Multimodal Child-Robot Interaction: Building Social Bonds

Tony Belpaeme, Paul Baxter, Robin Read, Rachel Wood
Plymouth University, United Kingdom
Heriberto Cuayáhuitl, Bernd Kiefer, Stefania Racioppa, Ivana Kruijff-Korbayová
Deutsches Forschungszentrum für Künstliche Intelligenz, Germany
Georgios Athanasopoulos, Valentin Enescu
Vrije Universiteit Brussel, Belgium
Rosemarijn Looije, Mark Neerincx
Organization for Applied Scientific Research, The Netherlands
Yiannis Demiris, Raquel Ros-Espinoza
Imperial College London, United Kingdom
Aryel Beck, Lola Cañamero, Antione Hiolle, Matthew Lewis
University of Hertfordshire, United Kingdom
Ilaria Baroni, Marco Nalin
Fondazione Centro San Raffaele del Monte Tabor, Italy
Piero Cosi, Giulio Paci, Fabio Tesser, Giacomo Sommavilla
National Research Council - ISTC, Italy
and
Remi Humbert
Aldebaran, France

For robots to interact effectively with human users they must be capable of coordinated, timely behavior in response to social context. The Adaptive Strategies for Sustainable Long-Term Social Interaction (ALIZ-E) project focuses on the design of long-term, adaptive social interaction between robots and child users in real-world settings. In this paper, we report on the iterative approach taken to scientific and technical developments toward this goal: advancing individual technical competencies and integrating them to form an autonomous robotic system for evaluation "in the wild." The first evaluation iterations have shown the potential of this methodology in terms of adaptation of the robot to the interactant and the resulting influences on engagement. This sets the foundation for an ongoing research program that seeks to develop technologies for social robot companions.

Keywords: Human-robot interaction, child-robot interaction, interaction design, adaptive social robotics, natural language, memory, engagement

Journal of Human-Robot Interaction, Vol.1, No.2, 2012, Pages 33–53, DOI 10.5898/JHRI.1.2.Belpaeme

1. Introduction

The ultimate goal of Human-Robot Interaction (HRI) is to have robots working in real-time with real people. Such aim presents enormous challenges, not the least of which is the necessity to have multiple behavior systems operating in synchrony. Humans interact socially, that is, behaving in ways which are contingent on the actions of those around us. Robots, if able to interact effectively with human users, also need to have the capacity of generating coordinated and timely behaviors predicated on their social surroundings. This presents an enormous challenge; acting socially means not just the performance of appropriate action sequences but also having a substantial degree of flexibility in the organization of behavior to produce timely responses that 'make sense' to a user. The technical difficulties associated with coordinating the functioning of multiple action systems to provide coherent, flexible, and timely behavior are enough to ensure that projects attempting HRI 'in the wild' (i.e. real robots interacting with real people in real time) are rare. Rarer yet are those HRI projects aiming to go beyond the level of moment-by-moment interchange to achieve long-term social coherence. While the interaction between children and robots has been studied (e.g. Draper & Clayton, 1992; Kanda, Hirano, Eaton, & Ishiguro, 2004; Tanaka, Cicourel, & Movellan, 2007), the Adaptive Strategies for Sustainable Long-Term Social Interaction (ALIZ-E) project[1] differs in its ambition as a multi-partner initiative focused on long-term adaptive social interaction between robots and child users built up through multiple sessions extended over a period of days.

The problems of configuring multiple and complex sub-systems to work effectively together are substantial and can get in the way of developing novel technologies for behavior production. In many cases it makes sense to take a 'divide and conquer' approach developing modular components that are only integrated into a single system at the end of the development cycle. Modular design is convenient and sensible, allowing expertise to be focused on making concurrent progress with several components. However, such an implementation approach does not provide any insight into the integration problem that must still be tackled if outcomes are intended for embodied robots in real-world settings. Conversely, a top-down approach to designing a complex system of interacting components, such as a cognitive architecture for HRI, leads to its own problems. In such systems the functioning of individual parts is typically constrained to fit with a centralized 'executive' control strategy and so flexibility of coordination may have to be sacrificed for robustness of performance. The ALIZ-E project seeks to steer a course between these extremes by marrying a focus on the implementation of novel technologies with a rolling program of *in situ* testing with child users from the earliest stages of development. Furthermore, the presence of a clinical application domain ensures that focus remains on a system capable of delivering definitive benefits to the child interactants. The project thus benefits from new approaches to solving the problems of cognitive architecture implementation, from individual sub-components through methods for the coordination of behavioral output. At the same time, the necessity of having testable system prototypes available from the earliest stages of development has led to an early focus on the issues associated with integration.

The objectives of ALIZ-E are long-term, adaptive social interaction with child users in a hospital setting, our test population being 8-11 year-olds diagnosed with metabolic disorders (diabetes and obesity). The aim of the project is to develop the science supporting robot companions able to interact with these children while they are in-patients at the hospital, acting as 'friends' and 'mentors' to improve the children's experience of a hospital stay, support their well-being and aid in their learning about the management of their health condition. Motivation for this approach comes from two diverse sources. First, previous research has demonstrated the efficacy of animal companions in supporting positive health outcomes for hospital in-patients (Fine, 2010). However such schemes are expensive and issues such as hygiene are not easy to overcome. Robots can potentially be an

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

alternative source of the welfare benefits provided by animal therapy and even beyond this, offer social, emotional and educational resources for hospital staff to use in patient care. The second motivation is a general willingness of children to engage with robots and treat them as social agents (Breazeal, 2003; Salter, Werry, & Michaud, 2008). ALIZ-E aims to use this imaginative capacity as a means to bootstrap social engagement into providing a platform for richer, temporally extended Child-Robot Interaction (CRI).

In order to evaluate the HRI aspects involved, a program of *in situ* hospital-based testing has been designed around a set of activities and games fitted to the robot's role in supporting users learning about and managing their medical condition. Games and activities include a quiz, a math game, an imitation game, a collaborative menu selection task (the 'SandTray'), and a dance game. These form the basis of the robot's interaction with each user (e.g. figure 1). The quiz serves as a non-physical interaction where the child and robot use language as the primary interaction channel. The math game is similarly verbal-interaction oriented, although the focus of this activity is in adapting the game to suit the child's performance in order to maximize learning outcomes. The imitation and dance games are more physically oriented and look into how children can be engaged in game play led by the robot. Finally, the SandTray activity serves as a way for the children to explore dietary requirements relevant for their medical condition, jointly with the robot. In each of these cases, the emphasis is not just on task completion, but on the social interactions that may arise while the task is being executed: these activities thus provide contexts for potential social interactions between the children and a robot. Additionally, they provide relatively constrained application contexts that facilitate development, testing and deployment of the various aspects of this technical system (section 2).

Figure 1. Young users with a robot engaging in a quiz game (left), an imitation game (centre) and a dance game (right).

It is in this respect, that the development of technologies capable of sustaining real-world interactions beyond the scale of minutes to multiple sessions over several days, defines a central aim of ALIZ-E. Human social behavior is predicated upon interaction histories enacted on multiple-timescales. Our responses to others are heavily influenced by our previous experiences in similar situations and by the unfolding of the current social exchange. To date, HRI has not been able to fully take into account interaction history as a means to coordinate and prime behavior in an embodied cognitive architecture. As such, ALIZ-E is developing methods to provide such a temporal embedding by implementing processing substrates to provide an experience-biased coordination of system information and behaviors.

While the central efforts in ALIZ-E, and the wider field of HRI, are geared towards autonomous social long-term interaction, it must be acklowledged that there remains a considerable effort necessary to develop supporting technologies to a state where such an integrated solution is possible. The aim of this paper, therefore, is to describe the developments made thus far in ALIZ-E towards the various technological domains that must underlie an integrated system focused on social CRI. Furthermore, given the emphasis on iterative development of testable system prototypes from an early

stage in the project, an overview of preliminary results involving integrated systems are presented that demonstrate the utility of the approaches taken in the challenging application domain.

The remainder of this paper is structured as follows. First, a review is provided of the novel technologies and approaches developed for the ALIZ-E project and the manner in which they can be evaluated (sections 2 and 3). Second, initial results are presented from experiments with children both in a hospital setting[2] and elsewhere (sections 4 and 5).

