Why That Nao? How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk

Hannah R. M. Pelikan and Mathias Broth

Pelikan, H. R. M., Broth, M., (2016), Why That Nao? How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk, *34TH ANNUAL CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, CHI 2016*, 4921-4932.
https://doi.org/10.1145/2858036.2858478

Why That Nao?

How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk

Hannah R. M. Pelikan
University of Osnabrück
Osnabrück, Germany
hpelikan@uos.de

Mathias Broth
Linköping University
Linköping, Sweden
mathias.broth@liu.se

ABSTRACT
This paper explores how humans adapt to a conventional humanoid robot. Video data of participants playing a charade game with a Nao robot were analyzed from a multimodal conversation analysis perspective. Participants soon adjust aspects of turn-design such as word selection, turn length and prosody, thereby adapting to the robot's limited perceptive abilities as they become apparent in the interaction. However, coordination of turns-at-talk remains troublesome throughout the encounter, as evidenced by overlapping turns and lengthy silences around possible turn endings. The study discusses how the robot design can be improved to support the problematic taking of turns-at-talk with humans. Two programming strategies to address the identified problems are presented: 1. to program the robot so that it will be systematically receptive at the equivalence to transition relevance places in human-human interaction, and 2. to make the robot preferably produce verbal actions that require a response in a conditional way, rather than making a response only possible.

Author Keywords
Human-robot interaction; recipient design; turn-taking; sequence organization; conversation analysis.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

INTRODUCTION
This paper is about how humans make sense of, and adapt to, the interactive abilities of a natural speech agent in the form of a robot. Robotics is currently developing quickly and it becomes increasingly common that humans have to deal with robots in everyday life. Humanoid robots are a special type of robot that resembles humans in its outer appearance as well as the abilities to walk, talk and see. While humanoid robots can perform various difficult activities such as playing football and standing on one leg, their verbal communicative competence is still at a basic level. As the robot's rule-governed behavior cannot be manipulated during the interaction, programmers have to assume what the future interaction will look like when designing robots for participation in interactive events. This poses a crucial challenge since the user should understand the machine's actions in the same way as intended by the designer to successfully interact with the robot [33]. When designing for spoken interaction, this becomes especially difficult, as turns-at-talk responding to a particular previous turn may be relevantly produced in many different ways. For instance, a response to a question may be more or less informative. Since most natural speech agents can only listen during specified time windows and are not able to produce sounds and listen at the same time, the designer has to predict and set specific time points at which the user will take the next turn-at-talk. In contrast to natural language user interfaces like Apple's Siri, turn-taking is not steered by the user who is pressing a button or giving a specific voice command to make the agent "listen" but when to listen is determined by the robot: it listens at specific time points that are specified by the designer and not modifiable by the user.

While human-robot interaction (HRI) has often been investigated by designing the robot in a specific way and evaluating how the different designs affect the interaction [e.g. 16,18,23,34], few studies have focused on humans and how they behave to make interaction work with the robot. Investigating how humans manage a first encounter with a robot, this paper documents the different ways in which humans adjust how they talk to the robot, based on their changing expectations over the course of the interaction. For this purpose, participants were filmed playing a charade game with Nao, a conventional humanoid robot. A crucial problem when designing a communicative intelligent system is that designers cannot know exactly how humans will interact with the system when starting to design and at the same time cannot design without some understanding of the interaction [10]. Fraser et al. [10] suggest that by simulating interaction in a Wizard of Oz paradigm

CHI'16, May 07 - 12, 2016, San Jose, CA, USA

ACM 978-1-4503-3362-7/16/05…$15.00
DOI: http://dx.doi.org/10.1145/2858036.2858478

designers can get an idea of what the interaction between human and system might look like. In real world settings however, the robot is mainly used by people who are uninformed by conversational theory. For instance, programmers decide intuitively what the robot should say when visiting school classes with it. Therefore, our robot's interactional skills were purposefully kept simple, relying on the system functionality of the robot. The data was analyzed from a multimodal conversation analysis perspective [see e.g. 11,12,28,33].

Based on the participants' conduct during their encounter with the robot, we identified some practices that humans employ to manage a successful interaction with it. We then use these findings to discuss straightforward ways in which design of natural speech agents could be improved, taking into account previous work on human interaction. Thus, our study adds to the understanding about how humans manage interaction with a robotic speech agent in real world settings and contributes to research on how turn-taking behavior in robots can be usefully implemented and improved.

RELATED WORK

After briefly introducing conversation analytic theory on some fundamental aspects of human talk-in-interaction, we move on to consider current knowledge on human-machine interaction.

Human Organization of Talk in Interaction

Sequential Organization

A central feature of human talk-in-interaction is its sequential organization, which means that turns at talk are not simply following one after another but that they project back on what has been said before and create expectations about relevant next turns. A recipient of another's turn naturally inspects it for its interactional import (this inference work is often formulated as "Why that now?" in the conversation analytic literature [32]). Speakers display their understanding of the prior turn through the design of their own contribution. This displayed understanding can then be (implicitly) approved, or alternatively repaired, by the first speaker in her or his "third turn" [32].

Recipient Design

A second important feature of talk-in-interaction is that it is designed not only with respect to the current conversational context, but also to its particular addressee. This phenomenon is referred to as "recipient design" [28] or "audience design" [5]. Humans adjust their actions to reflect and be suited for the assumed specific needs of the specific recipient. Assumptions about a recipient may concern properties such as his/her knowledge, motives and expectancies. Together, they constitute what may be called a "partner model" [4], which forms the cognitive basis of recipient design. However, as a speaker's assumptions may not be in line with the recipient's real knowledge, competence, feelings, etc., they are continuously liable to being updated in subsequent turns [2,4]. Apart from such incremental aspects of turn construction [11,25], many other aspects of talk such as word selection [25,27], and loudness [25] can also be adjusted in adaptation to the current recipient.

Turn-Taking

Humans overwhelmingly speak one at a time, with minimization of gap and overlap [28]. This social fact is interactively accomplished by a shared orientation to a set of normative rules that regulate turn-taking for conversation. If a speaker selects someone as next speaker, that participant has exclusive rights to the next "turn constructional unit", or "TCU". If no one has been selected when a current speaker arrives at a "transition relevance place" (TRP), the one who starts first to "self-select" as next speaker stands a good chance of getting exclusive rights to the next TCU. If no one self-selects, the current speaker may continue for another TCU. Turn-taking is thus a form of negotiation, and what is negotiated about, then, is the exclusive right and obligation to produce the next TCU in the emerging sequence of turns. Endings of TCUs and their corresponding upcoming TRPs are definable and projectable by syntactic, prosodic, pragmatic and embodied means [21].

Sequencing

Many turns at talk are produced as parts of "adjacency pairs", where some first pair part (FPP) makes a second pair part (SPP) "conditionally relevant" in a two-part action sequence [31]. For instance, after a first greeting there is normally a strong expectation that the recipient produces a return greeting; if no return greeting is produced, it is noticeably absent, and may be pursued by the producer of the first greeting.

