Children's Turn-Taking Behavior Adaptation in Multi-Session Interactions with a Humanoid Robot

Ivana Kruijff-Korbayová
DFKI
Saarbrücken, Germany
ivana.kruijff@dfki.de

Ilaria Baroni
FSR
Milan, Italy
baroni.ilaria@hsr.it

Marco Nalin
Telbios S.p.A.
Milan, Italy
marco.nalin@telbios.it

Heriberto Cuayáhuitl
Harriot Watt University
Edinburgh, U.K.
h.cuayahuitl@gmail.com

Bernd Kiefer
DFKI
Saarbrücken, Germany
bernd.kiefer@dfki.de

Alberto Sanna
FSR
Milan, Italy
sanna.alberto@fsr.it

ABSTRACT

We present results concerning the timing of children's verbal turn-taking behavior in quiz-game interactions with a humanoid robot, spread over three sessions on different days, in one of two conditions: the robot either gave explicit signals of being familiar with the user from previous interactions, or it did not. We found that communication problems such as speech overlaps and child speech ignored by the robot are decreasing across the three sessions. Moreover, these problems are fewer and decrease faster when the robot explicitly signals familiarity with the user.

Categories and Subject Descriptors

I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems—*Natural language interfaces*

Keywords

child-robot interaction; long-term interaction; verbal behavior adaptation; turn-taking; familiarity display; dialogue management; natural language generation

1. INTRODUCTION

As social robots are getting more commonplace, it is likely that they will interact with humans over longer stretches of time. Various experiments in long-term HRI have already been carried out [12, 11, 13, 16]. These have for example attempted to identify factors that contribute to long-term engagement. In order to enable robots to engage in and sustain effective communication, it is also important to understand how humans behave in interaction with robots and how their perception of and response to robot behavior develops over time in multiple sessions. One such aspect is adaptation of verbal behavior.

Interpersonal conversation is a dynamic adaptive exchange where an interlocutor's verbal and non-verbal signals are adjusted to the conversational partner (and the situation) in a way that fosters the predictability, intelligibility, and efficiency of communication, and also manages social impressions [10, 4]. Since it is by now also well established that humans tend to treat computers as social actors and respond to them as they would to another person [19] it can also be expected that humans adapt their conversational behavior to computers. And indeed, there is growing evidence that humans adapt various aspects of their verbal and non-verbal behavior to those of the computer interfaces they interact with. Concerning linguistic adaptation, for example, experiments with text-based human-computer interaction show lexical and syntactic adaptation of users to the system [9, 3, 2]. Systematic work on speech signal feature adaptation of users in spoken human-computer interaction is also starting to emerge [18]. However, verbal behavior adaptation in HRI remains to be studied. Moreover, human adaptation to systems has so far been studied in one-shot, relatively short encounters. Persistence of adaptation across sessions has not been addressed.

In our work we investigate adaptation of children in HRI across multiple sessions. In [17] we reported that children adapt various aspects of their verbal and non-verbal behavior, including speech timing, speed and tone, verbal input formulation, nodding and gestures. In this paper we present more detailed results concerning children's verbal turn-taking adaptation.

This work is carried out in the larger context of the project Aliz-E.[1] The goal of Aliz-E is to develop the theory and practice behind cognitive robots capable of maintaining believable any-depth affective interactions with young users over an extended and (possibly) discontinuous period of time. Different strategies for achieving this goal (with children) are studied in the project[1].

2. EXPERIMENT METHOD

2.1 Participants

19 children participated in an experiment (Italian, 11 male, 8 female; age 5-12), which took place on Saturdays in March

[1] http://www.aliz-e.org/

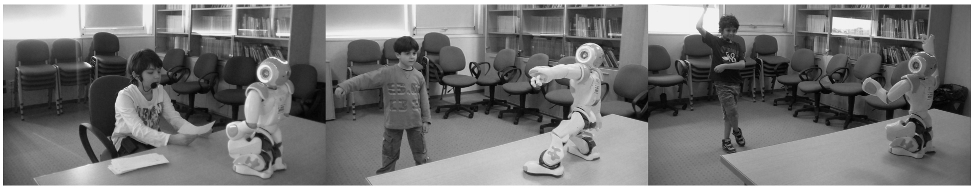

Figure 1: Children playing with the robot during the experiments. Left to right: quiz, dance, imitation.