2. Constituent technologies

The technical focus in ALIZ-E is on iteratively developing technologies fundamental to various aspects of HRI, and the implementation of integrated systems to achieve competence in a range of HRI interaction scenarios. In this section, four such fundamental technologies are explored; each may be regarded as a *cognitive modality*, or *cross-modality system*. For each of these, the contributions made to the respective fields are identified, and the means of validation and evaluation in an HRI context described: (1) natural language competencies; (2) memory structures; (3) user modeling; and (4) bodily expression and emotion. While in this section these are described individually, for an integrated system each component must coordinate information transfer and interact with each other (description of this process may be found in section 3).

2.1 Natural language interaction

In this section we highlight the most important aspects of the technologies we are developing for the processing of natural language in supporting long-term child-robot interaction. The processing of natural language input and output in our system follows the classical pipeline model consisting of automatic speech recognition (ASR), natural language understanding (NLU), dialogue management (DM), natural language generation (NLG) and text-to-speech synthesis (TTS). The quiz game forms an ideal basis for development and evaluation, given its inherent verbally interactive nature and its provision of a naturalistic constraint on the interaction structure (rounds of question-answer sequences).

2.1.1 Spoken input interpretation (ASR and NLU) Acoustic and linguistic characteristics of children's speech differ significantly from those of adult speech. We therefore trained acoustic models for Italian child speech using speech material from the ChildIt corpus (Gerosa, Giuliani, & Brugnara, 2007) and additional recordings made during ALIZ-E field evaluations, which extended these corpus resources.

A first ASR experiment was conducted using recordings made during experiments at the San Raffaele Hospital in March-April 2012, from which we extracted only those speech segments where users pose questions to the robot in the quiz game. The Open-Source Large Vocabulary Continuous Speech Recognition (LVCSR) Engine Julius[3] was used, with a 4-gram language model built with text data from the questions and answer-options in the quiz database. Results can be seen in Table 1.

The result of speech recognition (71% correct words and 42% word errors) is encouraging. As the children read questions from cue cards, the questions are known to the robot. This allows the NLU component to identify 56 out of the 64 questions correctly (87.5%) through fuzzy matching of the recognized content words against the quiz database entries.

For robust recognition of dialogue acts other than quiz questions and answers options, we use partial parsing with keyword spotting as a fall-back. Our partial parsing approach is based on chart

[2]Currently experiments have been run at the San Raffaele Hospital (Milan, Italy) and the Wilhelmina Kinderziekenhuis (Wilhelmina Children's Hospital, Utrecht, The Netherlands).

[3]The Sphinx3 ASR engine was evaluated, but proved difficult to implement features critical to our application domain (such as live decoding).

Table 1: Results of preliminary ASR test. "Expt. ID" is the experiment identification number; #Snt is the number of sentences; #Wrd is the number of words; WCR (Word Correct Rate) is the percentage of words correctly identified by the speech recognition; "Ins" is the percentage of words entered by speech recognition but not present in the transcript reference; WER (Word Error Rate) is the sum of words' substitution, deletion, and insertion percentages.

Expt. ID	#Snt	#Wrd	WCR	Ins	WER
1	4	22	81.8	40.9	59.1
2	6	82	73.2	35.4	62.2
3	5	40	52.5	15.0	62.5
4	7	63	73.0	1.6	28.6
5	15	114	90.4	7.0	16.7
6	4	49	61.2	12.2	51.0
7	12	107	58.9	1.9	43.0
8	11	84	65.5	10.7	45.2
Total	64	561	**70.6**	12.5	**41.9**

parsing, together with heuristic data to select best partial analyses, based on statistical methods to assess the quality of edges contained in the parsing chart. We used OpenCCG[4] as a basis, an open source natural language processing library, which provides parsing and realization services based on Combinatory Categorial Grammar (CCG) (Steedman, 2000b) with the multimodal extensions described in (Baldridge, 2002; Baldridge & Kruijff, 2003).

2.1.2 Dialogue Management (DM) The task of DM is to keep track of the state of interaction, to integrate interpretations of the user's spoken input (or nonverbal actions) with respect to this state, and to select the next communicative action. Recent years have seen a boom in research in spoken DM using probabilistic methods (Roy, Pineau, & Thrun, 2000; Williams & Young, 2007; Thomson, 2009; Young et al., 2010) and optimization of dialogue policies using reinforcement learning (Frampton & Lemon, 2009). We are contributing to this line of research by developing our framework for modeling human-robot dialogue under uncertainty as described below.

A human-robot dialogue consists of a finite sequence of verbal units. Assuming that the robot receives a numerical reward r_t for executing action a_t when the conversational environment makes a transition from belief state b_t to state b_{t+1}, a dialogue can be expressed as a sequence where $D = \{b_1, a_1, r_2, b_2, a_2, r_3, ..., b_{T-1}, a_{T-1}, r_T, s_T\}$, and where T is the final time step. A Reinforcement Learning (RL) agent uses such sequences to optimize the robot's dialogue behavior. We apply the learning approach developed in (Cuayáhuitl, 2011), which extends the state representation of Markov Decision Processes with relational representations and belief states.

We have developed methods for online learning of policies for flexible interaction (Cuayáhuitl & Dethlefs, 2011). Our approach extends the flexibility of learning dialogue systems in three ways. First, we introduce dynamic tree-based state representations that can grow (in the order of 10^4 states) during the dialogue according to the state variables used in the interaction. Second, rather than imposing strict hierarchical dialogue control, we allow users to navigate across the available sub-dialogues. Third, we represent the dialogue policies using function approximation in order to generalize the decision-making even for unseen situations. To this end, we use hierarchical RL algorithms with dynamic states, global state transitions and linear function approximation. This combination is our main mechanism to support dynamic adaptation.

[4] http://openccg.sourceforge.net/

Long-term human-robot conversational interaction requires the robot to keep track of the exhibited behavior and history of perceived observations for each set of interactions. We use our dialogue policy learning framework described above to infer the agent's behavior (or dialogue policy) from interactions within the environment. For the agent to adapt to newly gained knowledge of in real time, we investigate the use of probabilistic modeling and reasoning to incrementally track the beliefs of a given user, an build on memory systems and adaptive user modeling (see sections 2.3 and 2.2, respectively).

2.1.3 Spoken output production (NLG and TTS) In the next step, the communicative action selected by the DM is verbalized and then realized by speech, possibly also accompanied by nonverbal behavior. The DM determines the type of dialogue act and the values of a range of information state variables important for verbalization selection. The verbalizations are designed to foster a sense of familiarity to support long-term interaction based on common ground between interactants, including the use of names, references to previous encounters, etc. A range of interaction contexts are covered, such as greetings and introductions; activity-management moves (e.g., a request to play, a request to switch roles, a request for a user turn, etc.); asking questions (e.g., engagement in a game or a quiz question); and, providing instructions, information, and comments on the user's performance, in addition to various types of feedback and clarification requests.

We have developed an utterance planner that allows us to describe the construction of utterance verbalizations in a modular and flexible manner. The utterance planner engine is a general graph rewriting system. The rewriting rules have antecedents where basic tests for the presence or absence of substructures can be combined with Boolean operators to complex match conditions. The consequent, or rather, rewriting part contains addition and deletion instructions. Variables in the antecedents and consequent parts of rules make it possible to move existing information, i.e., subgraphs, into new locations. We currently use a canned-text approach to generation, where the utterance planning rules define rewriting of sub-graphs into natural language strings and combining of these strings with one another, to produce the output utterances. Alternatively, the output could be a logical form that would serve as input into a grammar-based lexical realization component using OpenCCG.

It is well known that (dialogue) system output can be tedious when it is repetitive. Since this could negatively influence engagement in studies with real users, we have invested considerable effort to implement a large range of verbalization variation. Selection among variants is either random or controlled by certain criteria, including contextual parameters (e.g. *is this the first interaction?*) and context characteristics (e.g. *has this quiz question already been asked?*) To assess the implemented verbalization variability, the utterance planner was allowed to generate verbalizations for all available dialogue acts with a variety of contextual parameter settings. A sequence of 40,000 iterations produced 59,296 unique utterances, with an evident convergence (figure 2). Users are not being exposed to repetitive system output since often-occurring dialogue acts have tens to hundreds of different verbalizations.

For speech synthesis we either use the commercial Acapella TTS system available on the Nao robot (Gouaillier et al., 2008), or the open source MARY TTS platform[5] (Schröder & Trouvain, 2003), with the latter allowing more control over prosody and voice quality. An Italian voice was constructed using the Mary TTS voice creation toolkit (Schröder, Charfuelan, Pammi, & Steiner, 2011). We are using the Statistical Parametric Speech synthesis approach (Masuko, Tokuda, Kobayashi, & Imai, 1996; Zen, Tokuda, & Black, 2009), given that Hidden Markov Model (HMM)-based voices are reaching high-quality synthetic speech.