Repair

Participants' practices to resolve trouble in interaction go under the heading of "repair" [31]. Problems can occur in speaking, hearing or understanding and may be treated and resolved in various ways. Self-repair is usually initiated and carried out before a next speaker's turn. Repair of another speaker's turn is generally initiated immediately after completion of the trouble source turn ("What do you mean?"), and then repaired by the producer of the trouble (other-initiated self-repair). Recipients of a trouble source turn may also both initiate and suggest repair of the trouble ("You mean X?").

Humans orient to speaking one at a time and to minimizing overlap and pauses. Thus, extended overlap and lengthy pauses constitute trouble in turn-taking. Speakers may repair this by cutting off their turn before it is finished, to later repeat or recycle it [28], or by beginning to speak in a growing pause.

Human-Machine Interaction

As in human interaction, actions can become conditionally relevant in human-machine interaction [33]. To perform an action registered by a machine, the user has to produce an adequate input that causes a state transition in order for the machine to proceed. Humans may understand the ensuing

response by the machine as an acknowledgement of their input and treat the lack of a reaction by the machine as a sign of incompleteness of their action. Repetition of the instruction may be interpreted as initiating repair (unless it is an iterative procedure). Humans tend to treat the machine's repetition as trouble in hearing and will thus repeat their previous action. If the human assumes that the problem lies in understanding, he or she may reformulate the initial action [33].

Assumptions about Natural Dialogue Systems

The design of talk-in-interaction based on assumptions that humans have about a human recipient [5,28], has been suggested to be applicable to interaction with artificial communicative partners as well [7,22]. For instance, humans are reported to adapt their utterances based on their beliefs and linguistic feedback that they receive from a robot [6,7] and in accordance with beliefs that they have about the linguistic capabilities of a computer [22]. Pitsch et al. [12] indicate that human expectations about a robot are shaped by the robot's conduct early in the interaction. A broader range of possible assumptions about the artificial conversational partner and stronger interpersonal variations than in human-human interaction have been reported for people interacting with a robotic wheelchair [6,7].

Machines as Social Actors

Whether humans orient to their machine communicative partner as a social actor is debated in the literature. Nass et al. suggest that humans treat computers as "fundamentally social" [20] and "mindlessly transfer" human social rules and expectations to computers [19]. Kiesler and Sproull [14] challenge this notion and point out that humans only treat computers "as though" they were humans and thus their behavior only resembles human social behavior.

Differential Human Behavior

Fischer suggests that humans adapt based on the feedback that they get from the robot [7]. They thus act differently, depending on whether they find the robot to be a social actor or more like a tool [6]. In a set of Wizard of Oz studies on phone calls to a flight data base Fraser et al. [10] discovered that humans change their turn-taking behavior when they think that they talk to a computer system: humans were found to allow longer silences to develop between turns with the system than when talking to flight service staff on the phone. Studying interaction with a humanoid robot that was gradually developing between sessions, Fischer and Saunders [8] found that the ways in which humans adapt converge to being more appropriate over the course of the interaction, provided that the robot provides sufficient feedback. Several studies suggest that humans align with a machine's verbal behavior by adapting a similar linguistic structure as the computer [1] or by copying gestures of an embodied conversational agent [15].

Insofar as robots have limited interactional capabilities that the human partner needs to adapt to, they have been compared to non-native speakers [6,8]. Prior expectations and goals of a native speaker may influence how interaction with non-native speakers is managed [6,25]. Native speakers have been shown to adapt to foreigners in terms of phonology (more and longer pauses, increased loudness, careful articulation, emphasis of information by stressing it), morpho-syntax (shorter utterances, less inversion, more questions) and semantics (limited lexicon, more content words, more nouns/verbs) [25]. In terms of interactional organization the choice of topic may be affected, and the interaction is often characterized by more question-answer pairs, more repetitions and increased application of embodied behavior [25]. As we will see, some of these adaptations also occur in human-robot interaction.

Human-Robot Interaction

Kiesler and Hinds [13] point out that being autonomous, fully mobile and the ability to make decisions distinguish robots from other interactive systems. They also stress that HRI should be investigated by the highly interdisciplinary HCI community, as it provides various other perspectives than an engineering one. So far research on HRI within the HCI domain has studied HRI in a variety of settings such as remote collaboration [17], teleoperation [9] and interaction with a robot museum guide [16,23,34]. Studies have focused on different communicative resources designed into robots, and point out the importance of gaze [18] and gesture [17] for easing human-robot interaction.

METHOD

As a commercial robot, Nao can be programmed for a multitude of usages and by many different kinds of users. The ways in which many, if not most, programmers design the robot's interactional abilities can therefore be assumed to be based on an ordinary understanding of how interaction works, rather than on expert knowledge. The interaction structure developed by the computer scientists in our case is characterized by an organization in full turns – with no human input registerable before their completion and human input required after them for Nao to proceed. Sound signals and glowing ears show and delimit the robot's "listening" phases and flashing eyes indicate successful speech recognition. These characteristics were purposefully kept the same in our design for the game that Nao would play with 13 humans (one at a time).

After introducing itself, the robot asks the participants for their names. Nao then asks whether they would like to play a game, and if they accept, it proceeds to explain the game. After confirmation from the participant that he or she is ready, Nao starts imitating things and animals and asks the participant to guess the terms that it just imitated. Using gestures and playing sounds, Nao imitates a plane, a horse, a flute, a saw, a clock, a monkey, a drum and a telephone. Depending on the correctness of the participant's answer, the robot then replies in different ways and proceeds to the next imitation. Finally, Nao announces the score and closes the interaction with a short closing sequence.

The Robot

A Nao robot by Aldebaran Robotics was used during the encounter (see Figure 1). This humanoid robot is 58 cm tall and has four built-in microphones that enable voice recognition and text-to-speech translation. Nao is already used in a range of institutional settings, such as elderly care, autism therapy and schools. Nao also serves as a bank assistant in a major Japanese bank.

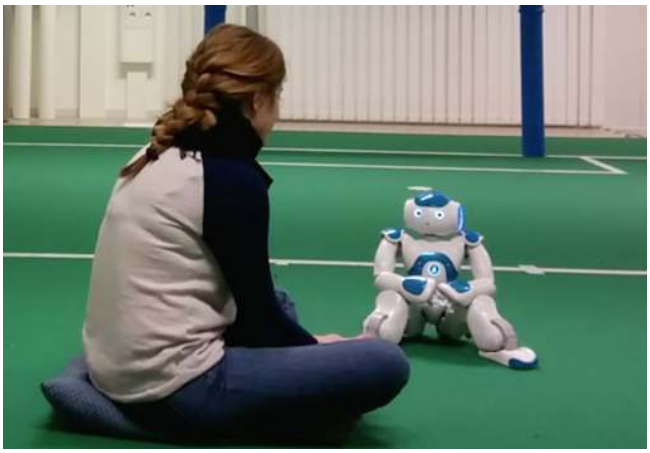

Figure 1. Sara and Nao robot before the start of the game.

