Child-Robot Interaction in The Wild: Advice to the Aspiring Experimenter

Raquel Ros Espinoza
Imperial College London
r.ros-espinoza@imperial.ac.uk

Marco Nalin
Hospital San Raffaele
nalin.marco@hsr.it

Rachel Wood
University of Plymouth
rachel.wood@plymouth.ac.uk

Paul Baxter
University of Plymouth
paul.baxter@plymouth.ac.uk

Rosemarijn Looije
TNO
rosemarijn.looije@tno.nl

Yiannis Demiris
Imperial College London
y.demiris@imperial.ac.uk

ABSTRACT

We present insights gleaned from a series of child-robot interaction experiments carried out in a hospital paediatric department. Our aim here is to share good practice in experimental design and lessons learned about the implementation of systems for social HRI with child users towards application in "the wild", rather than in tightly controlled and constrained laboratory environments: a trade-off between the structures imposed by experimental design and the desire for removal of such constraints that inhibit interaction depth, and hence engagement, requires a careful balance.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems—*software psychology*; I.2.9 [**Artificial Intelligence**]: Robotics—*commercial robotics and applications*; J.3 [**Computer Applications**]: Life and Medical Sciences—*health*

General Terms

Design, Experimentation, Human Factors, Theory.

Keywords

Child-robot interaction, experimental practice, human-robot interaction, interaction design, social robotics.

1. THE CHALLENGES OF LONG-TERM HUMAN-ROBOT INTERACTION

Over the last few decades electromechanical progress in robotics has outpaced developments in artificial cognition. This has been noticeable across the various sub-fields of robotics, but never more so than in social robotics. Robots embedded in a social environment not only need to deal with the challenges encountered by field robots, but in addition need capabilities for coping with highly dynamic and stochastic elements inherent in the social milieu. As such, social robots have to face a number of challenges which are to a large extent still beyond the current state of the art in science and technology. The three main challenges yet to be surmounted by social robotics are:

1. *Contingent responding in dynamic social environments:* In field robotics, an action by the robot does not necessarily induce a significant reaction from the environment: an unmanned vehicle driving around a rock, does not precipitate a response from the rock. In a social environment however, an action by a robot will almost always prompts some reaction from the user or users. Endowing robots with capabilities for contingent and appropriate handling of user responses, plus some capacity for the prediction of responses to their subsequent actions, is one of the main challenges in HRI.

2. *Interpretation of social sensory information:* in contrast to robots which do not engage in interaction over natural communication channels, social robots are required to make sense of social signals. For interactions that go beyond proto-social responses, this has proven to be a tremendous challenge. The multi-modal aspect of social interaction thus continues to elude us, but even the interpretation of single modalities, such as gesture or language, has proven to be extremely challenging. The early promise of multi-modal communication channels constraining the interpretation of social information has delivered limited progress.

3. *Designing sociable robots:* while progress in the electromechanical design of robots has been impressive and we seem to have a fairly clear understanding of which elements of the physical embodiment of a robot engage the user, the biggest challenge remains the creation believable, contingent behaviour. Arguably, this largely stems from our inability to endow robots with the capacity to correctly interpret the social world,

ICMI'11, November 14–18, Alicante, Spain.

thus hampering their ability to construct appropriate responses (i.e. linking points 1 and 2, above).

As such, the study of multi-modal human-robot interaction has been largely performed through using Wizard of Oz (WoZ) techniques, whereby an operator stands in for the socially cognitive capacities of the robot. Clearly, WoZ does not necessarily in itself form a problem: for some applications of human-robot interaction remote control suffices. The use of robots to interact with children with Autistic Spectrum Disorders, for example, seems promising and does not necessarily require the robot to be autonomous [13]. However there are application domains in which WoZ control is simply not sufficient or practicable for real-world applications.

We wish to use this paper to share good practice and insights gained from a series of human-robot interaction experiments executed in the paediatric department of a hospital. Too often, valuable practical information is glossed over in publications and is only accessible through word of mouth, the *practice* of experimentation being almost seen as trivial compared to the empirical data gathered from studies. We aim to redress this imbalance and in the next sections will give an overview of evaluation methods which we have successfully and less successfully used in child-robot interaction studies, hoping to provide some hard-won insights into the design of child-robot evaluation studies.