A problem which affects modern TTS is the lack of naturalness due to incorrect or ambigu-

[5] mary.dfki.de

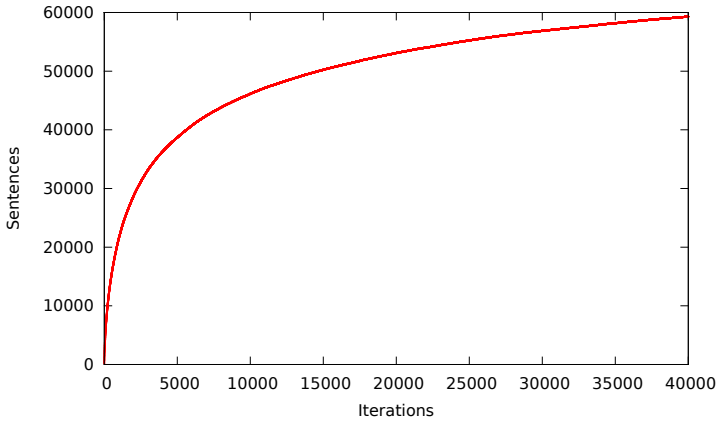

Figure 2. Unique sentence count per iteration of the full rule set in Natural Language Generation. Variation in the robot's responses is deemed necessary to maintain user engagement.

ous prosody of the generated sentence. Using the support MARY TTS provides for controlling the prosody of HMM-based voices with symbolic markup (Pammi, 2011), we implemented two prosody modifications. First, *prosodic prominence modification (stress)*: the NLG component labels focus words (Steedman, 2000a). The TTS component then modifies the prosodic realization by decreasing the speech rate and raising the pitch contour on the focus words. Second, *emotional prosody modification*: the dialogue manager decides when the system output should be rendered with (non-neutral) emotional coloring, either "sad" or "happy". The TTS component ensures the corresponding realization by increasing or decreasing the rate of speech and the pitch contour in the happy/sad case, respectively.

2.2 Memory structures for long-term social interaction

In attempting to achieve long-term social interaction, there is a clear necessity for memory to link prior experiences with ongoing and future episodes. Temporally extended social behavior also entails the coordination of multiple cognitive functions (Ros et al., 2011). In addition to performing the function of explicit semantic information storage (the primary function of existing robotic memory systems), memory may thus also be viewed as a means to achieve this coordination (Wood, Baxter, & Belpaeme, 2012). In this perspective, memory enables the storage of the relationship between multiple cognitive modalities, such that coordination between them may subsequently be achieved. This novel use of memory within an cognitive architecture focused on social interaction enables multimodal coordination to be a fundamental mechanism, integrated with the functioning of each modality, rather than a higher level capacity (Baxter, Wood, Morse, & Belpaeme, 2011).

Typically, robot control architectures (whether for HRI or other applications) treat memory as a passive storage device for symbolic encodings of environmental or agent states at a particular time (notionally a type of episodic memory). The information that is determined to be relevant is recalled on the basis of contextual cues for use in ongoing behavior. This approach has the benefit of providing explicit semantic information from prior experiences (e.g. *the ball I saw yesterday at 2pm was red*...), but requires information from multiple modalities to be encoded using some common, system-wide ontology, e.g. (Cassimatis, Trafton, Bugajska, & Schultz, 2004; Hawes & Wyatt, 2010). Additionally, this storage and recall of information does not typically influence the future operation of the different modalities themselves, but rather, only forms an input to current

processing. Recently, however, a novel framework for memory based on neuropsychological theory has been developed (Baxter, 2010; Wood et al., 2012). In this view supported by neurophysiological evidence (e.g. Fuster (1997) and Bar (2007)), memory is regarded as being fundamentally associative and via ontogenetic processes of construction provides a substrate for activation dynamics that shape subsequent behavior.

These principles combined with the spreading activation properties of Interactive Activation and Competition models (McClelland & Rumelhart, 1981; Burton, Bruce, & Johnston, 1990) can be applied to the design of robot control architectures enabling memory to provide 'soft' coordination between components. As such, there are a number of computational implementations consistent with these principles, as found by the work of (Morse, Greeff, Belpeame, & Cangelosi, 2010; Greeff, Baxter, Wood, & Belpaeme, 2012; Baxter, Greeff, Wood, & Belpaeme, 2012). By associating units of information in different modalities (e.g. the outputs of facial recognition processing and dialogue management) as they are acquired (and hence activated) through the robot's interaction with an environment, links are formed that may be reactivated in the future thus influencing ongoing processing (Baxter et al., 2011; Wood et al., 2012; Baxter, Cuayáhuitl, Wood, Kruijff-Korbayová, & Belpaeme, 2012). This can occur without requiring information to be translated into a common ontology, since the robot's memory is comprised only of encodings of associative relationships between modality-specific items.

Assessing the functionality and contribution of such a memory system to social human-robot interaction is problematic in that memory does not in itself have a direct interface with the environment: it necessarily interacts with sensory and effector modalities, which in turn provide the means by which outputs of memory are manifested. Indeed, this issue is present regardless of the methodology used to instantiate memory. The evaluation of such a system for cross-modal coordination therefore requires an interaction scenario that meets two requirements: first that there is an opportunity for the system to demonstrate behaviors that are not merely driven by a predefined script (thus allowing the influence of prior experience on ongoing behavior to be manifested), and second, that the conditions of the social interaction may be manipulated so that the influence of such a memory system on behavior may be observed. An activity that meets these requirements has been developed that emphasizes the facilitation of a social interaction between a robot and a human interactant by providing a context and a focus for the encounter, but without explicitly determining the interaction structure (Baxter, Wood, & Belpaeme, 2012).

Named 'SandTray', this scenario provides a collaborative activity in which a large touchscreen acts as an interaction medium for robot and child. This setup enables the pair to work collaboratively on activities such as an icon sorting task without imposing a rigid turn-based interaction format (figure 3). SandTray takes inspiration from the use of sandplay techniques in child psychotherapy (Lowenfield, 1939), where a sand filled tray and model characters are used to provide a nondirective context for facilitating communication and enabling the child to safely explore and express emotions (Hale, 2000).

In such a setting, robot and child can participate equally in performance of tasks by manipulating objects displayed on the screen where it is possible to explore a wider-range of interaction styles without the constraints of a turn-based game. The robot's behavior thus becomes central in building and maintaining engagement of the child, aided by the context that the task provides. SandTray supports manipulation of both task and interaction characteristics by providing a platform for exploring how soft memory-based coordination and priming can enable a richer repertoire of social behavior in the robot, hence a deeper level of user engagement. While full studies are still underway for this activity, pilot experiments have shown that child users respond positively to the collaborative setting provided by SandTray. It is also apparent that the lack of a fixed structure to the interaction provides more scope for flexibility in the robot's behavior which also has a positive effect on user

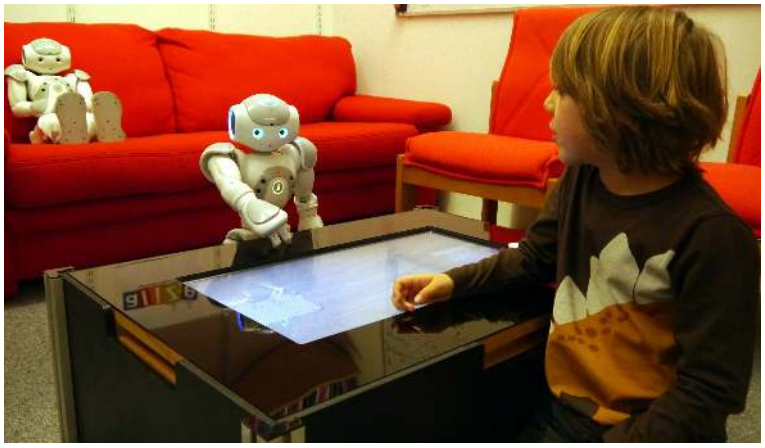

Figure 3. The 'SandTray' setup: social interaction between robot and child is facilitated by a touchscreen, which provides a collaborative activity where they can equally participate.

engagement.