After the participant was seated and cameras were switched on, the complete program was sent via Wi-Fi to the robot, which would then start to move. From this point on, the robot acted autonomously and was not controlled by the experimenter in any way.

Data Analysis

The data was analyzed from within an ethnomethodological and conversation analytic (EMCA) perspective on interaction [see e.g. 11,12,28,33]. This approach focuses on the social achievement of actions and activities, in and through the sequential organization of actions as these are produced in real time by the parties to the event. Salient patterns of practices and actions are identified by transcribing and analyzing video-recordings in detail. Thereby, the EMCA approach to interaction allows generalization from the data concerning just how specific types of actions may be achieved, without losing the empirical grounding in specific cases. As we hope to show, a close sequential analysis of how interaction unfolds allows us to understand just what resources and practices humans may mobilize and develop to deal with specific tasks and problems in interacting with a robot. The power and relevance for HCI of the methodology resides in its ability to produce analytic descriptions of what the interaction that emerges between the human and the robot really looks like, turn by turn, and just in what sequential contexts humans may run into trouble, and how they then try to address this trouble. In fact, our approach has already been suggested as a particularly suitable method for the investigation of human-robot interaction [23].

Procedure

After signing consent on video-taping, the participants were asked to sit down facing the robot. Participants were informed that the robot would take the initiative and that they should follow the robot's instruction. They did not receive specific instructions on how to interact with the robot. During the interaction, the experimenter sat behind a glass window hidden from the subjects by two big computer screens. Participants were informed that the experimenter would only come to help in case of greater trouble with the robot.

Two stable cameras were placed at different angles and distances to the scene. An external microphone was used to ensure sufficient sound quality of the recording as the moving joints of the robot caused some additional background noise.

Participants

The thirteen (5 female, 8 male) participants were all students at Linköping University in Sweden. None of them had interacted with a robot before and their interest in robotics in general was varying (3 specific interest, 4 no interest but friends/family interested in robots, 6 no interest). Similarly, their background in computer science was varying from programming on a regular basis (4), having basic programming experience (3) and no programming experience (6). Participants were from various Western countries and had different mother tongues. However, all were fluent in English.

Ethics

Participants were informed about the video recording and signed consent on the publishing of transcripts and pictures. Their names were substituted by pseudonyms according to conversation analytic standards.

ANALYSIS

Participants' initial varying assumptions about the robot are displayed in the first few turns and are then modified based on the robot's feedback. We provide a detailed analysis of the adaptation process during the course of interaction and identify difficulties that remain at the end of the game encounter.

Opening the Interaction: Initial Assumptions Displayed

In human interaction, first greetings require a return greeting [31]. The majority of the participants (10 out of 13) oriented to this requirement and immediately greeted the robot back by saying "Hello", "Hi" or "Hey". In three cases this was also accompanied by a "waving back" gesture. As they oriented to the robot's turns as making a next action on their part conditionally relevant, these participants can thereby be considered treating the robot as a form of social actor when beginning to interact with it.

The following excerpt (1) shows a typical opening sequence and the immediately following turns. In response to the robot's greeting and subsequent self-presentation, Gary produces a return greeting and states his name.

Excerpt 1. Nexus_3 [0:43-1:04] (transcription conventions: + denotes start and end of robot's embodied conduct, (n.n) silence in seconds, [a] overlapping talk, : lengthening of sound, ↑↓ intonation shifts, a stress, and >a< speedy talk)

```
01   Nao  +(0.6) hello:
     nao  +waving -->
02        (0.4)
03   Gar  >hi<
04   Nao  (0.5) i'm nao.
05        (0.8)
06   Gar  i'm+ gar[y]
     nao  -->+
07   Nao          ↑[i]'m a ro:bot
08   Nao  (0.4) an i'm four ↑years ↓old
09   Nao  (0.9) i come from fra:nce
10   Nao  (0.9) ↑what's ↓your name?
11   Nao  (0.4) da ↑dup
12   Gar  (0.7) >gary<
13   Nao  (0.9) da↓ dap
14   Nao  (0.3) nice to ↑meet ↓you (0.2) gary,
15   Nao  (1.6) i ↑love games,
```

After getting up from its pre-activation seated position on the floor, the robot initiates interaction by waving and greeting the participant (line 01). After a short pause (02), Gary takes the turn to produce a return greeting (03), the second pair part to the robot's first greeting, and thereby completes the sequence. Gary follows the normative rules that have been proposed for human turn-taking [28], which state as a first rule that a current speaker may select the next, who then has the right and obligation to speak. A common way to do so is to produce a first pair part of an adjacency pair and thereby address the next speaker as the one to continue. By answering the robot's greeting, the participant thus does not only acknowledge the conditional relevance of the robot's utterance but also displays that he assumes the robot to attend to the rules of human turn-taking. The other nine participants performing a verbal return greeting do so in a very similar way. They add the second pair part to the adjacency pair started by the robot to terminate the greeting sequence.

Nao then introduces itself by saying "I'm Nao" (04). After a slightly longer silence than when producing the return greeting, Gary reciprocates by also stating his name (06). In this way, he provides appropriate information in the sequential slot that is opening after Nao's turn. Sacks [26] points out that certain kinds of first actions make relevant particular kinds of next actions. Introducing oneself with one's name or some member's category for instance projects the same action by the other.

From a programming perspective, the robot does not perceive the information Gary produces, as it is in fact not "listening" for it in the pauses that are exploited by Gary. The robot's first uninterrupted unit was in fact programmed as "Hello (pause) I'm Nao (pause) What's your name?". However, by speaking in these unit-internal pauses, Gary displays the assumption that the robot can hear what he says as he produces a next action that would be perfectly in order in interaction with another human. What Gary says will only be perceivable by the robot once it finally ends its predefined turn in a question that asks for Gary's name and thereby selects Gary as the next speaker (10-11).

Although Gary has already provided his name, the robot's question needs to be answered appropriately in order to terminate the sequence (in interactional terms) and to cause a state transition in the robot (in programming terms). While Gary previously treated the pauses in the robot's initial speech unit as transition relevant places (TRPs), he now learns that his previous self-identification (06) was not perceived by the robot. After a brief hesitance, Gary states his name again (12). In contrast to the first time (06), he produces his name in a stand-alone fashion, not as part of a grammatical clause. He also utters it slightly faster than before. Arguably, this frames the re-giving of his name to address a problem in hearing.

Gary happens to be the only participant who immediately adds his name. However, three others also show a strong orientation to this form of implicit information request by stating where they come from. In contrast to Gary, these participants do not take the turn immediately but wait until Nao explicitly selects them as next speaker by addressing them with a question [28]. In this way, they withhold their next contribution until Nao lets them take the turn and thus orient to a larger unit in the interaction with Nao, and that coincides with the robot's preprogrammed unit.

