

Preschoolers' Engagement in Social Interaction with an Autonomous Robotic System

Phillip Lücking and Katharina Rohlfing and Britta Wrede and Malte Schilling
Center of Excellence 'Cognitive Interaction Technology' (CITEC),
Bielefeld University, P.O. Box 10 01 31, D-33501 Bielefeld, Germany
Email: mschilli@techfak.uni-bielefeld.de

Abstract—In this paper, we present an empirical study with children at the age of 4 and 5 years to reveal whether they engage with a robot in an interaction. For our analysis, we developed a score assessing the interaction level. It consists of emotional involvement, engagement and independence of the child in the interaction. For the interaction, an autonomous system was equipped with a designed dialogic structure. It consisted of a repertoire of (pre-recorded) interaction protocols that a robot can apply in an interaction. The results from the evaluation suggest that the implementation was successful as most of the studied children engaged in the interaction with the robot. The results also reveal some gender differences. The long-term aim of the study is to develop an autonomous system that can be applied in an interaction with young children.

I. INTRODUCTION

Although the effects of social robots have been shown in various areas of learning [1], [2], there is as yet no systematic design for how robots can engage in a dialogue with young children for a longer period of time. Indeed, Belpaeme and colleagues [3] identify the selection of a correct response in open and unconstrained environments as a technological challenge. From previous research [4], [2], we know that contingency is a key aspect eliciting children's attention, interest and reception. Contingency refers to a social responsiveness of the interaction partner on a local level [5], i.e. an immediate reaction following an action of the partner. With respect to the dialogue, on a global level [5], i.e. with respect to the communicative activity and its goals, Fridin proposed 'educational games' [6] as a tool to achieve some educational aims through social interaction. However, the proposed games lack interactional systematicity in the sense that they do not systematically elicit children's verbal knowledge but rather initiated a number of different exchanges: some were linked to the activity of story telling, some to the activity of guessing and answering. In this paper, we propose a concrete dialogue with concrete communicative purposes to investigate whether children at the age of 4 and 5 engage in an interaction with a robot and experience it as rewarding and fun by themselves.

In extension of the results presented recently by Breazeal and colleagues [2], our goal was to engage children in longer communicative exchanges. The design of the dialogue is motivated by language games as suggested by Steels [7]. These consist of structured interaction protocols between the dialogue partners [7]. In contrast to previous studies that tested learning performance in children, the focus of our provided

dialogue was to elicit interaction behavior in children. More specifically, following the idea of 'pragmatic frames' [8], we reasoned that children need not only to participate in an interaction but that the interaction needs to be structured by (i) a clear role that the child can fulfill and (ii) some joint goals defined by the task. By analyzing children's engagement and emotional involvement (see section II C), we assessed whether children found this kind of interaction enjoyable.

Against this background, in this paper, it is our goal to present the development of an autonomous system that can be employed in such a structured interactions. Two key aspects in human robot interaction are crucial for such an approach: *modularity* and *autonomy*.

Modularity is emphasized in order to allow accessibility for future work and the integration of lessons learned during the iterative development process. It also provides an approach to address multimodality.

The overall behavior emerges from the interaction of the different modules and the underlying framework is responsible for integration. Furthermore, the clear structuring of explicit high-level functions in modules facilitates the export towards access layers intended for the use by people with less experience in software engineering, especially relevant for multidisciplinary fields like child robot interaction (cHRI).

The autonomy of the robot in interaction with humans is an important topic for HRI [9]. While many layouts of human robot interaction systems today rely on a *Wizard-of-Oz* (WOZ) set-up, many researchers [10], [11] advocate a turn towards autonomous systems, even during preliminary evaluation.

The article, first, presents the autonomous system for stable cHRI. Second, this system is applied in a study showing that children between 4 and 5 years engage in interaction and that joint attention can be established.