2.3 User modeling through interaction

In order to support the general goal of *long-term* social interaction it is important that the robot adapts its behavior to each user by taking into account the specific semantic task and shared interaction information. In ALIZ-E this process is achieved through user modeling (and subsequent interactions with the memory system). The user model comprises three main components:

- General data: such as name, gender, age and whether the child has interacted with the robot before.
- Specific data: related to a particular activity (e.g. a record of performance in an activity such as dance or questions previously asked in the quiz).
- A decision making system: provides reasonsing about the goals of the activity and the behavior of the child. This module combines data from different sources and may send suggestions to other components of the integrated system.

The user model receives inputs from all components of the cognitive architecture (see section 3) and stores data pertaining to a particular user. Inferential processing of this data is used to generate inputs to other modules, thus shaping the robot's behavioral output, using the GOAL system (K. V. Hindriks, 2009). User data, for example, may guide the robot to maintain or change the current activity, to modify its speech content, or to change the emotional expressivity of its behavior (figure 4). The reasoning engine also decides on the social role of the robot, as to if the robot engages the child as an 'educator' or as a peer 'motivator'. The role depends upon the strategy chosen to achieve the goal. Prior work has demonstrated the applicability of GOAL to modeling user moods to inform behavior selection in an HRI setting (K. Hindriks, Neerincx, & Vink, 2012) - reasoning in the ALIZ-E user model is an extension of this application.

The role of the user model varies between activities; in the dance game and quiz scenarios, measures of the user's previous and ongoing performance (dynamic data) play an important part of shaping the ongoing interaction. In other contexts, static data pertaining to the user's interaction style and preferences may be more important. A number of experiments have been carried out to evaluate the most relevant user characteristics to model, on which to base user-specific adaptations.

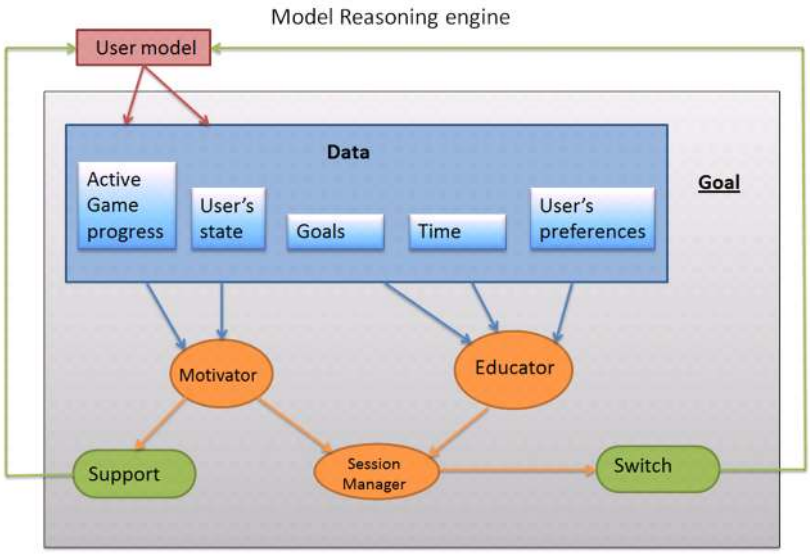

Figure 4. User Model based reasoning engine to determine the appropriate robot behavior strategy.

For largely practical reasons these experiments have been run independently of the hospital-based testing program using a population of non-clinical participants (see section 4.1).

2.4 Bodily expression of emotion

As an essential aspect of the ALIZ-E project, we investigate how to implement expressive gestures and behaviors to achieve sustained and meaningful interaction with child users. Part of this research looks at the generation of nonverbal expressive elements that are both grounded in the robot's particular embodiment, while being meaningful to the user.

In support of this goal, we conducted a perceptual study looking at how Italian children interpret key emotion poses displayed by the Nao robot (Beck et al., 2011). The aim was to investigate whether results obtained with adults could be replicated with children. We previously found that adults were able to interpret body poses displayed by the robot as conveying certain emotions, which is particularly relevant for the Nao robot used in ALIZ-E since it does not have facial articulation and as such is very limited in communicating emotion when not moving its body. We also noted that changing the robot's head position affects the expressiveness of the poses (Beck, Cañamero, & Bard, 2010; Beck, Stevens, Bard, & Cañamero, 2012). As with adults, it was found that moving the head upwards increased children's identification of emotions such as pride, happiness, and excitement, whereas moving the head downwards increased the correct identification for other displays (anger and sadness). Fear, however, was well identified regardless of head position.

These results have already been successfully integrated into an automated expressive system developed for ALIZ-E (Hiolle, Cañamero, Andry, Blanchard, & Gaussier, 2010) and used in various activities to give feedback to the user regarding the interaction.

The key elements of this expressive motion generation module are described below (also see figure 6).

- Motion Execution: this is the single interface used by other components.
- Emotional Key Poses: this is a dictionary of poses triggered either directly by the Motion

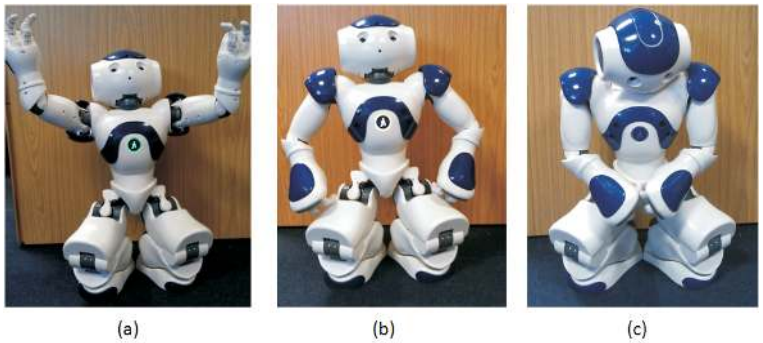

Figure 5. A Nao robot showing emotions through the use of body postures, these poses are consistently recognized by children and adults; (a) happiness (b) pride and (c) sadness.

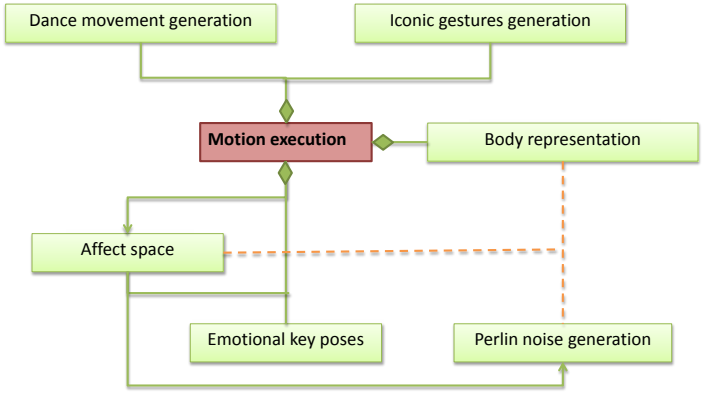

Figure 6. Structure of the Motion Generation system

Execution interface or through the Affect Space interface using valence and arousal as inputs. The list of joints which will be involved in the display is given by the Body Representation class.

- Perlin Noise Generation: this class generates trajectories for the different joints in order to animate the robot. Perlin noise adds a random rhythmic perturbation to the robot's pose and creates an illusion of the robot being alive. The list of joints to which Perlin noise is applied is given by the Body Representation class: Perlin noise cannot, for example, be applied to the legs and feet, as this has a negative impact on the robot's balance.
- Affect Space: this determines the emotional key pose to be displayed and the parameters of the Perlin Noise animation in order to express Valence and Arousal.
- Iconic Gestures Generation: this generates trajectories of predefined iconic gestures such as head nods, greetings and so on.

• Dance Movements Generation: this generates trajectories of predefined dance movements for the dance game.

• Body Representation: this class handles the joints of the robot and their use in the generation of movements and expressive displays. Each joint has several switches in order to control whether they will be part of a Perlin animation, a key pose, or a movement, permitting a fine control of expressivity and also enabling mixed emotional displays.

3. Evaluation platform and system integration

The ALIZ-E project makes use of the Nao humanoid robot as a common development and evaluation platform. Using such a shared platform facilitates the exchange of code and the transfer of results, which in turn allows quick prototyping and the adoption of agile project management. The Nao is a small humanoid robot, measuring 58cm in height, weighing 4.3kg and having 25 degrees of freedom. It has a range of sensors and actuators: 2 loudspeakers, 4 microphones, 2 cameras, a gyroscope, an accelerometer, and range sensors (2 IR and 2 sonar). The robot has an embedded computational core and connects externally via IEEE 802.11g WiFi or Ethernet. The Nao has a generally friendly and non-threatening appearance, which is therefore particularly well suited for child-robot interaction (Nalin, Bergamini, Giusti, Baroni, & Sanna, 2011).