– May 2012. 13 participants were able to come to three sessions on different days as foreseen in the protocol.

2.2 Procedure

The experiment consisted of three sessions that took place on different days one week apart. The first session consisted of a briefing about the experiment procedure and the child-robot activities, followed by filling in a pre-interaction questionnaire and then an interaction with the system featuring the activity that the child selected as main, and, time permitting, a second interaction with another activity of the child's choice. Each interaction was followed by filling in a questionnaire. The second and third sessions included no briefing, just the interaction involving the main activity and, time permiting, a second interaction with another activity, and the questionnaire(s). Each session was limited to one hour, including the questionnaire-filling time.

At the beginning of an interaction the robot greets and welcomes the child. In the very first interaction it introduces itself by name and asks for the child's name. The first time they do an activity the robot gives an explanation. Then they play. In follow-up interactions the robot greets the child, but does not repeat the name- and activity-introductions.

The child can end the interaction at any point. At certain points during the activity (e.g., end of a phase, or a game round) the robot explicitly asks whether the child wants to continue. The interaction duration is not fixed: the child may quit playing, or continue as long as it wants, up to a limit of 30 minutes (unless the interaction has to be ended earlier for technical reasons). If the child continues playing for 30 minutes, the robot apologizes that it needs to end the interaction to take some rest.

At the end of an interaction, the robot asks the child whether they liked playing, states that it enjoyed it and is hoping to play again, and gives the child good-bye. The child then fills in a post-interaction questionnaire for self-assessment of its engagement and relationship to the robot, and its opinions about the robot and the interaction.

The activities available for the children to choose from were a quiz game, a pose-sequence imitation game and a short dance-learning (Figure 1). In this paper, we present results base on the Quiz activity interactions.

In Quiz the child and the robot take turns in asking each other multiple-choice quiz questions from various domains, e.g., diabetes, nutrition, sports, geography, history, science. (The child gets a set of cards with the questions it can ask and the corresponding answers.) The asker provides correctness feedback. The asker can reveal the correct answer after two wrong attempts or upon request. The robot makes mistakes on purpose (with an answer error rate of about 30%), in order to avoid frustrating the child by a too good performance. At the end of each round the robot provides a summary of the number of correct and incorrect answers and a short evaluative comment. A round of quiz normally consists of four questions asked by the same asker, the child can however propose to switch roles at any time.

Besides activity-specific conversation, the interactions involve also a social component, such as greetings and introductions. When the robot provides performance feedback to the user during an activity, the social aspect requires careful handling of the evaluation process so as not to discourage the user with negative feedback. Preference is given to positive or encouraging comments on the child's performance. There is no comparison of the child's and the robot's performance.

2.3 Familiarity vs. Neutral Display Condition

Long-term interaction involves a series of encounters between the robot and a given user. As the robot interacts with a particular user, they become familiar with each other, i.e., they accumulate shared knowledge (a.k.a. shared history, personal common ground) [6] We can at least assume that they know each other's name, performance on a game, ways of speaking or nonverbal behaviors. Their mutual familiarity increases over time.

In the experiment we compared two versions of the system in a between-subjects design: In the *familiarity-display* (FD) condition, the robot tries to foster a sense of persistence and familiarity. It uses verbalizations explicitly acknowledging and refering to the shared history with a given user, thus showing that it is familiar with the user and remembers the previous encounters. Such verbal moves are accompanied by nonverbal behaviors showing familiarity, e.g., nodding, higher excitement. In the *neutral display* (ND) condition, the system uses verbalizations that are neutral with respect to familiarity, i.e., they do not signal familiarity. Table 1 shows examples of verbalizations from both conditions.