II. METHODS

A. System Overview

While many Child-Robot Interactions are built around "Wizard-of-Oz" (WoZ) scenarios, the long term goal of the cHRI field has to target more stable and reproducible settings. For such an approach, autonomy becomes an important issue. As one goal of the development of robots is to offer services to humans and to aid humans, it is mandatory that in the long-term robots do not require a lot of human intervention. Instead, robots should become autonomous and organize their actions

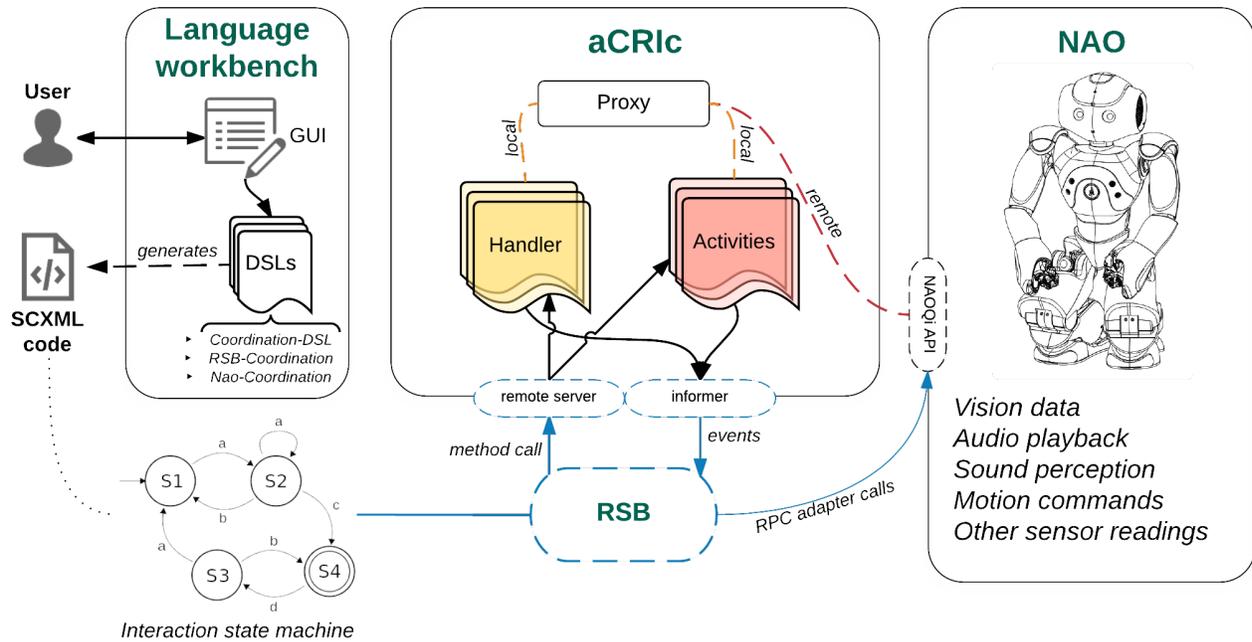


Fig. 1. Architecture of the cHRI system. In the middle, the main part of the system is shown which is connected with the robot on the right through the RSB protocol. The left shows the high level interface which allows to setup the structure of the control system using a DSL.

on their own. This is particular true in interactive scenarios which require the interaction partner to be truly embedded. While there are reasons to involve a distant teleoperator during interaction, it becomes more and more difficult to create this embeddedness and fluency which ultimately may cause the interaction to suffer: for example the timing would be affected which has shown to disrupt the interaction. It is therefore one of our goals to come up with an autonomous architecture which drives the robots interactive behavior in an adequate way.

As another requirement, the architecture has to deal with different modalities. On the one hand, multiple sensory inputs continuously provide information about the current state of the situation which has to be integrated in the unfolding interaction. On the other hand, the robot has different modalities to express himself which have to be coordinated, for example speech, motor behavior for locomotion, grasping or pointing movements as well as eye gaze.

In principle, the presented approach is following the idea of behavior-based robotics [12] as it is constituted of different modules working in parallel which are responsible for the different tasks. These different modules interact to organize the overall emergent behavior. As stated above, autonomy is a major goal for this system which poses further requirements in the development of the architecture:

- **Modularity:** A key concept of the approach is modularity as it allows a high degree of flexibility. The singular components comprising the behavioral repertoire of the robot

can be easily replaced or edited. This, on the one hand, allows to refactor processes and offers a good framework for further development. On the other hand, the whole system can be applied in different configurations and components can be reused. The goal is to realize this with a high level of description which can be used even by non-experts and which allows such non-experts to setup their own experiments. Therefore, we chose a highly modular approach and in addition introduced a high level domain specific language to describe an experiment and which modules should be included in what way.