1.1 Child-robot interaction

It is clear that social interaction with robots is challenging, but there is a point of entry to human-robot interaction which is, in some respects at least, easier. Children, for reasons not fully understood, respond much more readily and strongly to social robots. As such, human-robot engagement is significantly more easily attained with younger children than it is with adolescent or or adult users. Short-term human-robot interaction is the low hanging fruit of the field, toddlers and young children respond very well to robots with a very basic social repertoire. Several studies have even shown how such social interaction with robots can be temporally extended.

Tanaka et al, for example, showed how a small humanoid robot was perceived as social agent by toddlers (1.5 to 2 years of age) and how the quality of the interaction improved as the robot spent more time in the day care centre [19]. The robot had a wide repertoire of behaviours, but it appeared that simple reactive routines (for example, touching the head resulted in the robot playing a pre-recorded giggle) were surprisingly effective at engaging the young children.

Tanaka et al. concluded that for this particular age-group haptic interaction is highly salient, an interpretation supported by the finding that participants often touched and hugged the robot. Higher-level behaviour, such as linguistic interaction, was not implemented and seemed unnecessary to engage children in this age range. Kanda et al. introduced a humanoid robot in a primary school setting, with children aged between 10 and 11 years. The robot was able to identify children using RFID tags and used a gradual "unlocking" of new behaviours and sharing of "secrets" as children interacted with it as a means to maintain their engagement over an extended period of time. While some of the children stayed engaged over the full duration of the experiment (2 months), about 2/3 of the children gradually lost interest. It was unclear if this was due to the robot's behaviour, or due to other external factors [7].

These findings serve to demonstrate two fundamental issues in the design and implementation of robots intended to engage socially with human users. Firstly, the problem of age, young children are very willing endow the behaviour of artificial agents with social meaning and to interact with them accordingly. This finding may well reflect the predominance of imaginative play as a means for exploring and learning about the world in early childhood. Older children appear to be less likely to view a robot as a social actor and maintaining their engagement is more challenging.

The second issue concerns the interactive complexity of the system. Tanaka [19] found a simple, reactive giggle behaviour to be highly effective in engaging young children. It is reasonable to expect that in order to engage with older users a robot would require more complex behaviours and a richer repertoire of social responses. Yet Kanda's findings provide evidence for a phenomena which is easily observed in HRI, the more sophisticated the robot, the greater a user's expectations of it and thus the greater the risk of their disengaging on discovering the system's limitations.

2. THE ALIZ-E PROJECT AND INTERACTION SCENARIOS

The ALIZ-E project[1] seeks to address the challenges in human-robot interaction described above by emphasising the principle of coordinated multi-modal, any-depth, interaction. The aim is to develop the state-of-the-art in human-robot interaction through the application of this principle to the development of companion robots for children engaged in residential treatment programmes in a hospital environment, lasting on the order of four to five days. It is thus necessary to provide the companion robot with the capacity for extended interaction, and over multiple, separate interaction episodes. The maintenance of engagement with the child over this time is of central importance: the multi-modal approach must support an engaging, naturalistic interaction. There is an additional emphasis on robust interactions, where the robot should be capable of sustaining an interaction (albeit in a more limited fashion) in the face of partial disruption to some of its capabilities (i.e. any-depth).

Despite the particular application domain for this project being child-robot interaction in a hospital setting, the issues involved are general to the wider concerns of HRI. The question of how engagement should be maintained is brought into sharper focus with the emphasis on extended interactions, for two reasons. Firstly, and almost trivially, there is more time/scope for errors to occur, leading to a breakdown in engagement. Secondly, with multiple interaction episodes, with a potentially variable amount of time between them, comes increase variability in the context of interaction, requiring differing social behaviours, all of which must be handled. Indeed, the emphasis on robustness is important for a similar reason: since engagement over multiple interaction episodes is likely, for children in particular (see above, section 1.1), to be linked to the naturalistic, adaptive nature of the robot. Given that the goal is multi-modal interaction, should one of the interaction modalities fail (either through technical failure, or environmental constraints), then the interaction should not stop, but should gracefully degrade,

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

and rely on the remaining modalities. This must therefore be a general goal of human-robot interaction: if a robot in a noisy room cannot maintain even a rudimentary social interaction because the speech recognition system is impaired by transient environmental conditions, the robustness requirement is not addressed (with similar scenarios for other modalities).