The excerpt below (2) illustrates this practice. Sara attempted to take the turn at the two possible TRPs that Gary also oriented to in Excerpt 1. However, Sara's talk occurred in overlap with the robot's continuing talk, which she repaired by cutting herself off. When the robot finally hands over the turn to her by asking for her name, Sara produces a long answer.

Excerpt 2. BDMV_12 [1:37-1:59] (In addition to previously explained symbols, here * denotes start and end of the human's embodied conduct, … preparation and „„ withdrawal of a gesture, (.) a silence shorter than 0.2s, - cut off, = latching, h outbreath, and (h) breathiness; images are extracted at # signs in talk and numbered on a separate line)

```
01   Nao  (0.6) ↑what's ↓your name?
02   Sar  (0.3) [hh ha:- ]
03   Nao       *[d#a #↑du]p:#:
     sar       *... wave ,,,,, -->
     im           #1 #2     #3
```

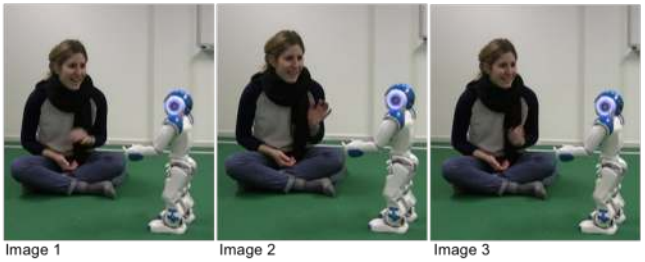

Image 1 Image 2 Image 3

```
04   Sar  (.)*(.) hi:=
     sar  -->*
05   Sar  =i'm# *s#ara↑:#**
     sar       *look face**...turn head-->
     im       #4  #5    #6
```

Image 4 Image 5 Image 6 Detail 7a

```
06   Sar  (.)+(.)#(.)+(.)*
     nao     +glw eye+ ((sign word recognition))
     im           #7
     sar    ,,,,,,,,,,-->*
07   Sar  i:['m from th]e
08   Nao    [da ↓dap:: ]
09   Sar  yu [es:]
10   Nao     [nic]e to ↑meet ↓you (0.2) sara,
11   Sar  (0.8) nice to meet you:[(hh) ]
12   Nao                          [i ↑lo]ve games,
```

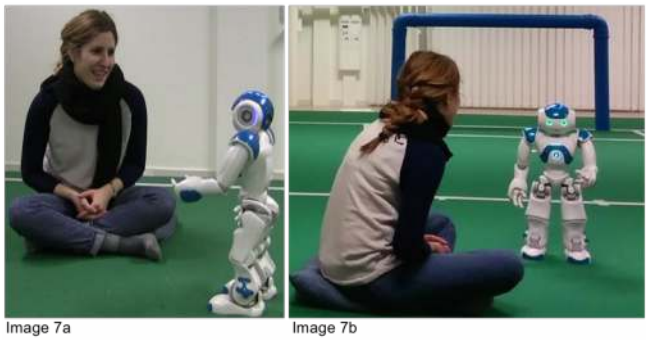

Image 7a Image 7b

When finally selected to speak by the robot (01), Sara does not immediately answer the question but tries to produce her return greeting again (02). She starts shortly after a possible completion of Nao's turn-at-talk, that is, at what would be treated as a transition relevance place in human interaction. However, this is still a bit too early for Nao, who just then is ending its current turn by producing the tone that indicates that it is now beginning to listen (03). Thus, Sara's speech overlaps with Nao's tone (02-03). Sara repairs this trouble by once more cutting off her talk and by withdrawing her small waving movement (03, img. 1-3). After a short pause, she restarts and finally manages to produce her full return greeting (04) and self-presentation (05).

By repeating the return greeting until it is produced without overlap, Sara orients to her overlapped talk as not being properly heard or understood by the robot [29]. Her repeated attempts to return the greeting display a strong orientation to the conditional relevance of producing this action. We also notice that, in her response, Sara reciprocates Nao's waving gesture, and thereby seems to embody her return greeting in a relevant way.

Sara goes on to provide an answer to Nao's question, and thereby complies with the conditional relevance of providing a second pair part of a type that is projected by the first pair part [31]. She prosodically marks the last syllable of her name by rising intonation. Peaks in pitch can indicate the end of the current turn and that a transition relevance place is up-coming [30]. Thus, rising intonation is a way to support the negotiation of the next speaker. While looking down during the first part of her turn (05, img. 4), Sara is visually orienting to the robot's face when uttering her name (img. 5). Producing the second syllable of her name, she also starts tilting her head towards her left shoulder and moves slightly forward while fixing her gaze on the robot's face (img. 6). As already Sacks et al. [28] point out, speaker change may be projected by addressing the other through gaze. By bringing a syntactic structure to its projectable end, prosodically emphasizing this end and visually orienting to Nao's face (eyes), Sara displays that she is ready to end her turn. Tilting her head could be a means to take a "closer" look at Nao's face, searching for an embodied cue for what to do next. Just when she starts to move her head straight again, Nao's eyes begin to glow bright green (06, img. 7b). This is a signal generated by the speech recognition module to show that a word (in this case Sara's name) was recognized. However, Sara is reacting to the signal with a slight smile (img. 7a) and then adds more words to her turn, thus rather orienting to it as a signal to proceed. She stretches the word "I" (07), seeming to still hesitate whether she could add a second turn constructional unit to her turn. Since Nao is still not taking the turn at this moment, she continues saying "I'm from the US". Just like Gary in Excerpt 1, Sara treats Nao's previous self-presentation as an implicit information request and provides information about where she is from. Projecting a certain response by stating one's name or category is thus a form of making a certain next action relevant, but without explicitly asking for it. Providing the expectable next action, both Sara and Gary display an orientation to the conditional relevance introduced by Nao's turns.

Interestingly, Sara is *not* orienting to the sound signals that indicate whether or not Nao is listening, and that overlap with her turn in progress (07-08). By continuing her turn she displays that she does not treat the signal as relevant for the ongoing turn-taking. In contrast, she does comply with the rule to minimize overlap [28] by stopping her turns when Nao starts speaking (09-10 and 11-12). While she clearly orients to resources used in human turn-taking (verbal, prosodic and embodied), she does not treat the sound signal as a relevant part of Nao's turns.

In the examples presented so far, participants were seen to approach the robot under assumptions pertaining to ordinary human interaction. In total 10 out of 13 participants provided a return greeting. Additionally, four participants oriented to the opportunities for taking a turn during silences following possible completions in the midst of the robot's preprogrammed turn, treating these moments as transition relevance places, as indeed would be in human interaction. The strong orientation to the conditional relevance of second pair parts by almost all participants stresses how first pair parts work as a means to project speaker change in human interaction. The difficulties encountered suggest that it is difficult for people to understand when the robot's preprogrammed speech unit ends and when it is appropriate for the human to take the turn. In fact, in human

interaction, "Hello" can either work as a stand-alone first pair part making a return greeting conditionally relevant, or as merely a first element in a more extended turn that together works as the FPP, as in "Hello I'm Nao, what's your name?". Which one it becomes is an interactional achievement on specific occasions, depending on prosodic and embodied cues and just where the recipient actually takes the turn. Further, as exemplified in Excerpt 2, eight out of 13 participants do not treat the sound signals of the speech recognition as displaying the end of the unit.