2.4 System

The experiment was carried using the humanoid robot Nao[2] with the HRI system described in [15, 14] (Figure 2).

We relied on a human Wizard to simulate the recognition and interpretation of the user's speech and gestures. After the Wizard selects an interpretation of the user's input in

[2] `www.aldebaran-robotics.com`

	Familiarity display (FD)	Neutral display (ND)
Use of user's name:	So, which answer do you choose, *Marco*?	So, which answer do you choose?
References to previous encounters	I am happy to see you *again*.	I am happy to see you.
and play experiences	It was nice playing with you *last time*.	–
References to previous performance:	Are you ready to play quiz *again*?	Are you ready to play quiz?
	Well done, you've done *better than last time*	Well done.
Reference to a quiz question familiarity:	The next question should sound familiar.	Now the next question.

Table 1: Examples of verbalizations in the FD and ND condition

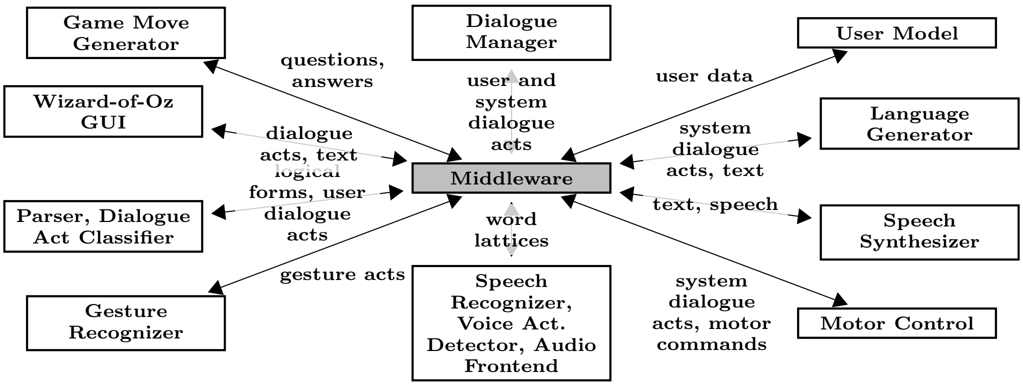

Figure 2: High-level architecture of the system used in the experiment

a GUI, the next system action is determined by the Dialogue Manager (DM), while the Wizard has the possibility to override the automatic selection if needed.

The DM carries the primary responsibility for controlling the robot's conversational behaviour. It keeps track of the interaction state, and integrates the interpretations of the user's input/actions with respect to this state. It queries and updates the game move generator and user model components, and selects the next action of the system as a transition to another state, making progress towards a goal. The next system action is selected according to a set of policies that specify a mapping from dialogue states describing situations in the interaction, to (communicative) actions [8]. The dialogue policies are learnt offline from a simulated environment partially estimated from real interaction data [7].

The dialogue act corresponding to the selected next system action is verbalized automatically by the natural language generation component. It takes as input the dialogue act and any additional relevant information from the DM and constructs text for the speech synthesizer.

To avoid repetitive verbalizations, we implemented a large range of verbal output variation. Selection among variants is either random or controlled by selection criteria. Some selection criteria refer to characteristics of the *content* to be conveyed, e.g., how many answer options a quiz question has and whether they are short or long. Other ones refer to various parameters of the *context*, e.g., the user's gender, how many quiz questions have already been asked, who is the current asker, etc. An important selection criterion is the familiarity display condition: only when the robot is to explicitly display familiarity, verbalization variants are used that include the child's name and/or explicitly refer to the interaction history (cf. Table 1). In this case utterance planning uses the information whether the current user interacts with the system for the first time or it is a subsequent encounter, whether they have already played the current game or it is new, whether the user's previous performance was good or not, etc.

In summary, these features describe the system used in the experiment: (1) Speech and gesture recognition simulated by a Wizard; (2) System action selection automatic with the possibility of Wizard override; (3) User barge-in: Interruption of the robot's speech by an early child response; (4) Automatically produced verbal output in Italian with many variations and expressive speech synthesis distinguishing sad, happy and neutral state; (5) Automatically produced head and body poses and gestures; (6) Persistent user-specific interaction profile.