The ALIZ-E project approaches these problems in an iteratively defined interaction scenario methodology, where a companion robot is designed to interact with a child in different social contexts, using a variety of multi-modal competencies [2]. The following section describes a number of these. In so doing, valuable lessons can be learned regarding the use of social techniques and of multiple modalities in the maintenance of engagement with real children: problems that may be glossed over in simulation studies are, in this context, brutally exposed.

The target group for the ALIZ-E project are diabetic children in the 8-10 years age range. The project aims to develop companion robots able to engage child users and support them in learning about and managing their metabolic condition. Thus a set of interaction scenarios were designed to provide structure for the implementation and evaluation of encounters between robot and child. The following scenarios were developed as a means to assess our progress towards this aim.

2.1 The Dance scenario

One of the desirable outcomes of interaction with the robot is that child users should be encouraged to be physically active as a means to manage weight (an important issue in diabetes care) and to support general health. Physical education also supports the acquisition and development of motor skills. The development of such skills may be influential in fostering a child's interest in a broader range physical activities, thus and more importantly, in a healthy lifestyle [1, 12]. For these reasons, it is fundamental to ensure adequate motor skill development during childhood.

Dance is a physical activity that not only promotes the development of motor skills, but also allows the acquisition of other important attainments such as social compentence (cooperation, leadership, communication, awareness of others, sense of belonging), emotional expression and creativity [16]. Current state-of-the-art humanoid robots, such as the NAO robot used in this project, are physically embodied agents capable of relatively complex patterns of movement such as dancing. These behaviours have the capacity to engage human users, particularly children: a dancing game was thus chosen as an interaction scenario [15].

Social components of the Dance scenario.

The scenario was constructed around the use of social-cues to encourage the child to participate and foster the formation of a temporally extensible social bond between child and robot. Robot dancers have been proposed previously (eg. [18, 20, 10]). However these approaches have focused on individual aspects of dance, such as rhythm, entertainment or the level of interaction. The main motivation here is to advance the study of child-robot social interaction, however we also aim at promoting the development of physical skills based on each individual child's capabilities.

To increase familiarity with the user during the overall interaction the robot refers to the user by her name from time to time. This information is given *a priori* to the robot. Regarding the user's feedback, it is manually typed in through the WoZ. Both positive (e.g. "*Nice!*") and negative (though positively valenced, e.g. "*That was close! Can you repeat it?*", with a maximum of three failures before another move is proposed) verbal feedback is provided to the child throughout the interaction session.

The dance activity was designed to involve three consecutive stages, each with a specific functional and social component: introduction, exploration, and dance. The child had the opportunity to halt the interaction at any point in this scheme.

1. Introduction: the robot greets the child and shows a sample dance. From the task performance perspective the goal is to exhibit the robot's motor capabilities to the user and to draw her attention to the task. From the social perspective the primary aim is to enjoin the child to participate in the activity (i.e. to dance together with the robot). This is achieved through three steps: initiate interaction (greeting), motivate (showing dance) and establish engagement (dancing together until the user elects to stop).

2. Exploration: at this point the child has agreed to begin the activity. The robot shows various different moves, one a time, to assess the child's capability to perform the various movements. As the robot executes a motion, it asks the child to execute it as well. This process is repeated, for each movement until an evaluation of the child's performance is obtained. Currently a WoZ is in charge of the evaluation. Based on this assessment the robot either proposes a new motion, if the evaluation was positive, or repeats the same motion, if the evaluation was negative. Three consecutive failures for a given motion are allowed before a new movement is automatically introduced.