The software integration relies on the Urbi framework, a robot-oriented middleware[6]. Urbiscript handles the orchestration of behaviors locally on the robot, but interfaces with software running on remote computers to handle computationally expensive tasks, such as natural language processing and computer vision that cannot be completed on-board in real time. Furthermore, it provides the substrate for event-handling that within the context of an integrated system readily allows environmental stimulus re-activity, which is of particular importance in the real-time context of HRI.

Above (section 2), we have described a selection of foundational technologies under development in ALIZ-E that provide competencies required of a robot designed to engage in long-term social interaction. In addition, the evaluation of these components individually has been discussed, and their performance reviewed. However, there are further components developed in ALIZ-E that have not been discussed here, but which form important prospective parts of an integrated system; e.g. voice activity detection (Dekens & Verhelst, 2011), emotion recognition from the user's voice (Wang, Verhelst, & Sahli, 2011), machine vision for emotion recognition from faces (Gonzalez, Sahli, Enescu, & Verhelst, 2011), gesture recognition (Oveneke, Enescu, & Sahli, 2012) and non-linguistic verbal expressions (Read & Belpaeme, 2012).

The integration of components (i.e. the manner and means of their interaction) is an important aspect of development for a system that is to be used in HRI evaluation studies. Using Urbi as the coordinating middleware, the developed architecture enables information to be exchanged between the different modules in an event-based manner (Kruijff-Korbayová et al., 2011), where the overall system behavior (including the coordination of the modules and sub-systems to be used) is achieved through the use of activity-specific controllers (figure 7). In this way, technology components may be reused for multiple interaction contexts, increasing flexibility of the architecture as a whole. In the following section (4), the systems used for child interaction evaluation utilise a subset of the potential components (figure 7), being focussed on activity-centric rather than component-centric evaluations.

4. HRI validation and exploratory experiments

ALIZ-E is currently in its second year of development so a number of the technologies required to support extended child-robot interaction are still in development. The development process is

[6]Open Source software, available from www.gostai.com/products/urbi.

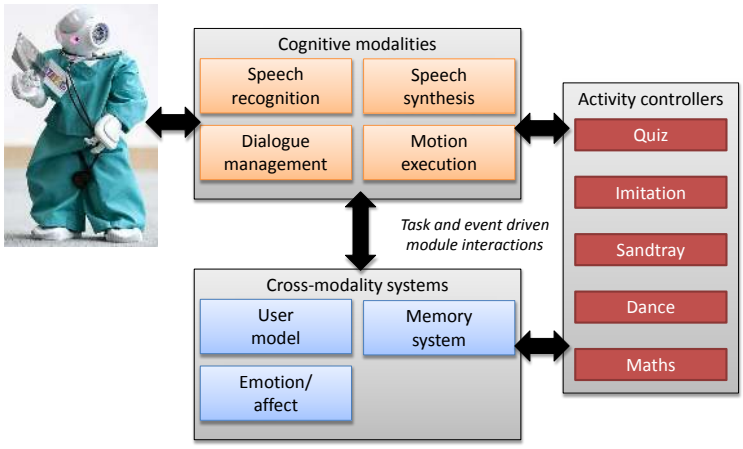

Figure 7. Overview schema of ALIZ-E system integration: functional interaction of the various modules (including both the *cognitive modalities* and *cross-modality systems*) is based on the task context (*activities*), and on event-based processing, handled in the robotic middleware Urbi.

guided by yearly testing cycles with children in the hospital alongside experiments carried out at other locations (including schools and summer camps). Testing in the hospital is exclusively focused on the interaction activities developed specifically for children with metabolic disorders and the technologies required to implement those interactions. Additional experiments carried out by project partners enable evaluation of the methods and approaches to be used in the development process. This coupling of *in-situ* testing of child-robot interaction activities with patients, backed up by experiments focused on individual aspects of the implementation (see section 2), is a valuable source of feedback on the efficacy of particular technologies and the overall functioning of the system during interaction with child users. In this section, an overview of the results obtained in this second testing cycle is provided: these focus on the manner in which the Nao robot is viewed by children in terms of role and engagement, and thus on its potential to act as a companion in support of learning metabolic disorder management.

4.1 Adaptation of robot to child

An early experiment focused on robot adaptation to personality traits of the user, specifically the extent to which introversion/extroversion would affect the enjoyment of the interaction. Betty et al. (2004) provide an overview of observational factors for the inference of personality from behavior (e.g. duration of eye contact and frequency of eye gaze). Parents and teachers were asked to answer a question assessing the personality of a child and the children were given the same question plus the Big Five Questionnaire for Children (BFQ-C), a short big five personality questionnaire for children in Dutch (Muris, Meesters, & Diederen, 2005). Fifteen participants, aged 10-11 (6 girls, 9 boys), played an imitation game with an 'introvert' and an 'extrovert' robot (Robben, 2011). The experiment used a Wizard of Oz (WoZ) paradigm, where the robot is controlled by the experimenter without the child's awareness. The human executed the speech recognition, but was only allowed

to choose between predefined utterances (with a view to automation of the whole process in the future). The actions and responses of the robot were scripted and dependent on the recognized speech input. The extrovert robot challenges the child and moves faster, while the introvert robot makes encouraging and comforting comments and moves slower. Each of the children played with both the introvert and extrovert robots: half of them first played with the introvert robot, followed by the extrovert, and the other half started with the extrovert robot before interacting with the introvert robot. Each interaction with the robot lasted approximately 10 minutes.

The results indicate that the child-robot interaction was very positively evaluated across all conditions to the extent that a ceiling effect was observed. There was no significant difference in evaluation between the two different robot 'personality' types, showing that the children were not sensitive to the personality as implemented on the robots. Indeed, a significant Pearson correlation between how children perceived the robots ($r=0.552$; $p<0.05$) indicates that children assign similar personalities to both robots. As a side result, we noticed that it proved difficult to derive participant personality types as there was no correlation between parent, teacher and self evaluations, and observational measures of participant extroversion/introversion.

A follow-up experiment (Janssen, Wal, Neerincx, & Looije, 2011) looked at task adaptation where the WoZ-controlled robot adapted the difficulty of a mathematical test based either on the child's performance or on a standard learning curve in order to test the effects of user-specific task adaptation on intrinsic motivation (Deci & Ryan, 1985). Mathematical skills and adding multiple digit numbers are important skills to have for diabetes management, and as such, we are interested in how a robot can contribute to teaching mathematics. Intrinsic motivation was measured using questionnaires and a free-choice period: after the obligatory interaction with the robot, the child could chose to carry on with the robot or do an alternative activity, such as playing with a portable game console or reading a comic. Twenty children, ages 9-10 years, completed mathematical tasks and an imitation game in which the robot adapted the difficulty of the task. Each child interacted with the robot in three separate sessions. The self-report questionnaires did not show any significant difference between conditions but the responses to the free-choice item (to carry on playing with the robot or change activity) showed that children who had played with the robot that adapted the task on the basis of their performance, chose to spend more time interacting with the robot. The children played three sessions each. After the first session the mean interaction time for both conditions in the free-choice period was similar (around 2.5min). In the second and third session the mean time for interaction with the robot in the personalized adaptation condition was 2.1 minutes, compared to 1 min for the standardized adaptation condition. This difference was significant ($p<0.05$).

These experiments demonstrate the benefits of evaluating user-derived adaptation factors prior to implementation and of the overall value of adaptation to user characteristics. The first experiments showed that the robot is evaluated very positively, but that implementing personality based adaptive behavior does not appear to be the most effective approach. In the second experiment, interaction times were greater in the second and third sessions for the personalized adaptive robot than they were with the non-personalized version, supporting the hypothesis that adaptation to user characteristics is an effective aid to engagement.