- **Robustness:** The presented autonomous system is intended to work for ten minute sessions during which problems in locomotion, conversation and perception may arise. As the system is autonomous there should be no operator intervention. This requires to consider a large number of possible interactions in between different modules and to provide general resolution strategies. Importantly, such problem resolution schemes can not be tailored towards singular faulty interaction between specific modules. Therefore, this requires to implement generally applicable recovery and fallback strategies. Furthermore, exhaustive testing becomes necessary.
- **Plasticity:** Modularity can provide a reasonable easy way to design scenarios. Ideally, the designing process should work like using a construction kit. This kit then contains predefined behavior building blocks which the scientists can arrange to their liking.

It is important to find the right balance between these requirements and a robust, but still flexible and autonomous system. To guarantee stability albeit the modular design, two preliminary evaluation stages were used during the development and refinement of the architecture. These provided plenty of insights taken into consideration before the system could withstand a quantitative study.

The cognitive autonomous architecture consists of, first, control structures to execute and coordinate behaviors. Secondly, knowledge representations containing the structure of interaction, behaviors and current state information. The architecture consists of different layers of abstractions which provide interfaces to the robot’s hardware. The hardware platform used in this work is the robot Nao [13]. Nao is a humanoid robot with a height of 58 centimeters and 25 degrees of freedom which allow Nao to exhibit a wide range of human-like motor actions. It is equipped with a variety of sensors for proprioception of, e.g., joints and for perception of its surroundings, e.g. camera, touch sensors, gyroscopes and microphones.

As an operating system the linux-based OpenNao runs on Nao which is interfaced through the Robotics Service Bus [14]. The aCRic framework is used to control Nao’s behavior and will be briefly introduced in the following sections.

1) *autonomous Child Robot Interaction control (aCRic)*: serves as the main component of the software model and is responsible for the coordination of the robots behaviors. Within the architecture, it realizes the cognitive, behavioral and sensorimotor layer. There are two main building blocks:

- *Handler* deal mainly with perceptions. They have access to the robots API (providing algorithms as well as sensor and actuator raw data). Individually, each module is responsible for the integration of related data retrieved from the robots sensors and preprocessing.
- *Activities* are the internal representation of specific, small cHRI scenarios and provide a fixed structure. They manage the coordination of the interaction. Activity modules can access handlers, but not vice versa. They are intended to keep track of and adjust to the current situation.

The modules are registered in a central proxy as clients which allows internal as well as external communication. The modular layout of the system has been used to interface all required functionality to the interaction modeling layer.

2) *Interaction Level – Scenarios*: One goal of our approach is to provide a framework that can be flexibly employed –even by non-experts– for novel scenarios and allows for reusability. Therefore, on a higher level a *domain specific language (DSL)* has been used to describe, first, which modules are used. Secondly, how the information flow is organized inside the system. The DSL translates the structured user input into an executable data type which contains a formalized description of the whole interaction by the means of the central components methods. Furthermore it integrates the components feedback depending on the scenarios layout (e.g. wait for vocal response). The systems architecture provides access to complex behaviors necessary for the concrete cHRI scenario.

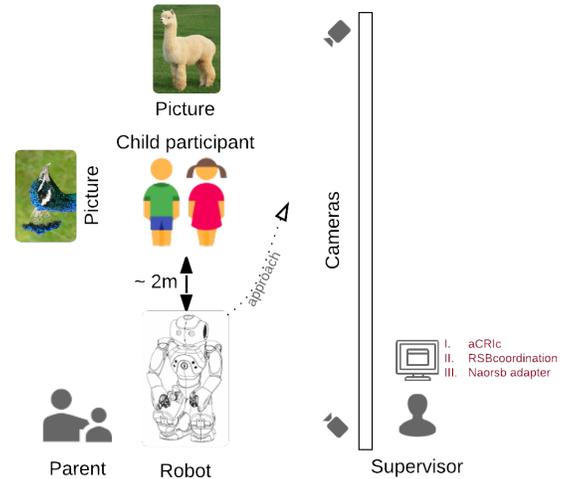


Fig. 2. Experimental setup. The child was seated on a red pillow at the indicated position and the robot approached it on the indicated path. The supervisor initiated the interaction and from there on had no control over the interaction.

B. Evaluation study

The robot framework has been applied in a study that focuses on the interaction between a robot and a child in a small guessing game. The game ended with a joint attention task. The experiment was done at the Cognitive Interaction Technology (CITEC) institute in Bielefeld, Germany. The room was divided into two visually separated parts (see fig. 2). On the one side, the interaction took place. The child was seated together with one of his or her parents in an area reachable by the robot. Two pictures were on the wall and door behind the child. On the other side, the supervisor was located behind an occluder. Two HD cameras were used for recording.