The child is encouraged and supported in continuing to dance through the robot's adaptation of the game to suit her abilities, thus fulfilling the general aim of encouraging physical activity in the user. From the social perspective the challenge here is to maintain the user's engagement with the robot and hence the task. The social aspect requires careful handling of the evaluation process so as not to discourage the user with negative feedback. Thus, all evaluation feedback is designed to be positively valenced. Three main sources of feedback are given: move description (a brief description of the motion being performed), evaluation outcome and motivational feedback ("Go ahead", "You can do it!") to encourage the user to keep trying the moves.

Because we are interested in long-term interactions, the set of moves performed are stored after each session. In terms of task performance this affects the selection of moves: thus subsequent sessions are based on the outcomes of previous interactions in order to gradually increase the complexity of the moves proposed to the user while still taking account of her individual abilities. The feedback given to the user also varies based on the history of interaction. Hence, if a movement demonstrated previously is proposed again, the robot can address the user making reference to that earlier experience (e.g. "Remember this move? Let's try it again!").

3. Dance: the robot creates a sequence of moves for the dance using only those successfully performed in the previous stage. The sequence is performed along with music as

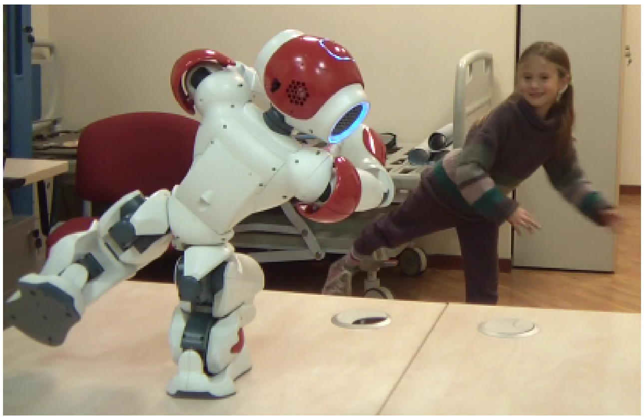

Figure 1: Child performing a motion with the robot during a dance session.

many times as the user desires (figure 1). Once the user decides to end the interaction, the robot thanks the user and the dance session ends. This stage of the session was found to be the most rewarding for the user and this was reflected in the greater levels of child-robot engagement.

2.2 'Simon Says' scenario

The Simon Game scenario is based on the "Simon Says" game in which players must do what 'Simon' tells them to do whenever asked with a phrase beginning with "Simon says". Thus the the player designated as Simon gives the directive "Simon says jump", the players must jump (those that do not that do not are eliminated). For the purposes of this interaction scenario the game was adapted to be played by two players (robot and child) and only arm movements were used, to ensure that the robot could perform all the moves adequately [5]. The game was further modified to include a memory component for the child such that the robot would demonstrate a move with one of its arms (left/right up, left-/right down) and the child had to copy this move and add another one. Next the robot repeated the moves from the child and added another one, up to a maximum of 5 movements. After this a new series began where the child started with the first move. In the case of an error a new series started.

When the robot moved its right arm up it was expected that the child should mirror the action. This was supported by the robot giving verbal descriptions of the moves that the child should make to mirror its actions (e.g. when putting its right arm up the robot would say "left up" to indicate the movement required from the child). The game was subsequently further adapted to incorporate simple mathematical exercises as a means to support the acquisition of competence in arithmetic which is a key skill for children with diabetes (e.g. for counting carbohydrates and measuring insulin).

Social components of the 'Simon Says' scenario.

The main focus of the Simon Game was to create an entertaining structure for interaction and to enhance companionship through shared activity, through the coordinated use of multiple modalities. The speech output of the robot was designed to support this. For instance the robot would explain that she and the child were secret agents that needed to learn a sign language and that after finishing a level they would be closer this shared goal. The complexity of the movements was incrementally increased, i.e. making it more challenging based on child performance, was envisaged to aid in the maintenance of engagement and motivation, which are proposed to be essential elements in making an activity enjoyable for the user. The general theme of the dialogue was motivational: the child was encouraged to go on and when she made an error the robot encouraged her to try again. In one experiment we included an additional manipulation providing different dialogues to children rated as extrovert and introvert in a pre-test evaluation. Previous research has shown that different forms of verbal encouragement are effective (e.g. "you can do better" vs. "I really believe you can do this") [14]. No significant difference between groups was observed in this experiment, however in future implementations we plan to adapt the dialogue component further to reflect the speech style of the child, as a additional means to foster feelings of companionship [8].