4.2 From healthy to sick children

To explore the applicability of findings from healthy child users to a clinical setting we performed additional experiments focused on a health management quiz administered to diabetic (Blanson Henkemans et al., 2012) and non-diabetic children (Zalm, 2011). Pairs of non-diabetic participants were given a quiz comprising questions on four health topics administered over two sessions. Participants played a WoZ quiz game with either a physical robot or an on-screen virtual robot, followed by another session of the quiz with the other robot. In each session the two players were in competi-

tion to achieve the best score but the robot (real and virtual) was configured to perform at a level comparable to that of the child. Ten children, 9-10 years old, were tested on their knowledge of health topics prior to playing two sessions of the quiz game on the same topics. When participants were retested after the quiz sessions, no difference in retained knowledge was found for the two conditions. However, participants were found to have paid more visual attention to the real robot. There was a significant main effect of condition on duration of looking toward the physical robot ($F(1,7)=87.4$, $p<0.001$) and on looking toward the virtual agent ($F(1,7)=11.97$, $p<0.02$) with children looking longest at the robot ($r=0.94$). Furthermore, among the children who looked more often toward the robot, the difference with the virtual agent condition was significant ($p=0.05$). There was also a significant main effect of condition on the number of looks directed at the opponent ($p<0.05$, $F(1,8)=5.94$), again with more looks toward the robot than toward the virtual agent ($r=0.65$).

In a procedure with diabetic children (Blanson Henkemans et al., 2012), five diabetic children, 8-12 years old, played three diabetes-based quiz games against either an adaptive or a non-adaptive robot. The adaptive robot asked their name, favorite sport and favorite color. It referred to this information during the interaction and quiz. It also asked the child questions about how they were enjoying the game during play. We measured knowledge about diabetes, the duration of the quiz, and the subsequent attitude towards the robot and the quiz. Results showed an increase in diabetes knowledge ($p<0.05$) across both conditions with a trend toward a greater knowledge increase in the adaptive robot condition. There was also a trend for participants who had experienced the adaptive robot to give higher enjoyment ratings than those who experienced the non-adaptive robot (though not at significant levels).

Taken together these results indicate that physical robots can effectively engage the attention of young users and that adaptation to user characteristics can be a useful tool in supporting sustained interaction. Of particular interest is the observed trend for interaction with an adaptive robot leading to more effective knowledge acquisition. This potential effect is the subject of ongoing research, where we look at which aspects of adaptivity are most effective to support learning.

4.3 Toward child-robot engagement: perception of the robot

A round of hospital evaluations carried out at San Raffaele Hospital, Milan, has examined the robot's ability to establish and maintain a social bond with a child user. For this protocol, 19 children, ages 7-12 years (11 males, 8 females), were recruited, and 13 of them completed the full evaluation cycle of three interactions (full details may be found in (Nalin et al., 2012)). Of these children, six were affected by diabetes type I. Each child was allocated a period of three hours to interact with the robot, spread over three different sessions (on separate days). The robot was able to play three games (quiz, imitation, and dance), and the child was requested to select one of these games to be played in each session. The child could stop the game and/or interaction at any point. Since a single game was shorter than one hour, teh children were allowed to also try the other games for the remainder of their time with the robot. The three sessions allow the children to become accustomed to interacting with the robot and allow time for any intrinsic novelty effects to diminish.

In all games the robot's speech generation and nonverbal behavior is autonomous but a Wizard of Oz (WoZ) approach was used to support speech interpretation. At the end of each interaction session with the robot, a battery of self-report questionnaires were administered to assess, for example, the child's mood, enjoyment of the interaction, perception of the robot as a social actor, and feelings about the games played.

Some interesting observations can be made regarding the data obtained. Overall, participants report very positive reactions to the robot; children described themselves as feeling 'happy' in its company and found the interaction entertaining (e.g., figure 8b and figure 9). These data indicate that participants continue to be interested in the robot and to enjoy interacting with it even during the

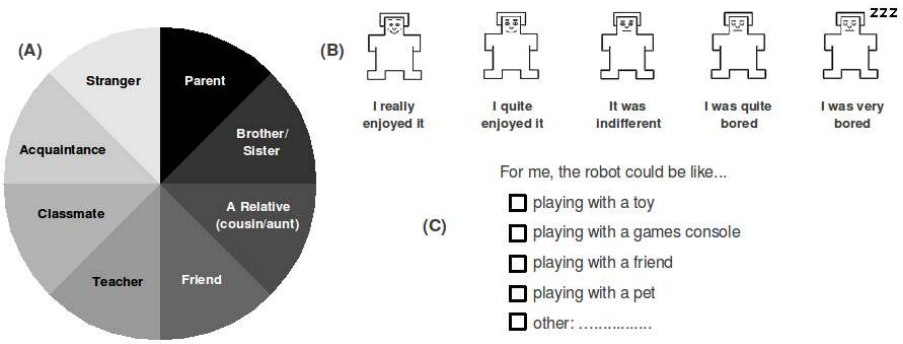

Figure 8. Examples of questionnaire questions administered to children post interactions: (A) pie chart describing potential social relationships; (B) 1-from-5 'enjoyment' rating with pictorial description; (C) forced-choice between descriptors. *Translated from the original Italian.*

third encounter when any novelty effects are likely to have been significantly reduced. The majority of children also reported that they would like to interact with the robot again, an effect which was observed regardless of which of the games were played. Over the course of the three sessions participants appear to become more used to interacting with the robot, as seen by an increase in the number of children describing themselves as feeling 'very relaxed' (see figure 9). This effect is likely to be in part explained by their repeated exposure to the robot, but may also reflect a development of social relationship between the pair. This conjecture is explored in the following paragraphs.

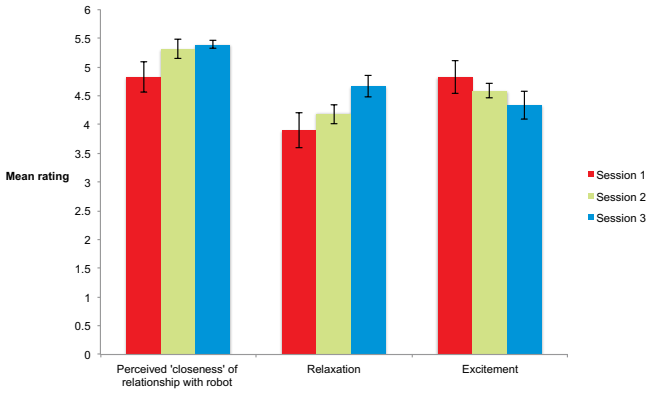

Figure 9. After each session children were asked to rate the interaction in terms of how close they perceived their relationship with the robot to be, how relaxed they felt in the interaction and how exciting they found it.

As a means to assess children's perception of the robot as a social actor we asked them to select a forced-choice label to describe their relationship with the robot. Participants were given a pie chart diagram with segments representing different social relations. They were asked to mark the segment which best fits how the robot appears to them and to write down the reasons for their choice

(figure 8a).

In the data gathered so far, the majority of children rate the robot as a close peer with 'friend' being the most common descriptor selected, followed by 'brother or sister' and 'classmate' (see figure 9). The labels 'acquaintance' and 'stranger' were those least often used. Labels were scored in descending order of closeness with a maximum score of 8 for parent down to 1 for stranger. These data suggest that the child developed a social bond with the robot. Interestingly, the labels 'acquaintance' and 'stranger' were usually selected by participants who experienced a robot malfunction during their interaction. Further scrutiny of these findings is required to test the strength of any correlation between technical malfunctions experienced during the interaction and subsequent assessments of the robot as an autonomous social actor.

Participants were also asked to assess their experience of playing with the robot by making a forced choice between comparative descriptions (figure 8c). Again, the majority of participants chose a label indicating that the interaction was viewed as being comparable to playing with another child, a finding which is supported by the data showing that a very high percentage of the children who indicated that they enjoyed talking to the robot and said they would respond honestly if it asked them an intimate question (which would not be expected if the robot were perceived as being inanimate). Participants were also asked to rate the extent to which they felt the robot understood the games they played and understood what they were saying to it. Children indicated that they thought the robot understood them, which is to be expected as speech interpretation was under WoZ control; however they also rated it as understanding the game, again supporting the view that participants saw the robot as an active social participant having agency. In many cases it was also observed that participants adjusted their own speech output (especially pauses and turn-taking in speech) to that of the robot, and sometimes even made gestures used during dialogue with the robot, such that speech interactions were observed to run more smoothly the longer they were sustained (Nalin et al., 2012). This alignment was informally observed but warrants further study.

5. Discussion and prospects

The research, development and evaluation in ALIZ-E provides useful results emerging from the cycle of testing at the hospital and in other settings. The various studies over a period of two years in school and hospital settings with children between 7 and 10 years of age has shown that children respond best to a robot which adapts its behavior to the young user. The robot as a physical embodied agent receives more attention than an on-screen avatar does, opening promising avenues for education and social interaction.