The participants were twelve preschool children all considered as being typically developed. Participant’s mean age was 58 months ($SD = 4.99$; range = 51–65). They were equally divided among their age in years and gender.

At the beginning of the study, the robot was briefly introduced as “Nao” and the children were told that it is able to move and speak. The accompanying person was asked to sit down at a table in close proximity to the child during the interaction. They were given the written project description, information on the data use and a declaration of consent that would allow the use of the video and audio material.

The interaction between robot and child lasted around ten minutes and was subdivided into several stages, following a classical scheme of HRI. At the outset, Nao was located in the social space of the child in a kneeling position exhibiting already slight secondary behaviors, as if breathing. Such secondary behaviors appear to be useful in order to increase the credibility of a robot as an autonomous agent [15]. After

spotting the participant in front of it, the robot waved, said that it can see them and started approaching. The robot was introducing himself verbally and asked the children for their name. At the end of the introductory phase, Nao proposed a game. To confirm their participation, the child was inquired to touch the robot's head. As the main phase of the interaction, a small language game was played: the "animal sounds" scenario. The scenario was designed as an inquiry-response game with two phases. First, in a learning stage the children were familiarized with the sounds and the style of Nao's interaction. The robot initially presented the child with all animal sounds along with the corresponding species name. In this phase, the participant was not yet actively involved in the conversation. Most of the sounds represent animals commonly known to children of this age like dogs, cats and cows. Second, during the recall-stage the robot was taking the role of a quiz master and the child became actively involved in a turn-taking activity solely based on the verbal exchange. The robot presented the child with an approximately 3 second sound clip of a sound a certain species produces and awaited an answer. The robot switched into a visually salient state of listening during which the child could provide an answer. In this state the robots LEDs in the ears were flashing, the robot shifted weight from one leg to another, pushed its hands against its hips and looked around the room which provided the child with an additional cue according to which they could align their moment of response. Any utterance of the child triggered the robot's perception. Neither the semantical nor syntactical content of the message was analyzed by the system. Following the recognition, the robot provided a positive feedback. The animal sound scenario focused on turn-taking between child and robot and asked the question if such an interaction can be continuously established in a simple, autonomous system and if the robot is capable to continuously engage the children.

As a second aspect, the experiment focused at the end on joint-attention as a fundamental mechanism for communication. Joint attention describes the shared focus of two individuals on an object in the same reference frame and can be realized by several verbal or non-verbal indications [16]. Thus, to test the ability of the autonomous system to establish this, the robot tried to guide the child's attention towards two images of animals inside the room. The robot asked the child to look at one image. First, the robot described the position verbally relative to the child and in reference to the room. Secondly, the robot turned on the spot and formed a pointing gesture with one arm in the direction of the image which was kept for several seconds. Afterwards, the robot turned back and provided the child with information about the animal. Again, as a research question we were interested if the children followed the interaction. Was the robot able to guide the joint-attention and did the children follow the cues given by the robot?

C. Evaluation Method

The goal of the evaluation process is to determine whether children were engaged in the interaction and whether they ex-

perience the encounter as rewarding and fun over the duration. This has been evaluated through analysis of the video material. It was a twofold manual analysis of the available data. The annotation of the video has shown to be reliable as a random sample of a quarter of the videos had been annotated twice, once by the main researcher and once by a student intern. The percentage of agreement for the scores was $p = .847$. The only training the intern received was the following description of the scores.HRI

As a first measurement, an *interaction level* (IL) was defined in order to access how much the children were engaged at different points of time. Based on a set of variables which individually rated key aspects of human-robot or human-human interaction, it provided a score which expresses the ability of the autonomous robotic system in combination with the interaction scenario alone to provide an enjoyable and motivating environment for the participants. The calculated score was based on the work of Fridin et al. [17]. Looking explicitly at the development of the *interaction level* over the course of the individual stages, allowed us to analyze whether the child was engaged during the interaction. The child-robot *interaction level* (IL) at stage s is given by:

$$IL_s = E_s^{I_s} * sign_s * A_s \quad (1)$$

The proposed formula evaluates the following variables:

- *Engagement, E*: Engagement has been described in literature as "the process by which two (or more) participants establish, maintain and end their perceived connection." [18]. The variable *engagement* measures the empirically observable attentiveness towards the robot. One promising cue when rating engagement during an interaction in a post-hoc video data analysis is the eye gaze of the participant [19]. Additional hints that indicate engagement are deliberate utterances, answering to a posed task and nonverbal cues like head nods which have been identified as backchannels [20]. Engagement was annotated as not given ($E=0$) if the child did not interact; it was considered as medium quality ($E=1$) if a single cue for engagement (gazing at the robot; usage of nonverbal (backchanneling) behavior; deliberate utterances; taking the turn when invited by answering questions) was observable or it was considered highest ($E=2$) if multiple cues had been found. This more fine-grained engagement score differs from Fridin et al. [17] who equated high engagement solely with the child looking at the robot.
- *Emotional Involvement – Valence and Affect*: Emotional involvement indicates whether the interaction does have positive or negative effects on the emotional state of the child. Following Fridin et al. [17], the expressions of the child were examined for observed valence. *Valence* is defined as the direction or the intrinsic attractiveness or aversiveness [21] of an emotion, ranging roughly between pleasant and unpleasant. Valence here was assumed as either good or bad and was reflected in *Sign* which can take on the values (+1) or (-1). *Affect, A*: represents a range of affective factors measuring whether there is an

TABLE I
THE MEAN AND STANDARD DEVIATION OF THE RESULTS FOR THE RESPECTIVE GROUPS OF THE PARTICIPANTS. RESULTS INDICATING SIGNIFICANT DIFFERENCES ARE HIGHLIGHTED.

Groups	N	Interaction level (IL)		Performance	
		Mean	SD	Mean	SD
Overall	12	5.16	2.206	1.35	0.416
Girls	6	3.83	1.885	1.22	0.474
Boys	6	6.48	1.721	1.48	0.339
Age four (σ 4.48)	6	5.60	1.984	1.14	0.439
Age five (σ 5.25)	6	4.71	2.510	1.56	0.291

emotional reaction during the interaction or not (ranging from 1 to 2 which is already reduced compared to the score applied by Fridin et al.). Modalities considered here were facial (e.g. smiling), bodily (e.g. clapping hands) and vocalized (e.g. laughing) expressions of affect.

- *Independence, I*: As a third aspect it is factored in if the engagement is happening independently from the surrounding persons. The autonomous system should ideally provide an environment in which the child is able to immerse in the interaction without being dependent on reassurance coming from the attending person or the examiner. It is known that preschool children are fixated and dependent on their parents [22] and that this behavior can be seen as normal in unknown environments. Independence was scored low ($I=1$) when the child depends on repeated help by the accompanying person or even stays in physical contact. It was accessed as high ($I=2$) when the child was acting on its own during the interaction and needed little or no affirmation.

In the original score, the “independence” and “engagement” terms were combined into one term. Here it was evaluated separately because there might have been children who constantly reassured themselves but at the same time were highly engaged in the interaction (as experienced in some cases in a pre-study). The division of engagement and independence as a consequence allowed to examine this in more detail.

As a second measurement, a performance rating was used which serves as an interaction metric that monitors the compliance with robot suggestions and instructions. As the tasks were considered easy for the children, the performance rating reflected if the children actually followed the interaction and not their abilities. Performance rating ranged from zero (not accomplishing the task), over one (requiring help) up to two (independent and deliberate accomplishment of the task).

III. RESULTS

The applied autonomous system is the result of continuous development and integration as proposed by [15]. Before the main study was conducted the system has been employed in two pilot studies which helped to discover errors and possible problems during the ongoing interaction. The iterative test process [10] lead to a stable and robust platform.

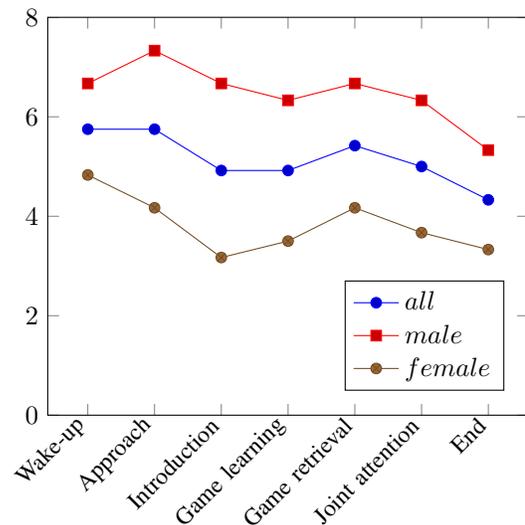


Fig. 3. Development of IL for the “girls” and “boys” group during the study in the course of the whole designed interaction.