2.3 Quiz scenario

In this interaction scenario the robot played quiz master with the child answering the questions. The questions were all multiple choice and related to health, with the goal of supporting the educational goal of the child in the hospital [8]. In the case that the child answered incorrectly she was given another turn, if the question was answered incorrectly a second time the robot provided the correct answer and a new question. After each correct answer the child received a point.

Social components of the Quiz.

The main focus of the quiz was educational but again using a game format to motivate the child and foster feelings of social engagement with the robot thus supporting sustained interaction. This scenario was primarily aimed at testing the capacity of a companion robot to provide support to a child in learning about their medical condition. The feedback given to the child after an incorrect answer was designed to have a positive valance and verbal feedback was provided after each answer provided by the child, regardless of whether its was correct or incorrect in order to facilitate the individual learning process. In future implementations we would like to increase the element of joint activity by changing the arrangement from robot as quiz master to a trivial pursuit setup in which both child and robot alternate in the quiz master role and they compete for the best score.

3. LESSONS LEARNED

Testing human-robot interaction in a hospital setting with young patients is certainly no trivial undertaking. There are several obstacles to be overcome and a number of issues that must be carefully considered, both during the design and the execution of possible experiments. In this section we try to provide a comprehensive summary of these aspects, based on experience gained in the ALIZ-E project, during testing of first year prototypes in the paediatric department of the San Raffaele Hospital, in Milan.

3.1 Legal and administrative issues

One important consideration for HRI experimentation is the regulation of new technologies in institutional settings such as hospitals. These regulations vary between settings and across Europe but certain constraints apply almost universally. One such relates to CE[2] marking of devices. Problems arise for the experimenter when prototypes are used with modifications of the original hardware or software which effectively invalidate any CE mark the product may have. Introducing non-certified devices that can come in contact with end-users (patients) in an hospital environment (for example) is a major issue and in this case the prototype must be exhaustively documented and additional electromagnetic and mechanical tests may be required before local ethical approval can be obtained.

Another issue concerns *confidentiality.* When experimenting with participants drawn from this population two issue arise 1)participants are minors and 2) participants are patients. It is thus crucial that data are collected in compliance with existing confidentiality procedures, and that data collection is limited to the information strictly necessary for the experimental hypothesis under test. Particular issues can arise with the collection of audio/video data, but also in case of collection of sensitive patient information which the experimenters may wish to share with other research institutes, for example in multi-centre studies. An additional issue can arise through the use of sub-symbolic network architectures for instance as used in ALIZ-E for implementation of memory system functionality [3]. Clearly there is a requirement for experimenters to be fully cognisant of the confidentiality requirements of the all the institutions involved plus any relevant legislation and to have robust procedures for anonymising their data.

3.2 The experimental environment

The studies described here were carried out in a hospital setting which presents some additional challenges to the experimenter, while the majority of these problems are present to some extent in all institutional settings (e.g. schools and day-care facilities). There are some aspects of the hospital environment which are more or less unique, for example the fact that hospitalization is very often unavoidably traumatic, and the patient may have to undergo intrusive or painful procedures administered by unfamiliar adults. For diabetic children (the focus group for ALIZ-E), hospital also means an intensive regime of dietary control, blood sugar monitoring and insulin administration. All of these factors render the hospital environment a particularly challenging one in which to implement and test social engagement. On the other hand, if the particularly stringent requirements for setting up non-trivial levels of social interaction between child and robot can be met in a hospital setting then this provides a very strong indication of the efficacy and robustness of the approach taken.

One important issue in designing studies with children and robots concerns the setting for the experiment, finding a neutral space (i.e. in hospital avoiding areas associated with treatment or other potentially traumatic events). The appearance of the room, of the robot and even the experimenters needs careful control in order to avoid biasing the child's initial impressions and expectations of the interaction. Clearly the ideal is to create a friendly environment in which the child can feel comfortable, however these effort must necessarily be balanced against technical requirements of the interaction (e.g. neutral background colours to support the functioning of vision systems on the robot, minimising extraneous noise to support speech processing and so on) It is certainly valuable for experimenters to distinguish themselves as far as possible from medical staff adopting a *casual* style of dress to avoid the putative 'white-coat effect'[3].