The challenges of CRI 'in the wild' are significant with both technical and pragmatic issues to be faced. Children are not 'mini-adults' and this fact is very much apparent in the context of CRI. Children bring an imaginative investment to encounters with robot agents that is immensely valuable in the exploration of how we can develop technologies and systems for social interaction. Conversely, they have often had exposure to highly sophisticated toys with complex (yet rigid) behavior patterns, and so the interest of a child is easily lost when the limits of a robot's responsiveness are discovered. Thus we have come to understand that sometimes the less complex but more robust and flexible behaviors are those which produce the best results with users.

The central feature of ALIZ-E is a coupling of innovative technology with embedding in a real-world application domain. Thus the project involves the development of new solutions to a number of significant issues in social HRI, some of which have a particular focus on child users e.g. novel technologies for parsing child speech and aspects of user modeling as applied to children. Other technologies are applicable across the range of social HRI applications such as flexible dialogue strategy management and associative memory-driven coordination of behavior. These methods and solutions are being developed and evaluated in parallel, and continuously integrated for testing with

child users both in a hospital setting and elsewhere. This rolling program of integration and testing obviously provides invaluable feedback on the real-time performance of the system. At the same time, each interaction session enables us to study the particular characteristics of this user group, learning more about what a robot needs to do in order to establish and maintain a social bond with a child.

One fundamental goal of ALIZ-E is to implement behavior to enable *social* CRI, i.e. the robot must be able to function as more than simply an object or toy in its interactions with a user. The pair should be able to play together. The robot must be able to initiate, participate, and collaborate in the interaction. Thus we aim for the robot to take a role similar to that of a peer or companion - with all of the user expectations this entails. By extension, for the interaction to be meaningfully social it must be maintained beyond the scale of moments meaning that the robot should be able to adapt its behavior to the needs of a child over multiple interaction episodes. In the ALIZ-E project to date we have laid the groundwork for these requirements to be met, demonstrating through the use of the various interaction activities that an adaptive embodied agent elicits responses consistent with the social role of peer or companion. The next stage of development and testing (of which the central developments have been outlined in this paper) will show the new technologies which have been developed, implemented and evaluated further to establish how far we are able to go in building robots capable of forming and sustaining social bonds with children.

6. Acknowledgements

This work is supported by the EU Integrated Project ALIZ-E (FP7-ICT-248116).

References

Baldridge, J. (2002). *Lexically specified derivational control in combinatory categorial grammar*. Unpublished doctoral dissertation, University of Edinburgh, United Kingdom.

Baldridge, J., & Kruijff, G.-J. (2003). Multi-modal combinatory categorial grammar. In *Proceedings of the 10th Annual Meeting of the European Association for Computational Linguistics.* Budapest, Hungary.

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in cognitive sciences*, *11*(7), 280-289 http://dx.doi.org/10.1016/j.tics.2007.05.005.

Baxter, P. (2010). *Foundations of a constructivist memory-based approach to cognitive robotics*. Unpublished doctoral dissertation, University of Reading, United Kingdom.

Baxter, P., Cuayáhuitl, H., Wood, R., Kruijff-Korbayová, I., & Belpaeme, T. (2012). Towards augmenting dialogue strategy management with multimodal sub-symbolic context. In *Proceedings of the 35th German conference on Artificial Intelligence (KI)* (p. 49-53).

Baxter, P., Greeff, J. de, Wood, R., & Belpaeme, T. (2012). "And what is a seasnake?": Modelling the acquisition of concept prototypes in a developmental framework. In *Proceedings of the international conference on development and learning and epigenetic robotics.* IEEE Press.

Baxter, P., Wood, R., & Belpaeme, T. (2012). A touchscreen-based 'Sandtray' to facilitate, mediate and contextualise human-robot social interaction. In *IEEE/ACM International Conference on Human-Robot Interaction (HRI2012)* (p. 105-106). http://dx.doi.org/10.1145/2157689.2157707.

Baxter, P., Wood, R., Morse, A., & Belpaeme, T. (2011). Memory-centred architectures: Perspectives on human-level cognitive competencies. In P. Langley (Ed.), *Proceedings of the AAAI Fall 2011 symposium on Advances in Cognitive Systems* (p. 26-33). AAAI Press.

Beck, A., Cañamero, L., & Bard, K. (2010). Towards an affect space for robots to display emotional body language. In *Ro-man 2010.* http://dx.doi.org/10.1109/ROMAN.2010.5598649.

Beck, A., Cañamero, L., Damiano, L., Sommavilla, G., Tesser, F., & Cosi, P. (2011). Children interpretation of emotional body language displayed by a robot. In *Proceedings of the International Conference on Social Robotics (ICSR2011).*

Beck, A., Stevens, B., Bard, K., & Cañamero, L. (2012). Emotional body language displayed by artificial agents. *Transactions on Interactive Intelligent Systems (Tiis)*, *2*(1), 2–29, http://dx.doi.org/10.1145/2133366.2133368.

Betty, H., France, L., Heisel, A., & Beatty, M. (2004). Is there empirical evidence for a nonverbal profile of extraversion?: A meta-analysis and critique of the literature. *Communication Monographs*, *71*(1), 28–48, http://dx.doi.org/10.1080/03637520410001693148.

Blanson Henkemans, O., Bierman, E., Janssen, J., Neerincx, M., Looije, R., Bosch, H. v. d., et al. (2012). A personalized robot contributing to enjoyment and health knowledge of children with diabetes at the clinic: a pilot study. *Patient Education and Counseling*, *accepted*.

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, *59*(1-2), 119–155, http://dx.doi.org/10.1016/S1071-5819(03)00018-1.

Burton, A., Bruce, V., & Johnston, R. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, *81*, 361-380, http://dx.doi.org/10.1111/j.2044-8295.1990.tb02367.x.

Cassimatis, N., Trafton, G., Bugajska, M., & Schultz, A. (2004). Integrating cognition, perception and action through mental simulation in robots. *Robotics and Autonomous Systems*, *49*(1-2), 13-23, http://dx.doi.org/10.1016/j.robot.2004.07.014.

Cuayáhuitl, H. (2011). 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* (pp. 9–15).

Cuayáhuitl, H., & Dethlefs, N. (2011). Optimizing situated dialogue management in unknown environments. In *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH), Florence, Italy* (pp. 1009–1012).

Deci, E., & Ryan, R. (1985). *Intrinsic Motivation and Self-determination in human behavior*. New York: Plenum Press.

Dekens, T., & Verhelst, W. (2011). On noise robust voice activity detection. In *12th Annual Conference of the International Speech Communication Association (INTERSPEECH-2011), Florence, Italy* (p. 2649-2652).

Draper, T., & Clayton, W. (1992). Using a personal robot to teach young children. *Journal of Genetic Psychology*, *153*(3), 269-273.

Fine, A. F. (Ed.). (2010). *Handbook on animal-assisted therapy: Theoretical foundations and guidelines for practice (3rd edition)*. London: Academic Press.

Frampton, M., & Lemon, O. (2009). Recent research advances in reinforcement learning in spoken dialogue systems. *Knowledge Engineering Review*, *24*(4), 375–408, http://dx.doi.org/10.1017/S0269888909990166.

Fuster, J. (1997). Network memory. *Trends in neurosciences*, *20*(10), 451-9, http://dx.doi.org/10.1016/S0166-2236(97)01128-4.

Gerosa, M., Giuliani, D., & Brugnara, F. (2007). Acoustic Variability and automatic recognition of children's speech. *Speech Communication*, *49*(10-11, http://dx.doi.org/10.1016/j.specom.2007.01.002).

Gonzalez, I., Sahli, H., Enescu, V., & Verhelst, W. (2011). Context-independent facial action unit recognition using shape and Gabor phase information. *Affective Computing and Intelligent Interaction*, *6974*, 548–557, http://dx.doi.org/10.1007/978-3-642-24600-5_58.

Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J., Lafourcade, P., et al. (2008). *The NAO humanoid: A combination of performance and affordability*.

Greeff, J. de, Baxter, P., Wood, R., & Belpaeme, T. (2012). From penguins to parakeets: A developmental approach to modelling conceptual prototypes. In *PG Conference on Robotics and Development of Cognition at ICANN 2012* (p. 8-11).