2.5 Collected Data

The data collected in the experiment consists of the pre- and post-interaction questionnaires, video and audio recordings of the interactions and system logfiles.

3. ANALYSIS AND RESULTS

As we reported in [17], the children seem to adapt various aspects of their verbal and non-verbal behavior, including speech timing, speed and tone, verbal input formulation, nodding and gestures. Here we present the results of a systematic analysis of verbal turn-taking behavior using video data from all children who completed three quiz interactions. In this case we were also able to study the effect of the FD vs. ND condition. Data from N=10 children (equally distributed over FD/ND condition) were included in this analysis, a total of 9.5 hours of video material.

3.1 Data Coding

We coded *child speech segments* (CSS). Any occurrence of child speech was considered a CSS. A CSS could contain silence between stretches of child speech, as long as there is no robot's speech in between. It could be a single complete utterance or a sequence of utterances (e.g., a quiz question followed by listing the answer options), but also just an utterance fragment or a short acknowledgement or feedback. It could also be a sequence of repetitions. A CSS could be the realization of one or more dialogue moves (e.g., a quiz question plus a reguest for answer, or an acknowledgment plus the next dialogue move, etc.). CSSs were identified manually by the coders and the following attributes were coded for each CSS:

Start time The CSS onset time relative to the beginning of the quiz interation.

Timing An abstract characterization of the timing of the CSS w.r.t. the robot's speech. This attribute has three possible values:

Overlap The child and the robot speak simultaneously (at some point) during the given CSS. Overlaps are coded irrespective of which interlocutor started speaking first.

Forced The child clearly waits with its speech until the robot finishes speaking, or even until the robot produces a particular prompt, for example a request for the next quiz question. The child waits, even though it does not have to, since it knows what to say next, and it could barge in. Only clear cases of the child obviously delaying its speech are coded with this value.

Timely The CSS comes in a timely fashion, resulting in smooth turn-taking (without an overlap or forced waiting). It might be that the child waits a little with their speech, but not obviously so.

Robot's reaction Whether the robot appears to take the CSS into account for its next action. This attribute has two possible values:

Ignore The CSS has no or only a partial effect on the next action of the robot, the robot carries on with the interaction as if (a part of) the CSS did not occur. This often leads to the child repeating (part of) their speech. An example of a partial effect is when the child presents the next quiz question along with the answer options, but the robot still asks for the latter. Ignoring a CSS is not a decision made by the Wizard: most of time it is a consequence of a delay in the system (from the moment when the wizard sends the command to the actual execution by the robot), the child's barge-in and thus speech overlap, or the child's move to a dialogue state out of the coverage of the system.

Not-ignore The robot's next action is a coherent continuation of the interaction given the CSS. The robot either immediately responds to the CSS (e.g., answers a quiz question), or it moves on to an appropriate next step.

Alignment Whether the child's verbal behavior aligns with the robot's expectations (i.e., the implemented strategy), in other words, whether the child adheres to the foreseen interaction script. This is an attribute derived on the basis of the other two, in order to see their combined effect. It has two possible values:

Not-aligned The CSS has problems either in timing (overlap) or in the robot's reaction (ignore), or both.

Aligned The CSS has no overlap and is not ignored by the robot.

Two independent coders (two of the authors) performed the coding. To check inter-annotator agreement the two coders coded independently the same 36 minutes of video of the same child to identify overlap- and ignore-CSSs. They reached Cohen's κ of 0.94, indicating very good reliability.

3.2 Results

Tables 2 – 5 show the distributions of the values coded in the data for the factors of timing, robot's reaction and alignment. We report the mean of each factor averaged over the 10 children, separated per session and per condition. For the analysis of the effect of growing familiarity across the three sessions and of the FD/ND condition we use two-way Analysis of variance (ANOVA), where each factor is a dependent variable and the session number and FD/ND condition are independent variables.