Continuing Interaction: Adapting to Nao's Capabilities

During the course of the interaction with the robot, participants clearly adjust their turn design with respect to what they progressively discover, in interacting with it, the robot's needs and capabilities to be. The adaptation process is occasioned by individual trouble and success and may be carried out in varying ways. Participants were observed to modify word selection, turn length and prosody after just a few exchanges with the robot.

The following excerpt (3), exemplifies loudness adaptation as a form of changing prosody. Mateo is waiting for the robot to proceed after stating his name and needs help from the experimenter to speak up.

Excerpt 3. BDMV_9 [1:21-1:57] (°a° denotes silent speech, A loud speech, and brackets uncertain hearing)

```
01   Nao  ↑what's ↓your name?
02   Nao  (0.3) da ↑dup::
03   Mat  (0.3) °mateo°
04        (9.8)*(0.3)#(0.9)#(1.0)#(0.2)*(1.7)
     mat       *inspect ear          -->*
     im                 #1    #2    #3
05   Exp  °(i think that)°
06   Exp  you need to speak up a bit
07   Mat  (0.6) >mA<TEo
08   Nao  (0.5) da↓ dap::
09   Nao  (0.2) nice to meet you (0.2) mathew
10   Nao  (1.4)     i ↑love games
11   Nao  (0.4) ↑would you ↓like
12   Nao  to play a game with me:?
13   Nao  (0.2) da ↑dup::
14   Mat  (0.2) Y:Es
15   Nao  (0.8) da ↓dap::
16   Nao  (0.7) nice
17   Nao  (0.4) let's start
```

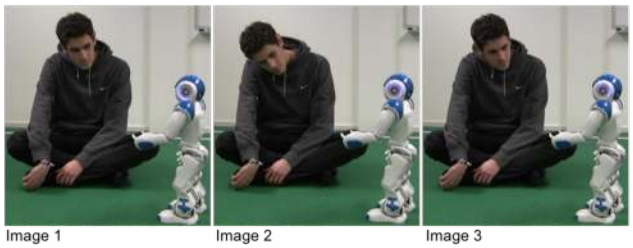

Image 1 Image 2 Image 3

When Nao asks Matteo for his name, he also states it, soon after the robot's listening signal (03). However, the name is produced too silently to be perceived by the robot, and as it was programmed to listen until it receives some input to the question, the robot does nothing. Since Mateo stays silent for a long time, almost 14s (04), during which he inspects the robot's ears, the experimenter provides help, telling him to speak up (05-06). Mateo then repeats his name with increased loudness (07). This time the name is perceived by the robot, which can then proceed in the program. Following the robot's next question, Mateo produces his response in a similarly loud way (14), thus now orienting to the robot's limited auditory capabilities.

At first, Mateo does not treat Nao's inactivity during its 14s of silence as indicating any incompleteness of his prior action and that would require repair [33]. Instead, he is leaning forward and looking at the robot's head, inspecting the robot's face and the rotating lights in its "ears" (img. 1-3). Whereas the latter were designed to indicate continuous listening for information, Mateo rather treats the rotating light in the robot's ears as a signal of processing within the robot, and not as calling for further actions on his part. Thus, Mateo orients to the robot's embodied signals to try to understand how to continue the interaction. However, he fails to interpret them in a correct and relevant way. Only when complying with the experimenter's suggestion to speak up does he treat the robot's inactivity as indicating an insufficient response.

Six other participants encountered similar trouble and had to learn to adjust their speech amplitude in a similar way to continue the interaction with Nao. Interestingly, they all permanently changed the loudness of their turns, and thus adapted to the robot's limited hearing capabilities. As Mateo in this example adjusts his loudness not only locally but also during the remaining parts of the encounter, he redesigns his turns with respect to the robot based on his early interactional experiences with Nao.

All studied participants end up producing very short turns by the end of the interaction. While participants like Sara (see Excerpt 2) produce rather long turns initially, which are then gradually reduced, Mateo keeps his turns short from the very beginning and thus does not need to modify this feature of turn design.

Over the course of the encounter with Nao, participants also adapt in terms of word selection. Before proceeding to the game, participants have to answer two yes-no questions. As the robot was programmed to only accept "yes" as a positive response, humans who used other forms, such as "sure" or "of course", soon ran into trouble.

In excerpt 4 below, Jessica tries to answer the robot's question whether she is ready to play.

Excerpt 4. BDMV_5 [2:09-2:17]

```
01   Nao  are you ready?
02   Nao  (0.2) da ↑d[up::]
03   Jes             [yeas]:
04   Jes  (0.3) °(h)hh° (1.4)
05   Jes  ye:s, i'm ready
06   Jes  (.) e(h)hh
07   Nao  (0.8) da↓ dap::
08   Nao  (0.7) goo:d
```

After a short pause, Jessica provides a first positive answer to the robot's question, thus avoiding a lengthy gap between

the two turns at talk (03). However, her "yes" overlaps with the robot's sound signal, which from the robot's perspective ends its turn (02-03). Jessica nevertheless orients to the possibility that Nao could have registered her contribution, because, after her turn, she demonstrably waits for a reaction on from Nao (04). Just like Sara (excerpt 2) Jessica does not treat the tone signal as part of Nao's turn. It is only after a significant silence that she orients to a possible incompleteness of her action [33], which she then tries to repair by modifying her turn. She produces a more proper "yes", to which she also adds "I'm ready" (05). This works as an appropriate response for the robot, and it proceeds with the next action.

Like other participants encountering this problem (eight out of 13), Jessica is trying to produce the relevant action by changing the design of her turn. In most cases this adaptive strategy is necessary for the robot to understand, and thus very useful. However, in this case the first agreeing "yes" would probably have been sufficient for the robot to understand if produced when it is listening, i.e. after the sound signal. Thus, this example also shows how humans treat verbal units spoken by Nao as turn constructional units, ending in transition relevance places where taking the turn is a relevant thing to do as soon as such a TCU has been completed. Clearly, it is difficult for participants to learn that Nao's turns always end with a sound signal. Since Nao accepts her answer the second time around (05-08), Jessica could understand the cause for trouble here as related to word selection. She does not seem to draw inferences about the trouble in turn-taking, which would be the main source of trouble here.

