbook of conversation analysis*, Malden, Wiley, p. 395-414.

HEATH C., LUFF P. (2018), The Naturalistic Experiment. Video and Organizational Interaction, *Organizational Research Methods*, vol. 21, n° 2, p. 466-488.

HERRING S. C. (Ed.). (1996), *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*, Pragmatics and Beyond series, Amsterdam, John Benjamins.

HIRSCHAUER S. (1999), “Die Praxis der Fremdheit und die Minimierung von Anwesenheit. Eine Fahrstuhlfahrt”, *Soziale Welt*, vol. 50, n° 3, p. 221-245.

KOBAYASHI Y., SHIBATA T., HOSHI Y., KUNO Y., OKADA M., YAMAZAKI K. (2010). ““I will ask you”. Choosing Answerers by Observing Gaze Responses using Integrated Sensors for Museum Guide Robots”, *in Ro-Man 2010*, Viareggio, p. 687-692.

KOBAYASHI Y., YAMAZAKI K., YAMAZAKI A., GYODA M., TOMOYA T., KUNO Y., SEKI Y. (2012), “Care Robot Able to Show the Order of Service Provision Through Bodily Actions in Multi-Party Settings”, *CHI 2012*, Austin, Texas, p. 1889-1894.

KRUMMHEUER A. (2011), “Künstliche Interaktionen mit Embodied Conversational Agents, Technikfolgenabschätzung”, *Theorie und Praxis*, vol. 20, n° 1, p. 32-39.

KUNO Y., SADAZUKA K., KAWASHIMA M., YAMAZAKI K., YAMAZAKI A., KUZUOKA H. (2007), “Museum Guide Robot Based on Sociological Interaction Analysis”, *CHI 2007*, San Jose, p. 1191-1994.

LUFF P., HINDMARSH J., HEATH C. (2000), *Workplace Studies: Recovering work practice and informing system design*, Cambridge UK, Cambridge University Press.

MEISNER R., VOMLEHN D., HEATH C., BURCH A., GAMMON B., REISMAN M. (2007), “Exhibiting Performance. Co-participation in science centres and museums”, *International Journal of Science Education*, vol. 29, n° 12, p. 1531-1555.

MICHALOWSKI M. P., SIMMONS R., KOZIMA H. (2009), “Rhythmic attention in child-robot dance play”, *RO-MAN 2009*, Toyama, Japan, p. 816-821.

MONDADA L. (2018), “Questions on the move. The ecology of question-answer sequences in mobility settings”, *in* A. Deppermann, J. Streeck (Eds.), *Time in Embodied Interaction. Synchronicity and sequentiality of multimodal resources*, Amsterdam/Philadelphia, Benjamins, p. 161-202.

PITSCH K. (2015), “Ko-Konstruktion in der Mensch-Roboter-Interaktion. Kontingenz, Erwartungen und Routinen in der Eröffnung”, *in* U. DausendschönGay, E. Gülich, U. Krafft (Eds.), *Ko-Konstruktionen in der Interaktion. Die gemeinsame Arbeit an Äußerungen und anderen sozialen Ereignissen*, Bielefeld, Transkript, p. 229-258.

PITSCH K. (2016), “Limits and Opportunities for Mathematizing Communicational Conduct for Social Robotics in the Real-World?”, Towards enabling a Robot to make use of the Human’s Competences », *AI & Society*, vol. 31, n° 4, p. 587-593.

PITSCH K. (2019), “Referential practices. Human-Robot-Interaction as Methodological Tool to Investigate Multimodal Interaction”, in *Presentation at the workshop, Interacting with Social Robots and Virtual Agents?, at MuC 2019*, Hamburg.

PITSCH K., DANKERT T., GEHLE R., WREDE S. (2016), “Referential practices. Effects of a museum guide robot suggesting a deictic ‘repair’ action to visitors attempting to orient to an exhibit”, *Ro-Man 2016*, New York, p. 225-231.

PITSCH K., DANKERT T., GEHLE R., WREDE S. (2017), “Interactional Dynamics in User Groups. Answering a Robot’s Questions in Adult-Child Constellations”, *HAI 2017*, Bielefeld, p. 393-397.

PITSCH K., GEHLE R., WREDE S. (2013), “Addressing Multiple Participants: A Museum Robot’s Gaze Shapes Visitor Participation, in

Workshop 'Robots in public spaces, Towards multi-party, short-term, dynamic human-robot interaction'", *ICSR 2013*, 4 p.

PITSCH K., KOCH B. (2010), "How infants perceive the toy robot Pleo. An exploratory case study on infant-robot-interaction", in *New Frontiers in Human-Robot Interaction, Symposium at the AISB 2010 Convention 2010*, Leicester, UK, p. 80-87.

PITSCH K., VOLLMER A.-L., 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", *Interaction Studies*, vol. 14, n° 2, p. 268-296.

PITSCH K., VOLLMER A.-L., ROHLFING K., FRITSCH J., WREDE B. (2014), "Tutoring in adult-child-interaction: On the loop of the tutor's action modification and the recipient's gaze", *Interaction Studies*, vol. 15, n° 1, p. 55-98.

RAMMERT W. (2008), "Where the action is: Distributed agency between humans, machines, and programs", *in* U. Seifert, J. H. Kim, A. Moore (Eds.), *Paradoxes of Interactivity. Perspectives of Media Theory, Human Computer Interaction, and Artistic Investigations*, Bielefeld, Transcript, p. 62-92.