Timing. The relative number of CSSs with forced waiting is increasing over the three sessions ($F(2, 29)=5.185$, $p=0.032$), and it is increasing more in the FD condition ($F(1, 29)=4.570$, $p=0.021$). Furthermore, the children in the FD condition tend to force themselves to wait at least twice as much as children in the ND condition.

The relative number of CSSs with overlaps appears to decrease across the three sessions from 14.15% to 7.63% in the FD condition, and from 19.93% to 12.82% in the ND condition. While the statistical significance of this improvement between sessions is only weak ($F(2,29)=2.586$, $p=0.096$), the difference between the FD and ND condition shows higher statistical significance ($F(1, 29)=4.375$, $p=0.047$).

Robot's reaction. The relative number of CSSs ignored by the robot drops across the three sessions: from 23.05% to 9.05% in the FD condition and from 28.2% to 12.89% in the ND condition. There is statistical significance in both the improvement across sessions ($F(1, 29)=10.608$, $p=0.001$) and the difference between the FD and ND condition ($F(1, 29)=5.121$, $p=0.033$).

Alignment. Combining the above aspects, the relative number of CSSs that are aligned with the foreseen interaction script increases across the three sessions from 68.78% to 85.95% in the FD condition and from 62.16% to 79.44% in the ND condition. Also these improvements show statistical significance in both the improvement across sessions ($F(1, 29)=9.436$, $p=0.001$) and the difference between the FD and ND condition ($F(1, 29)=5.514$, $p=0.029$).

	CSS timing								
	Forced (%)			Timely (%)			Overlap (%)		
Session	1	2	3	1	2	3	1	2	3
FD cond.	04.17	10.98	15.90	81.68	77.50	76.47	14.15	11.52	07.63
ND cond.	00.94	05.03	07.83	79.13	79.07	79.36	19.93	15.91	12.82

Table 2: Distribution of CSS timing values across sessions and conditions

	Robot's reaction					
	Ignore (%)			Not-ignore (%)		
Sessions	1	2	3	1	2	3
Familiar	23.05	09.67	09.03	76.95	90.33	90.97
Non-Familiar	28.20	19.13	12.89	71.80	81.87	87.11

Table 3: Robot's reactions to child's speech

	Alignment with the dialogue managed by the robot					
	Aligned (%)			Not-aligned (%)		
Sessions	1	2	3	1	2	3
Familiar	68.78	83.40	85.95	31.22	16.60	14.05
Non-Familiar	62.16	73.07	79.44	37.84	26.93	20.56

Table 4: Alignment of chid speech segments

It is also interesting to look at the improvements in alignment across sessions. We performed both the Tukey-Kramer test for differences between means and the Scheffe test for contrasts among pairs of means, using an α=0.05 for both tests, and the result was the same: There is a significant difference between the first and the second session (Sheffe statistic 3.10, critical value 2.59; Tukey-Kramer statistic 4.384, p=0.0131), and between the first and the third session (Sheffe statistic 4.19, critical value 2.59; Tukey-Kramer statistic 5.919, p=0.0010), but not between the second and the third session (Sheffe statistic 1.09, critical value 2.59; Tukey-Kramer statistic 1.535, p=0.5322).

Child Speech Segment Rate. The number of CSSs in a session per minute appears to be decreasing from 3.13 to 2.78 in the FD condition, and from 4.71 to 3.5 in the ND condition. While the decrease across the sessions is itself not significant (F(2, 29)=2.272, p=0.125), but the difference between the FD and the ND condition is (F(1, 29)=12.511, p=0.002). Apparently, children in the ND condition produce almost 40% more CSSs than children in the FD condition.