Understanding that Nao, unlike human interactional partners, is not prepared for responses at TRPs and is only listening when this is indicated by a sound signal and rotating lights in its "ears", is not easy. Not treating the sound signals as part of the robot's turns, participants tend to speak before the robot is ready to listen, which causes trouble in robot aural perception. The robot signals this kind of trouble by staying in the "listening" mode or by repeating its previous turn verbatim. Generally, participants encountering turn-taking trouble like Jessica often make adjustments in turn design such as changing word selection or turn-length. Since this works as a sufficient repair, the underlying trouble in hearing is occluded in most cases. Instead of dealing with trouble as related to hearing by repeating the utterance, participants treat the trouble as due to understanding problems, which they try to resolve by reformulating their utterances.

End of the Interaction: Remaining Trouble

While reformulating utterances when faced by Nao's repetitions and silences earlier in the interaction, participants learn to repeat their turns verbatim towards the end of the game. This suggests that they learn to treat repetitions and silence as indicating trouble in hearing. However, even towards the end of the interaction, four out of 13 participants do not orient to the speech recognition sounds as parts of the robot's turns and thus encounter turn-taking trouble.

Excerpt 5 shows a typical case of a participant repairing trouble by repeating the turn in a slightly modified form.

Excerpt 5. BDMV_12 [4:15-4:41]

```
01   Nao  (0.5) what am i?
02   Nao  (0.3) [da] ↑d[up::]
03   Sar        [ a]   [ clo]ck?
04   Sar  (1.4) *#a cl↑Oc#k?*
     sar        *lean forwrd*
     im         #1      #2
05   Nao  (0.6) da ↓dap::
06   Nao  (0.9)+(2.5)
     nao       +clapping--> ((clapping sound))
07   Nao  c*on↑gratulations (0.3) *+
     sar   * align with clapping  *
     nao                        -->+
08   Sar  (.) °th↑ank ↓you°
09   Nao  (0.5) you are doing goo*:d
     sar                          *raise brows -->
10   Nao  (0.5) ↑h*ere comes another one
     sar       -->*
```

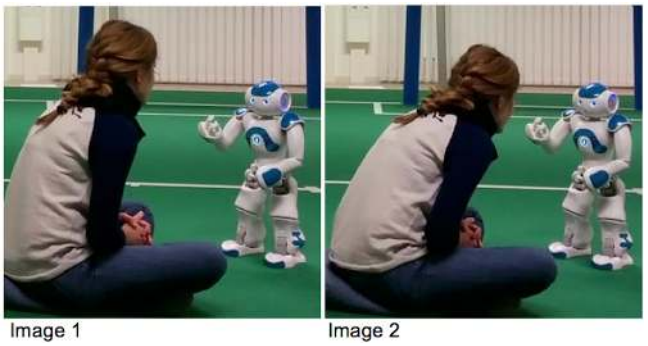

Image 1 Image 2

Sara speaks too soon and her talk overlaps with the sound signal (02-03). This suggests that she has not fully understood that the robot has not yet finished its unit-in-progress. However, she is more confident about the kind of trouble she is encountering and simply repeats her previous turn when the robot stays silent (04). Her slight prosodic modifications work as an emphasis of the word "clock" and display that she is treating the trouble as trouble in hearing. This is also underlined by her leaning towards the robot's "ears" (04, img. 4-5), as leaning closer to the robot also works as a means to overcome the slight "deafness" of the robot. Several other participants that encounter trouble with loudness also apply this technique.

Sara's repair is sufficient to satisfy the robot's needs, as Nao confirms by clapping and nodding. Sara aligns with the clapping and copies the robot's movement for a short while. Seven participants align with the movements of the robot that signal approval or rejection of the guess, which can be interpreted as orienting to the robot rather as a social actor than a mere tool [6]. When the robot utters "congratulations", Sara waits until the robot is finished and then answers "thank you" (08). As her speech is relatively silent, it is not completely clear whether she is actually addressing the robot or rather talking to herself. Three other participants also respond to the robot's compliments, even if they

reported afterwards that they knew that the robot could not hear them. So still towards the end of the interaction, some participants treat the robot's utterances as making certain next actions relevant. They produce those actions in the silences within the robot's units, thus treating these silences like transition relevance places in human interaction.

While Sara still has trouble with overlapping turns, she reduces her turn size and uses a simplified vocabulary. She is employing prosodic means to emphasize her responses in a suitable way for Nao and now knows how to treat trouble in hearing. Her assumptions about the robot's unit boundaries are not in line with the reality of the robot but she found sufficient means to repair the trouble that occurred.

DISCUSSION

This paper has explored how humans manage interaction with a humanoid robot. Displaying their initial assumptions in the first turns of the interaction, participants orient to the conditional relevance created by the robot's turns. Participants speak during silences in the robot's speech unit-in-progress, thus treating the silences like transition relevance places in human interaction. Participants are quick in adapting to the robot's needs and capabilities by modifying their turn design. As the robot only proceeds when a word is produced that it can understand, adaptation in word selection is fast and learning takes place mainly in the beginning of the interaction. When answering the yes-no questions, nine out of 13 participants learn to abandon other equivalent forms like "sure" or "of course". In the later parts of the game, they use short and simple words when answering the puzzles. Turn size generally converges towards condensed units or single words. While some participants produce short turns already at the beginning, six out of 13 participants clearly adjust in this respect during the interaction (e.g. Sara, Excerpt 2 & Excerpt 5). Seven speakers encounter trouble in making their utterances heard by Nao, and adapt after initial trouble by speaking up (e.g. Mateo, Excerpt 3) or by employing other prosodic means such as adapting pitch to emphasize important information.

In summary, turn design soon converges to short turns, simple words, and clear prosodic marking. This is in line with previous findings that show convergence of the behavior during the course of the interaction [8]. However, our findings are only partly in line with the suggestion by Fraser et al. [10] that humans accept longer gaps in the interaction with a machine than with a fellow human. Our study shows that participants orient to long pauses as turn-taking trouble, and try to deal with them by repeating their utterances. Repetition is also mobilized to repair instances of overlap. Our participants still do not display a full understanding of the differing conception of unit boundaries in the robot versus human interaction, though.

Following the omni-relevant practice of adjusting contributions to the recipient, novice participants succeed in interacting with the robot without the interaction breaking down. As suggested by research using a different robot [7], sufficient transparency of the robot's capabilities is crucial for a successful adaptation to take place. The closer participants' assumptions match with the real properties of the interactional partner, the better recipient design can be adjusted. When understanding the trouble source correctly, humans are very good at making the interaction work, as for example displayed in word selection adaptation. As trouble in turn-taking is recurring and not improving significantly in the studied encounters, we suggest that the rules and methods by which the robot manages turn-taking are not sufficiently transparent for participants to adapt. We now turn to make suggestions on how the mismatch between the boundaries of the robot's preprogrammed, and thus static units on the one hand, and the TRPs that humans perceive on the other hand can be tackled.