RAMMERT W., SCHULZ-SCHAEFFER I. (2002), "Technik und Handeln. Wenn soziales Handeln sich auf menschliches Verhalten und technischeAbläufe verteilt", *in* W. Rammert, I. Schulz-Schaeffer (Eds.), *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*, Frankfurt/Main, Campus, p. 11-54.

ROBINS B., DAUTENHAHN K., DICKERSON P. (2009), "From Isolation to Communication: A Case Study Evaluation of Robot Assisted Play for Children with Autism with a Minimally Expressive Humanoid Robot*", 2009 Second International Conferences on Advances in Computer-Human Interactions,* Cancun, p. 205-211.

SCASSELLATI B., ADMONI H., (2012), "Matari M. Robots for use in autism research", *Annual Review of Biomedical Engineering*, n° 14, p. 275-294.

SCHEGLOFF E. A. (1996), "Issues of Relevance for Discourse Analysis:

Contingency in Action, Interaction, and Co-Participant Context", *in* E. H. Hovy, D. R. Scott (Eds.), *Computational and Conversational Discourse: Burning Issues: An Interdisciplinary Account*, Springer, p. 3-38.

SCHNIER C., PITSCH K., DIERKER A., HERMANN T. (2011), "Collaboration in Augmented Reality. How to establish coordination and joint attention?", *in* S. Bødker, N. O. Bouvin, W. Lutters, V. Wulf (Eds.), *ECSCW 2011: Proceedings of the 12th European Conference on Computer Supported Cooperative Work*, 24-28 September, Aarhus Denmark, p. 405-416.

SHIOMI M., KANDAT., KOIZUMI S., ISHIGURO H., HAGITA N. (2007), "Group attention control for communication robots with wizard of OZ approach", *ACM/ IEEE International Conference on Human-Robot interaction,* Washington, DC, 8 p.

STIVERS T. (2012), "Physician-child interaction. When children answer physicians questions in routine medical encounters", *Patient Education and Counseling*, vol. 87, n° 1, p. 3-9.

STIVERS T., HAYASHI M. (2010), "Transformative answers. One way to resist a question's constraints", *Language in Society*, vol. 39, n° 1, p. 1-25.

STIVERS T., ROBINSON J. D. (2006), "A preference for progressivity in interaction", *Language in Society*, vol. 35, n° 3, p. 367-392.

STIVERS T., SIDNELL J., BERGEN C. (2018), "Children's responses to questions in peer interaction. A window into the ontogenesis of interactional competence", *Journal of Pragmatics*, n° 124, p. 14-30.

SUCHMAN L. (1987), *Plans and Situated Actions. The problem of human machine communication*, Cambridge, Cambridge University Press.

SUCHMAN L. (2006), *Human and Machine Reconfigurations: Plans and Situated Actions*, Cambridge, Cambridge University Press.

TAMURA Y., KIMOTO M., SHIOMI M., IIO T., SHIMOHARA K., HAGITA N. (2017), "Effects of a Listener Robot with Children in Storytelling", *HAI 2017*, Bielefeld, p. 35-43.

WEINGARTEN R. (1989), Die Verkabelung der Sprache. Grenzen der Technisierung von Kommunikation, vol. 418, Fischer Taschenbuch Verlag.

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”, *Paper presented at the ECSCW 2007*.

YAMAZAKI A., YAMAZAKI K., BURDELSKI M., KUNO Y., FUKUSHIMA M. (2010), “Coordination of verbal and non-verbal actions in human-robot interaction at museums and exhibitions”, *Journal of Pragmatics*, vol. 42, n° 9, p. 2398-2414.

YAMAZAKI A., YAMAZAKI K., IKEDA K., BURDELSKI M., FUKUSHIMA M., SUZUKI T., KOBAYASHI Y. (2013), “Interactions between a quiz robot and multiple participants: Focusing on speech, gaze and bodily conduct in Japanese and English speakers”, *Interaction Studies*, vol. 14, n° 3, p. 366-389.

YAMAZAKI K., YAMAZAKI A., IKEDA K., LIU C., FUKUSHIM, M., KOBAYASHI Y., KUNO Y. (2016), ““I’ll Be There Next”. A Multiplex Care Robot System that Conveys Service Order Using Gaze Gestures”, *ACM Transactions on Interactive Intelligent Systems*, vol. 5, n° 4, Art. 21.

YOUSUF M. A., KOBAYASHI Y., KUNO Y., YAMAZAKI A., YAMAZAKI K. (2012), “Development of a mobile museum guide robot that can configure spatial formation with visitors”, *in* D.-S. Huang et al. (Eds.), *Intelligent Computing Technology*, Berlin Heidelberg, Springer-Verlag, p. 423-432.

YOUSUF M. A., KOBAYASHI Y., KUNO Y., YAMAZAKI A., YAMAZAKI K. (2013), “How to move towards visitors: A model for museum guide robots to initiate conversation”, *RO-MAN, 2013*, Gyeongju, South Korea, p. 587-592.

DOI: 10.3917/res.220.0113
URN: urn:nbn:de:hbz:464-20200924-155938-6

Postprint of an article published in: Réseaux, 2020/2 (No 220-221), p. 113-150.
DOI: 10.3917/res.220.0113.