	CSS rate		
Sessions	1	2	3
Familiar	3.13	2.89	2.78
Non-Familiar	4.71	3.85	3.50

Table 5: Child speech segment rate

3.3 Discussion

There is a clear change in the children's speech timing and their adherence to the interaction script. Whereas many synchronization problems occur in their first session with the robot, the second and third session are smoother. In particular, the children are often waiting for the robot to talk, even if they know how to continue without the robot's turn. For example, in the first interactions, having asked a multiple-choice question, the children often go on to read the list of possible answers, thus causing the robot to barge-in with the possible answers request, while in the subsequent interactions, they wait for the request from the robot before reading the list. Conversely, when the robot asks a question, children might answer straightaway in their initial interaction, again causing the robot to barge-in, whereas in the later interactions they wait for a prompt from the robot. To summarize, children seem to adapt the timing of their speech to the robot's non-adaptive dialogue strategies, so as to avoid speech overlaps. Similar channel exclusion phenomena have been observed in another study of human turn-taking in HRI: subjects waited for the robot to finish speaking before they spoke and tended to avoid simultaneous speaking after a simultaneous start [5]. While other researchers have also studied user speech timing adaptation, they focused on different aspects, e.g., user response latency decrease with practice during a single session [5] or user response latency adaptation to the system's extrovert/introvert style [18].

In the FD condition there are fewer overlaps between child and robot speech, and forced waiting of the children for the robot to speak is twice as frequent. These children are apparently more lenient with the robot when it makes speech timing mistakes. They adapt their behavior more, for the sake of smooth turn-taking.

The children also adhere more to the foreseen interaction script in the FD condition, as shown by a lower relative number of speech segments to which the robot does not react. A child's speech segment is ignored either because it is out of the currently implemented domain of interaction (e.g., the child confides about belly ache to the robot), or because the child "runs ahead" of the implemented script, and provides information that the robot did not prompt for yet in a situation where the robot's implemented strategy is not flexible enough to react to this, and thus the wizard cannot select an appropriate next dialogue move. The children in the FD condition seem more committed to respect the robot's expectations concerning the interaction script, once they understand them.

The children in the ND condition appear to produce more speech segments. What our analysis does not make clear is whether this is a difference in the amount of speech or only in the number of speech chunks. The latter could be a consequence of there being more speech overlaps in the neutral display condition, and thus the children's speech is more fragmented, and we therefore count more child's speech segments. Moreover, since these children tend to deviate more from the foreseen interaction script, resulting in a higher number of speech segments to which the robot does not react, they may (have to) repeat their input more often (until the appropriate system prompt appears).

What it is that leads to these effects is not clear yet. Data from the post-questionnaires concerning the children's relationship to the robot indicate that all the children felt a strong connection with the robot and perceived it as a peer. We observed informally that the use of the child's name in the familiarity display condition seems to catch their attention. This might result in more concentration on the interaction with the robot. We speculate that this might, consciously or not, lead to the children's higher commitment to the (efficiency of) the interaction.

The fact that the change in alignment is larger between the first and the second session than between the second and third session seems to indicate that the children adapt their behavior to the interaction with the robot quite fast, and this level of adaptation persists. It is not clear whether the children's adaptation is just a consequence of becoming trained in "the rules of the (interaction) game", or it could be linked to social aspects of the interaction, and particularly the children's perception of and interaction with the robot as a social partner. The effect of the familiarity display condition on the adaptation seems to corroborate the latter.

4. CONLUSIONS

We presented results on children's verbal turn-taking behavior in quiz interactions over three separate sessions. In particular, we analyzed the timing of the children's speech, and whether or not the robot reacted to a child's turn. We found that adaptation increases across multiple sessions. Moreover, we found that overall there is more adaptation in the condition where the robot gives explicit signals of familiarity with the child across sessions, as opposed to the condition where the robot's behavior is neutral in this respect.

Our findings have implications for the need for and the benefits of persistent memory of robots. A robot explicitly displaying familiarity thus seems to elicit more cooperation from a (young) user leading to a smoother communication. There might also be more tolerance towards such a system, despite its inevitably imperfect interaction capabilities.