A. Continuous Engagement of Children

The participating children were equally divided among their age in years and gender as visible in Table I. The average interaction level of the children was constantly high throughout the experiment (as visible in Fig. 3). With one exception¹, all participants persevered until the end of the interaction. They participated in the conversation with the autonomous robot platform and exhibited positive arousal.

The data was analyzed for a gender effect between boys and girls during the interaction. A Mann-Whitney test revealed that the distributions regarding IL in the two gender groups differed significantly ($U = -2.3$, $n1 = n2 = 6$, $P = 0.015 \leq 0.05$ two-tailed). Thus the “boys” group was associated with significantly higher IL. Additionally, they acquired an insignificant higher performance score ($U = -0.801$, $P = 0.485$) on average.

As a result of the preliminary study, effects of age towards the engagement were anticipated. Spearman’s rho did not indicate a significant correlation between IL and the participant’s age ($r_S = .210$, $p = .512$, $N = 12$). The data suggests that the younger children gained slightly higher IL while the performance score in comparison to the five years old group was slightly lower ($p = .20$).

B. Guiding Joint-Attention

The results of the performance score (Fig. 4) further stress that the children are involved in the task and follow the turn-taking during the “animal noises” game. Further, the performance score for the following phases highlights that the children stay in the interaction during the subsequent phases. In these phases the robot tries to guide the joint attention to other parts of the room, the pictures on the wall. The children

¹One child did not follow the interaction, but rather talked freely with Nao.

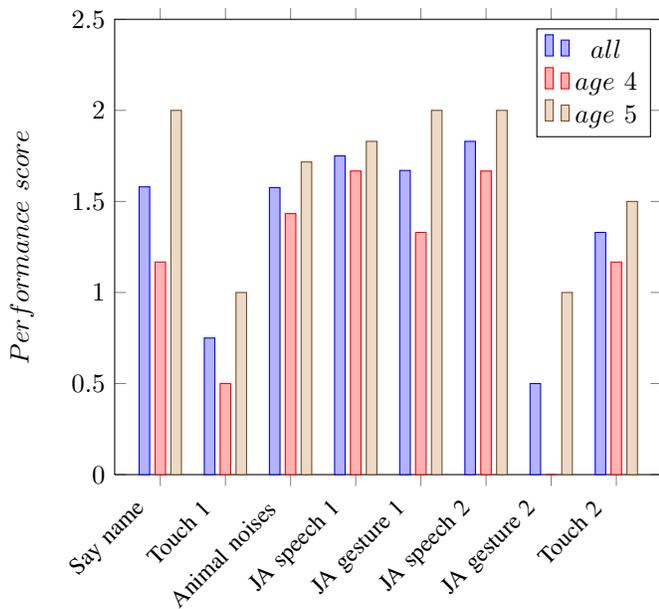


Fig. 4. Performance score results for the different tasks.

perform well for the first three joint attention tasks as shown in Fig. 4. All children followed the robot’s verbal inquiry to look at either picture whereas the successful establishment of joint attention by gesture was predominantly only realized in the first case (one explanation might be that the children had to turn away from the robot not recognizing the gesture of the robot anymore which only came later). Joint attention is established there by speech and gesture. Again, there is no significant difference between the age groups.

IV. DISCUSSION

The aCRIC system offers a framework for setting up experimental scenarios, even for non-experts as it provides a high level domain specific language to describe the modular structure of the framework. In the experiment, it is demonstrated that the system is able to engage children in an ongoing dialogue over a longer time span and further to direct attention through speech and (at least partially) through gesture.

Importantly, the system is not teleoperated but is autonomous as advocated by Belpaeme et al. [23]. Many other examples of cHRI are used in WoZ scenarios. One example is given by the approach in [24]. The system in that approach is divided into different layers. While the lower levels are constituted by sensorimotor and behavior levels similar to the approach presented here, there are also higher explicit supervision, control and cognition layers which are build on top of the other layers. The cognitive layer contains an explicit user model and is tasked with activity tracking as well as selection. The user model can be accessed on the higher supervision layer which also provides a WoZ user interface which enables the operator to exert almost holistic control over the robot’s communication and actions predefined by the behavior library.