Design Implications

Since signaling the boundaries of the robot's units generally seems to be a crucial problem that results in overlapping turns, we suggest that more listening components in between the artificial speaker's turns be introduced. Additionally, we suggest that the sound signals used to show that the robot is listening should be dropped, as their relevance is not transparent and they rather cause trouble in turn-taking. Both these changes would make the robot's behavior more human-like. As we pointed out earlier, sequential organization with the robot is very different from interaction with natural language user interfaces, in which turn-taking is controlled by the human alone. So using a similar sound as in such dialogue systems seems to be a rather bad design choice, as humans might assume that they alone can determine when the robot should listen, thus building a partner model that is incoherent with reality. Activating the recognition at every transition relevance place would come closer to human interaction in which speakers' utterances can be heard at any time and thus speakers can take turns more freely.

We found that participants display a strong orientation to the conditional relevance of adjacency pairs throughout the interaction, even when explicitly reporting after the game that they knew that the robot did not perceive the second pair parts they produced. Thus, we suggest that adjacency pairs, as a strong means to address the next speaker [28], could be exploited for signaling the end of a unit like in human interaction and thereby, turn-taking with robotic dialogue systems may be improved. Possible time points to introduce listening components would for instance be after every first pair part of an adjacency pair or any utterance that might open a slot for a possible next action of the interaction. Setting a short time-out that is oriented at human silence duration, the dialogue system would simply continue with a proper question if the human does not treat its utterance as making a next action conditionally relevant.

As a consequence, the robot's predefined units would become much shorter. Designing robots in a way that displays their capabilities is reported to ease the adaptation

[7], especially when doing so early in the interaction [24]. Therefore, introducing more listening opportunities and thus shortening the robot's speech units can have the positive side effect of reflecting the turn length that a dialogue system can process. This may help humans in building a more coherent partner model. Thus, listening phases in between the turn constructional units could not only make interaction more natural but also ease the adaptation to the robot recipient as it displays that the robot cannot follow a long story.

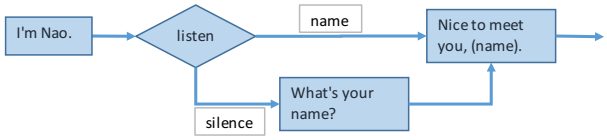

Figure 2. Example for making the interaction more dynamic

Figure 2 exemplifies this by using the first utterance that Nao produces. The sequential nature of interaction is exploited to a greater extent, allowing speakers to add information that they feel is made relevant by the robot's previous turn. The activation time of the speech recognition should be specified to be short to let the robot proceed if the human does not add information at a possible transition relevance place. It allows the robot to explicitly ask for information that has not been produced automatically but is necessary to accomplish a state transition. Similarly, the robot could be programmed to listen for return greetings or responses to compliments (e.g. "thank you"). Gestures and other embodied means may be used to display subsequent acknowledgement of the human responses by the robot.

However, as Button and Sharrock [3] point out, one should be careful when trying to derive immediate design choices using human conversation as a model to imitate. As they point out, the normative rules of human conversation cannot be formalized in a computer. Turn *design* might indeed be hard to formalize as one specific turn type (such as a positive answer to a question) can be instantiated in various ways (e.g. "yes", "sure", "of course", etc.). Our findings nevertheless suggest that taking the organization of human interaction in turn constructional units and transition relevance places into account may be crucial in designing for human-robot interaction, as our participants clearly oriented to these features in interacting with Nao. Thus, basic turn-*taking* rules can and should be considered in the design of human-robot interaction.

Limitations and Future Work

The following limitations of the current study need to be addressed in future research. First, we included participants with various different socio-demographic backgrounds and different programming experience. The reason for this was that we were interested in general practices used to successfully manage interaction with a humanoid robot. A more systematic investigation of the effect of different cultures and competencies as well as experience with natural language interfaces such as Siri, Cortana and Google Now should be investigated. Participants may find this knowledge useful when building a partner model of the robot, as may be indicated by the fact that participants with programming skills seemed to have less trouble in taking turns with Nao. This may be due to their clearer understanding of the input/output structure of the robot's programmed ability to participate in interaction.

Regardless of these limitations, this study contributes a detailed analysis of a playful encounter between humans and a Nao robot. We find that humans are able to manage the interaction, as they are adapting to the robot as a specific recipient. We provide evidence of how turn-taking mechanisms and other features of human speech directly influence interaction with a conventional robot. The found mismatch between Nao's unit boundaries and projected human transition relevance places suggests that a careful consideration of human interactional practices is essential when designing for human-robot interaction. Furthermore, the study demonstrates the value of multimodal conversation analysis for robot interaction designers.

CONCLUSION

The study has shown how humans adapt to the robot and make interaction work. Starting off with variable assumptions about the robot partner's competence, participants quickly adapt to the robot's needs and capabilities. Although individual differences persist, the participants' recipient design soon converges towards the application of simple words, the production of short turns and the appropriate adjustment of prosody. The adaptation of word selection, turn size, and prosody works relatively fast and without great effort. This implies that the human practice of designing turns with respect to a specific recipient is also useful in the interaction with an artificial partner. Adaptation to the turn-taking behavior of the robot is more difficult and problems mainly occur because the boundaries of the robot's verbal units are not transparent to the human user. The difficulties suggest that the robot's way to signal the end of a turn-at-talk could be improved. We propose that the introduction of more speech recognition blocks (thus shorter robot turn constructional units) can facilitate turn-taking. Listening blocks should for instance be introduced after first pair parts of adjacency pairs and other utterances that make a next human action relevant, as this can support the signaling of a robot transition relevance place.

ACKNOWLEDGMENTS

We thank all the volunteers, the anonymous reviewers and publications staff for providing helpful comments on previous versions of this document. We would like to express our gratitude to Barry Brown for giving feedback on an earlier version of the manuscript. Fredrik Heintz, Fredrik Löfgren and Jon Dybeck at Linköping University deserve special thanks for supporting the programming of the robot and giving access to the humanoid robots.