5. REFERENCES

[1] T. Belpaeme, P. Baxter, R. Read, R. Wood, H. Cuayáhuitl, B. Kiefer, S. Racioppa, I. Kruijff-Korbayová, G. Athanasopoulos, V. Enescu, R. Looije, M. Neerincx, Y. Demiris, R. Ros-Espinoza, A. Beck, L. Cańamero, A. Hiolle, M. Lewis, I. Baroni, M. Nalin, P. Cosi, G. Paci, F. Tesser, G. Sommavilla, and R. Humbert. Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, 1(2):33–53, 2013.

[2] S. Brennan. Lexical entrainment in spontaneous dialogue. In *Proceedings of the International Symposium on Spoken Dialogue (ISSD-96)*, pages 41–44, 1996.

[3] S. Brennan and J.O. Ohaeri. Effects of message style on user's attribution toward agents. In *Proceedings of CHI'94 Conference Companion Human Factors in Computing Systems*, pages 281–282. ACM Press, 1994.

[4] J. K. Burgoon, L. A. Stern, and L. Dillman. *Interpersonal Adaptation: Dyadic Interaction Patterns.* Cambridge Univ. Press, 1995.

[5] C. Chao, J. Lee, M. Begum, and A. L. Thomaz. Simon plays simon says: The timing of turn-taking in an imitation game. In *Proceedings of the 20th IEEE International Symposium on Robotics - RO-MAN*, 2011.

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

[7] H. Cuayáhuitl. Learning dialogue agents with Bayesian relational state representations. In *Proceedings of the IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems (IJCAI-KRPDS), Barcelona, Spain*, pages 9–15, Jul 2011.

[8] H. Cuayáhuitl and I. Kruijff-Korbayová. An interactive humanoid robot exhibiting flexible sub-dialogues. In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Montreal, Canada*, pages 17–20, Jun 2012.

[9] Elizabeth Z. Ford. How to get people to say and type what computers can understand. *Int. J. Man-Mach. Stud.*, 34(4):527–547, April 1991.

[10] Howard Giles, Anthony Mulac, James J. Bradac, and Patricia Johnson. Speech accommodation theory: The first decade and beyond. In Margaret L. McLaughlin, editor, *Communication Yearbook 10*, pages 13–48. SAGE Publications, 1987.

[11] R. Gockey, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. C. Schultz, and J. Wang. Designing robots for long-term social interaction. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems IROS'05*, pages 1338–1343, 2005.

[12] T. Kanda, T. Hirano, and E. Daniel. Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, 19:61–84, 2004.

[13] C. D. Kidd. *Designing for Long-Term Human-Robot Interaction and Application to Weight Loss.* Ph.d. dissertation, Massachusetts Institute of Technology, 2008.

[14] I. Kruijff-Korbayová et al. A conversational system for multi-session child-robot interaction with several games. In *German Conference on Artificial Intelligence (KI)*, 2012. system demonstration.

[15] I. Kruijff-Korbayová, H. Cuayáhuitl, B. Kiefer, M. Schröder, P. Cosi, G. Paci, G. Sommavilla, F. Tesser, H. Sahli, G. Athanasopoulos, W. Wang, V. Enescu, and W. Verhelst. Spoken language processing in a conversational system for child-robot interaction. In *Workshop on Child-Computer Interaction*, 2012.

[16] I. Leite, C. Martinho, A. Pereira, and A. Paiva. As time goes by: Long-term evaluation of social presence in robotic companions. In *Proceedings of the 18th IEEE International Symposium on Robotics - RO-MAN.* IEEE Computer Society, 2009.

[17] M. Nalin, I. Baroni, I. Kruijff-Korbayová, L. Cañamero, M. Lewis, A. Beck, H. Cuayáhuitl, and A. Sanna. Children's adaptation in multi-session interaction with a humanoid robot. In *Proceedings of the Ro-Man Conference*, Paris, France, 2012.

[18] S. Oviatt, C. Darves, and R. Coulston. Toward adaptive conversational interfaces: Modeling speech convergence with animated personas. *ACM Trans. Comput.-Hum. Interact.*, 11(3):300–328, September 2004.

[19] B. Reeves and C. Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge University Press, 1996.