In contrast, aCRIC is dealing right now with a simpler behavioral repertoire, and the dialogue system is not taking the semantics of the answers into account. However, we should emphasize that aCRIC is completely autonomous which allows for easy deployment and reusability of the system. The presented experiment shows that such a simple autonomous system can motivate children to interact with it at least by speech even though the system is not concerned with the semantic content of the answers. We should point to the limitations of our system that does not focus on long-term interaction and therefore the system’s memory is restricted and no learning is possible. These are certainly aspects that will play a key role in further developments.

Interestingly, our results indicate a significant difference in the interaction niveau depending on the gender of the participants but not regarding the performance score. Lots of literature has pointed out the formation of gender identity during early childhood [25]. Thus, the difference in children’s behavior depending on their gender is not surprising. Until the age of three, children start to play with gender-specific toys, mostly engage with playmates of their own sex and often have a stable sense of their gender identity (which often aligns with their assigned biological sex). They learn role behavior according to socialized aspects of their gender. To further investigate this question, the children were asked to answer some very general questions concerning the robots perceived animacy and gender after the experiment. In order to determine the perceived gender, the children were asked whether the Nao engages in some kind of activities stereotypically ascribed to boys or girls. The questionnaire was intended to incorporate their subjective perception. There was no significant result found, but typically male associated activities were favored slightly by both, male and female participants with respect to the robot.

cHRI [26] as a field is strongly connected to the field of human-robot interaction, but differs in the sense that in children a more playful interaction style can be observed. One important advantage in cHRI is that adults are more strongly culturally biased on how robots should behave or perform compared to children who have usually lower expectations. This provides a particular opportunity as children with a more playful approach are more likely to compensate for deficiencies in the cognitive abilities of the robot. Scientists can focus on the “*inherently social*” [10] nature of children’s interaction with robots and up to a certain degree neglect the demand for high-end technology. Therefore, cHRI does not only pose an entry point for HRI but also provide means and models to scientifically investigate the emergence of fundamental social behaviors during (early) development. In the present study, the autonomous system was able to interact successfully with the children and engage them continuously without interpreting the answers of the children. However, the data analysis focussed on social interaction. In the future, the dialogical framework should be extended and include more modules allowing to test more differentiated behaviors in children.

ACKNOWLEDGMENT

This research/work was supported by the Cluster of Excellence Cognitive Interaction Technology ‘CITEC’ (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