REFERENCES

1. Holly P. Branigan, Martin J. Pickering, Jamie Pearson, and Janet F. McLean. 2010. Linguistic alignment between people and computers. *Journal of Pragmatics* 42, 9: 2355–2368. http://doi.org/10.1016/j.pragma.2009.12.012
2. Mathias Broth, Jakob Cromdal and Lena Levin. In press. Starting out as a driver. Instructed pedal skill progression over a series of trials. In *Memory Practices and Learning. Interactional, Institutional and Sociocultural Perspectives*, Åsa Mäkitalo, Per Linell, and Roger Säljö (eds). Information Age Publishing, Charlotte, NC, USA.
3. Graham Button and Wes Sharrock. 1995. On simulacrums of conversation: Toward a clarification of the relevance of conversation analysis for human-computer interaction. In *The Social and Interactional Dimensions of Human-Computer Interfaces*, Peter J. Thomas (ed.). Cambridge University Press, New York, USA, 107-125.
4. Arnulf Deppermann. 2015. When recipient design fails: Egocentric turn-design of instructions in driving school lessons leading to breakdowns of intersubjectivity. In *Gesprächsforschung* 16: 63-101.
5. Nicholas J. Enfield. 2006. Social consequences of common ground. In *Roots of Human Sociality: Culture, Cognition and Interaction*, Nicholas J. Enfield and Stephen C. Levinson (eds.). Berg, Oxford, UK, 399-430.
6. Kerstin Fischer. 2011. Interpersonal variation in understanding robots as social actors. In *Proceedings of the 6th International Conference on Human-Robot Interaction* (HRI '11). ACM, New York, NY, USA, 53-60. http://doi.acm.org/10.1145/1957656.1957672
7. Kerstin Fischer. 2011. How people talk with robots: Reduce user uncertainty. *AI Magazine* 32, 4: 31–38.
8. Kerstin Fischer and Joe Saunders. 2012. Getting acquainted with a developing robot. In *Human Behaviour Understanding*, Albert A. Salah, Javier Ruiz-del-Solar, Çetin Meriçli, and Pierre-Yves Oudeyer (eds.). Springer, Berlin, 125-133. http://doi.org/10.1007/978-3-642-34014-7_11
9. Terrence Fong, Charles Thorpe, and Charles Baur. 2003. Collaboration, dialogue, human-robot interaction. In *Robotics Research*, Raymond A. Jarvis and Alexander Zelinsky (eds.). Springer, Berlin, 255-266. http://dx.doi.org/10.1007/3-540-36460-9_17
10. Norman Fraser, Nigel Gilbert, Scott McGlashan, and Robin Wooffitt. 1997. *Humans, Computers and Wizards: Human (Simulated) Computer Interaction*. Routledge, London, UK.
11. Charles Goodwin. 1979. The interactive construction of a sentence in natural conversation. In *Everyday Language: Studies in Ethnomethodology*, George Psathas (ed.). Irvington, New York, USA, 97-121.
12. Christian Heath and Paul Luff. 2000. *Technology in Action*. Cambridge University Press, UK.
13. Sara Kiesler and Pamela Hinds. 2004. Introduction to this special issue on human-robot interaction. *Human-Computer Interaction* 19, 1-2: 1-8.
14. Sara Kiesler and Lee Sproull. 1997. "Social" human-computer interaction. In *Human Values and the Design of Computer Technology*, Bataya Friedman (ed.). Center for the Study of Language and Information, Stanford, CA, USA, 191-199.
15. Stefan Kopp. 2010. Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. *Speech Communication* 52, 6: 587–597. http://doi.org/10.1016/j.specom.2010.02.007
16. Yoshinori Kuno, Kazuhisa Sadazuka, Michie Kawashima, Keiichi Yamazaki, Akiko Yamazaki, and Hideaki Kuzuoka. 2007. Museum guide robot based on sociological interaction analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07). ACM, New York, NY, USA, 1191-1194. http://doi.acm.org/10.1145/1240624.1240804
17. Paul Luff, Christian Heath, Hideaki Kuzuoka, Jon Hindmarsh, Keiichi Yamazaki, and Shinya Oyama. 2003. Fractured ecologies: Creating environments for collaboration. *Human-Computer Interaction* 18, 1: 51-84. http://dx.doi.org/10.1207/S15327051HCI1812_3
18. Bilge Mutlu, Takayuki Kanda, Jodi Forlizzi, Jessica Hodgins, and Hiroshi Ishiguro. 2012. Conversational gaze mechanisms for humanlike robots. *ACM Transactions on Interactive Intelligent Systems*. 1, 2, Article 12, (33 pages). http://doi.acm.org/10.1145/2070719.2070725
19. Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of Social Issues* 56, 1: 81-103. http://doi.org/10.1111/0022-4537.00153
20. Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers are social actors. In *Conference Companion on Human Factors in Computing Systems* (CHI '94), Catherine Plaisant (ed.). ACM, New York, NY, USA, 72-78. http://dx.doi.org/10.1145/259963.260288
21. Elinor Ochs, Emanuel A. Schegloff, and Sandra A. Thompson (eds.). 1996. *Interaction and Grammar*. Cambridge University Press, New York, USA.
22. Jamie Pearson, Jiang Hu, Holly P. Branigan, Martin J. Pickering, and Clifford I. Nass. 2006. Adaptive language behavior in HCI: How expectations and beliefs about a system affect users' word choice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '06), Rebecca

Grinter, Thomas Rodden, Paul Aoki, Ed Cutrell, Robin Jeffries, and Gary Olson (eds.). ACM, New York, NY, USA, 1177-1180. http://dx.doi.org/10.1145/1124772.1124948

23. Karola Pitsch, Hideaki Kuzuoka, Yuya Suzuki, Luise Süssenbach, Paul Luff, and Christian Heath. 2009. "The first five seconds": Contingent stepwise entry into an interaction as a means to secure sustained engagement in HRI. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication* (RO-MAN 2009), 985-991. http://dx.doi.org/10.1109/ROMAN.2009.5326167
24. Karola Pitsch, Katrin S. Lohan, Katharina Rohlfing, Joe Saunders, Chrystopher L. Nehaniv, and Britta Wrede. 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 the IEEE International Symposium on Robot and Human Interactive Communication* (RO-MAN 2012), 974-981. http://dx.doi.org/10.1109/ROMAN.2012.6343876
25. Jörg Roche. 1998. Variation in Xenolects. *Sociolinguistica*, 12: 117-139.
26. Harvey Sacks. 1992. *Lectures on Conversation* (Vol. 1). Gail Jefferson (ed.). Blackwell, Oxford.
27. Harvey Sacks, and Emanuel A. Schegloff. 1979. Two preferences in the organization of reference to persons in conversation and their interaction. In *Everyday Language: Studies in Ethnomethodology*, George Psathas (ed.). Irvington, New York, NY, USA, 15-21.
28. Harvey Sacks, Emanuel A. Schegloff and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 4: 696-735.
29. Emanuel A. Schegloff, 1987. Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organisation. In *Talk and Social Organisation.* Graham Button & J. R. E. Lee (eds.). Clevedon, UK, 70–85. [Originally written in 1973].
30. Emanuel A. Schegloff. 1998. Reflections on studying prosody in talk-in-interaction. *Language and Speech*, 41, 3-4: 235-263.
31. Emanuel A. Schegloff. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge University Press, UK.
32. Emanuel A. Schegloff and Harvey Sacks. 1973. Opening up closings. *Semiotica* 8, 4: 289-327.
33. Lucy A. Suchman. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, UK.
34. Keiichi Yamazaki, Akiko Yamazaki, Mai Okada, Yoshinori Kuno, Yoshinori Kobayashi, Yosuke Hoshi, Karola Pitsch, Paul Luff, Dirk vom Lehn, and Christian Heath. 2009. Revealing Gauguin: Engaging visitors in robot guide's explanation in an art museum. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09), 1437-1446. http://dx.doi.org/10.1145/1518701.1518919