Hale, R. (2000). Book review: Sandplay therapy with children and families. *The Arts In Psychotherapy*, *27*(1), 75-76, http://dx.doi.org/10.1016/S0197-4556(99)00050-7.

Hawes, N., & Wyatt, J. (2010). Engineering intelligent information-processing systems with CAST. *Advanced Engineering Informatics*, *24*(1), 27-39, http://dx.doi.org/10.1016/j.aei.2009.08.010.

Hindriks, K., Neerincx, M., & Vink, M. (2012). The icat as a natural interaction partner. In F. Dechesne,

H. Hattori, A. Mors, J. Such, D. Weyns, & F. Dignum (Eds.), *Advanced agent technology* (Vol. 7068, p. 212-231). Springer Berlin Heidelberg, http://dx.doi.org/10.1007/978-3-642-27216-5_14.

Hindriks, K. V. (2009). Programming rational agents in goal. In A. El Fallah Seghrouchni, J. Dix, M. Dastani, & R. H. Bordini (Eds.), *Multi-agent programming* (p. 119-157). Springer, http://dx.doi.org/10.1007/978-0-387-89299-3_4.

Hiolle, A., Cañamero, L., Andry, P., Blanchard, A., & Gaussier, P. (2010). Using the interaction rhythm as a natural reinforcement signal for social robots: A matter of belief. In *Proceedings from the International Conference on Social Robotics (ICSR2010).*

Janssen, J., Wal, C. van der, Neerincx, M., & Looije, R. (2011). *Motivating children to learn arithmetic with an adaptive robot game.* Amsterdam, The Netherlands.

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: a field trial. *Human-Computer Interaction*, *19*(1), 61–84, http://dx.doi.org/10.1207/s15327051hci1901&2_4.

Kruijff-Korbayová, I., Athanasopoulos, G., Beck, A., Cosi, P., Cuayáhuitl, H., Dekens, T., et al. (2011). *An event-based conversational system for the Nao robot.* http://dx.doi.org/10.1007/978-1-4614-1335-6_14: Springer.

Lowenfield, M. (1939). The world pictures of children: A method of recording and studying them. *British Journal of Medical Psychology*, *18*(1), 65–101, http://dx.doi.org/10.1111/j.2044-8341.1939.tb00710.x.

Masuko, T., Tokuda, K., Kobayashi, T., & Imai, S. (1996). Speech synthesis using HMMs with dynamic features. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing Conference* (Vol. 1, pp. 389–392). IEEE, http://dx.doi.org/10.1109/ICASSP.1996.541114.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1, an account of basic findings. *Psychological Review*, *88*(5), 375-407 http://dx.doi.org/10.1037/0033-295X.89.1.60.

Morse, A. F., Greeff, J. d., Belpeame, T., & Cangelosi, A. (2010). Epigenetic robotics architecture. *IEEE Transactions on Autonomous Mental Development*, *2*(4), 325-339.

Muris, P., Meesters, C., & Diederen, R. (2005). Psychometric properties of the big five questionnaire for children (bfq-c) in a dutch sample of young adolescents. *Personality and Individual Differences*, *38*(8), 1757–1769, http://dx.doi.org/10.1016/j.paid.2004.11.018.

Nalin, M., Baroni, I., Kruijff-Korbayová, I., Canamero, L., Lewis, M., Beck, A., et al. (2012). Childrens adaptation in multi-session interaction with a humanoid robot. In *Proceedings of the IEEE RoMan Conference* (p. abstract). http://dxdoi.org/10.1109/ROMAN.2012.6343778.

Nalin, M., Bergamini, L., Giusti, A., Baroni, I., & Sanna, A. (2011). Children's perception of a robotic companion in a mildly constrained setting: How children within age 8-11 perceive a robotic companion. In *Proceedings of the Children and Robots workshop at the IEEE/ACM International Conference on Human-Robot Interaction (HRI2011).* Lausanne, Switserland.

Oveneke, M. C., Enescu, V., & Sahli, H. (2012). Real-time dance pattern recognition invariant to anthropometric and temporal differences. In *Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS2012).* 10.1007/978-3-642-33140-4_36.

Pammi, S. (2011). *Prosody control in HMM-based speech synthesis* (Tech. Rep.). DFKI Speech Technology Lab, ALIZ-E Project, Saarbrücken, Germany.

Read, R., & Belpaeme, T. (2012). How to use non-linguistic utterances to convey emotion in child-robot interaction. In *Proceedings of the IEEE/ACM International Conference on Human-Robot Interaction (HRI2012), Boston, MA.* 10.1145/2157689.2157764.

Robben, S. (2011). *Facilitate bonding between a child and a social robot: Exploring the possibility of a robot adaptive to personality*. Unpublished master's thesis, University of Nijmegen, The Netherlands.

Ros, R., Nalin, M., Wood, R., Baxter, P., Looiije, R., Demiris, Y., et al. (2011). Child-robot interaction in the wild : Advice to the aspiring experimenter. In *Icmi.*

Roy, N., Pineau, J., & Thrun, S. (2000). Spoken dialogue management using probabilistic reasoning. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics (ACL2000)* (p. 93-100). Hong Kong, http://dx.doi.org/10.3115/1075218.1075231.

Salter, T., Werry, I., & Michaud, F. (2008). Going into the wild in child-robot interaction studies: Issues in

social robotic development. *Intelligent Service Robotics*, *1*, 93–108, http://dx.doi.org/10.1007/s11370-007-0009-9.

Schröder, M., Charfuelan, M., Pammi, S., & Steiner, I. (2011). Open source voice creation toolkit for the MARY TTS platform. In *Proc. interspeech.* Florence, Italy.

Schröder, M., & Trouvain, J. (2003). The German text-to-speech synthesis system MARY: A tool for research, development and teaching. *International Journal of Speech Technology*, *6*(4), 365–377, http://dx.doi.org/10.1.1.122.9187.

Steedman, M. (2000a). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, *31*(4), 649–689, http://dx.doi.org/10.1.1.192.4027.

Steedman, M. (2000b). *The syntactic process*. Cambridge, MA: MIT Press.

Tanaka, F., Cicourel, A., & Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences*, *104*(46), 17954–17958, http://dx.doi.org/10.1073/pnas.0707769104.

Thomson, B. (2009). *Statistical methods for spoken dialogue management*. Unpublished doctoral dissertation, University of Cambridge, United Kingdom.

Wang, F., Verhelst, W., & Sahli, H. (2011). Relevance vector machine based speech emotion recognition. In *Proceedings of the 4th international conference on affective computing and intelligent interaction - volume part ii* (pp. 111–120). Berlin, Heidelberg: Springer-Verlag.

Williams, J., & Young, S. (2007). Partially observable Markov decision processes for spoken dialog systems. *Computer Speech and Language*, *21*(2), 393-422, http://dx.doi.org/10.1016/j.csl.2006.06.008.

Wood, R., Baxter, P., & Belpaeme, T. (2012). A review of long-term memory in natural and synthetic systems. *Adaptive Behavior*, *20*(2), 81–103, http://dx.doi.org/10.1177/1059712311421219.

Young, Y., Gasic, M., Keizer, S., Mairesse, F., Schatzmann, J., B., T., et al. (2010). The hidden information state model: A practical framework for pomdp-based spoken dialogue management. *Computer Speech and Language*, *24*(2), 150-174, http://dx.doi.org/10.1016/j.csl.2009.04.001.

Zalm, A. van der. (2011). *Help I need some body! The effects of embodiment on learning in children*. Unpublished master's thesis, University of Utrecht, The Netherlands.

Zen, H., Tokuda, K., & Black, A. W. (2009). Statistical parametric speech synthesis. *Speech Communication*, *51*(11), 1039 – 1064, http://dx.doi.org/10.1016/j.specom.2009.04.004.

Author contributions: P.B., T.B. and R.W.: memory model, H.C., B.K. and I.K.K.: natural language interaction, G.A. and V.E.: audio-visual processing, R.L. and M.N.: user modeling, Y.D. and R.R.E: machine learning, R.R., A.B., L.C., A.H. and M.L.: affective interaction, I.B. and M.N.: user evaluations, P.C., G.P., F.T. and G.S.: speech recognition, and R.H.: integration framework. T.B. is the project coordinator, R.W., P.B. and T.B. coordinated and edited the paper, contact: Tony Belpaeme, Centre for Robotics and Neural Systems, Plymouth University, United Kingdom. Email: tony.belpaeme@plymouth.ac.uk.