REFERENCES

- [1] F. Tanaka, *Cybernetics: Fusion of human, machine and information systems*. Tokyo: Springer Japan, 2014, ch. Robotics for Supporting Childhood Education, pp. 185–195.
- [2] C. Breazeal, P. L. Harris, D. DeSteno, J. M. Kory Westlund, L. Dickens, and S. Jeong, “Young children treat robots as informants,” *Topics in Cognitive Science*, pp. 1–11, 2016.
- [3] T. Belpaeme, P. Baxter, J. Greeff, J. Kennedy, R. Read, R. Looije, M. Neerinx, I. Baroni, and M. C. Zelati, *Social Robotics: 5th International Conference, ICSR 2013, Bristol, UK, October 27-29, 2013, Proceedings*. Cham: Springer International Publishing, 2013, ch. Child-Robot Interaction: Perspectives and Challenges, pp. 452–459.
- [4] F. Tanaka, A. Cicourel, and J. R. Movellan, “Socialization between toddlers and robots at an early childhood education center,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 46, pp. 17954–17958, 2007.
- [5] U. Quasthoff, *Narrative development: Six Approaches*. Mahwah, NJ: Erlbaum, 1997, ch. An interactive approach to narrative development, pp. 51–83.
- [6] M. Fridin, “Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education,” *Computers and Education*, vol. 70, pp. 53–64, 2014.
- [7] L. Steels, “Language games for autonomous robots,” *IEEE Intelligent Systems*, vol. 16, no. 5, pp. 16–22, 2001.
- [8] K. J. Rohlfing, B. Wrede, A.-L. Vollmer, and P.-Y. Oudeyer, “An alternative to mapping a word onto a concept in language acquisition: Pragmatic frames,” *Frontiers in Cognitive Science*, p. 470, 2016.
- [9] M. A. Goodrich and A. C. Schultz, “Human-robot interaction: a survey,” *Foundations and trends in human-computer interaction*, vol. 1, no. 3, pp. 203–275, 2007.
- [10] T. Salter, I. Werry, and F. Michaud, “Going into the wild in child–robot interaction studies: issues in social robotic development,” *Intelligent Service Robotics*, vol. 1, no. 2, pp. 93–108, 2008.
- [11] T. Belpaeme, P. Baxter, J. De Greeff, J. Kennedy, R. Read, R. Looije, M. Neerinx, I. Baroni, and M. C. Zelati, “Child-robot interaction: Perspectives and challenges,” in *Social Robotics*. Springer, 2013, pp. 452–459.
- [12] M. J. Mataric, “Behavior-based robotics,” in *MIT Encyclopedia of Cognitive Sciences*, R. A. Wilson and F. C. Keil, Eds. MIT Press, Cambridge, MA, 1999, pp. 74–77.
- [13] D. Gouaillier, V. Hugel, P. Blazevic, C. Kilner, J. Monceaux, P. Lafourcade, B. Marnier, J. Serre, and B. Maisonnier, “The nao humanoid: a combination of performance and affordability,” *CoRR abs/0807.3223*, 2008.
- [14] J. Wienke and S. Wrede, “A middleware for collaborative research in experimental robotics,” in *System Integration (SII), 2011 IEEE/SICE International Symposium on*. IEEE, 2011, pp. 1183–1190.
- [15] R. Ros, M. Nalin, R. Wood, P. Baxter, R. Looije, Y. Demiris, T. Belpaeme, A. Giusti, and C. Pozzi, “Child-robot interaction in the wild: advice to the aspiring experimenter,” in *Proceedings of the 13th international conference on multimodal interfaces*. ACM, 2011, pp. 335–342.
- [16] M. Tomasello and M. J. Farrar, “Joint attention and early language,” *Child development*, pp. 1454–1463, 1986.
- [17] M. Fridin, “Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education,” *Computers & education*, vol. 70, pp. 53–64, 2014.
- [18] C. L. Sidner, C. D. Kidd, C. Lee, and N. Lesh, “Where to look: a study of human-robot engagement,” in *Proceedings of the 9th international conference on Intelligent user interfaces*. ACM, 2004, pp. 78–84.
- [19] G. Castellano, A. Pereira, I. Leite, A. Paiva, and P. W. McOwan, “Detecting user engagement with a robot companion using task and social interaction-based features,” in *Proceedings of the 2009 international conference on Multimodal interfaces*. ACM, 2009, pp. 119–126.
- [20] V. H. Yngve, “On getting a word in edgewise,” in *Chicago Linguistics Society, 6th Meeting*, 1970, pp. 567–578.
- [21] N. Frijda, *The Emotions*, ser. Studies in Emotion and Social Interaction. Cambridge University Press, 1986.
- [22] S. P. Shelov, *Caring for Your Baby & Young Child: Birth to Age 5*. Oxford University Press, 1997.
- [23] T. Belpaeme, P. E. Baxter, R. Read, R. Wood, H. Cuayáhuil, B. Kiefer, S. Racioppa, I. Kruijff-Korbayová, G. Athanasopoulos, V. Enescu *et al.*, “Multimodal child-robot interaction: Building social bonds,” *Journal of Human-Robot Interaction*, vol. 1, no. 2, pp. 33–53, 2012.
- [24] A. Coninx, P. Baxter, E. Oleari, S. Bellini, B. Bierman, O. B. Henkemans, L. Canamero, P. Cosi, V. Enescu, R. R. Espinoza, A. Hiolle, R. Humbert, B. Kiefer, I. Kruijff-Korbayova, R. Looije, M. Mosconi, M. Neerinx, G. Paci, G. Patsis, C. Pozzi, F. Sacchitelli, H. Sahli, A. Sanna, G. Sommariva, F. Tesser, Y. Demiris, and T. Belpaeme, “Towards long-term social child-robot interaction: Using multi-activity switching to engage young users,” *J. of Human-Robot Interaction*, 2015.
- [25] C. L. Martin and D. Ruble, “Children’s search for gender cues cognitive perspectives on gender development,” *Current directions in psychological science*, vol. 13, no. 2, pp. 67–70, 2004.
- [26] K. Dautenhahn, “Robots as social actors: Aurora and the case of autism,” in *Proc. CT99, The Third International Cognitive Technology Conference, August, San Francisco*, vol. 359, 1999, p. 374.