Another key aspect of experimentation which is particularly emphasised in the hospital setting is the *recruitment* of participants for the experiments. In-patients have a schedule of procedures during the day which is very dynamic and dependent on factors which are hard to predict (e.g., availability of X-Ray machines, magnetic resonance, surgery theatres, etc.). In addition, medical personnel will discharge patients as soon as they are deemed fit enough to return home, these decisions often cannot be predicted in advance and this cannot easily be predicted in advance. Thus suitable participants who have agreed to take part may be discharged from hospital before testing can take place and this is another factor which experimenters must take into account when scheduling periods of experimentation with clinical populations.

A further, important factor, is the requirement for experimenters to liaise with the persons responsible for a child's care in order to fully assess their suitability for the study, in a hospital this may entail discussions with not just doctors and nurses but also psychologists and other therapists. In the experiments at San Raffaele Hospital, many apparently eligible children had to be excluded from the study for reasons other than their primary medical condition. This factor can introduce significant delays into the running of experiments and while a particular issue in hospital based studies can also be relevant in schools and other institutional settings.

3.3 Working with children

Conducting studies involving child participants places significant constraints on the experimenter. In the context of HRI, younger children are generally very willing to view the robot as more than a simple mechanical object (readily anthropomorphizing artefacts [21]). As has been previously demonstrated [11], children have the tendency to attribute *human characteristics and behaviors* to robots (at least to anthropomorphic robots). For this reason, where the aim of the experiment is to foster *social* interaction between robot and child, it is important that the procedure for introducing the robot to the child should take this into account. For example, the naming of the robot and the way in which it is first presented to the child can have a profound impact on the way in which it subsequently perceived and the 'believability' of the interactions. Older children may attribute social capacities less readily to a robot, making it all the more important that the initial encounter between robot and participant is carefully managed. Of particular importance are behavioural expectations such as contingent responding and behavioural robustness (e.g. not falling when walking) which if violated tend to significantly diminish the quality of subsequent interaction.

The *briefing* of the child plays an important role and should if possible prepare the child for potential problems in the robot's behaviour, for instance if it stops moving or

[2] *Conformitè Europèenne*: declaration of product conformity to the relevant European directives

[3] The association between the doctor's traditional white-coat and stress.

runs out of battery power. Ideally these issues should be dealt with with an explanation from the robot itself but it is also helpful if the child is given some indication of such possibilities in advance so as to avoid leaving the participant unsure as to what they should do in such a situation. In the current series of experiments an adult was present ion the room throughout and was able to mediate the interaction in the case of mechanical and or software problems. This person did not interact with the child or the robot during the experiment, but was present solely to intervene in the event of a technical problem, The adult chaperone was instructed to pretend not to be following the interaction (being seated at a desk out of the child's sight-line and apparently doing some paperwork) so as to allow the child to respond as spontaneously as possible. The number of times the participant referred to the adult in the room (with speech, or with glances, etc.) for clarification during the scenario was also taken as an additional metric for assessing the quality of interaction with a greater number of clarification gestures from the child taken as an indicator of reduced engagement. As a further means to support the child in feeling comfortable in the experimental set up she was invited to participate, briefed and accompanied in the experiment by the same person. Clearly there is a risk that the presence of this 'known' adult might introduce some unspecified bias in the child's response to interaction with the robot. However, in this context, with a specific aim of fostering *social* interaction, we considered it more important to ensure that the child was at ease than to adhere to the strictest tenets of empirical protocol. This is another area where a trade-off must be devised between the requirements for setting up and maintaining an interaction and the necessity for experimental rigour.

The *debriefing* is also an important aspect of the protocol. It is very likely that experimenters will need children to fill in questionnaires as a means to assess their experience of an interaction. In a our hospital based experiments this did not prove to be a major issue as patients are well used to filling forms and answering questions. In some cases this was even taken by participants as an enjoyable extension, somehow, of their experience with the robot. these effect might be explained by the fact that very often in hospital children are quite bored and welcome any variation of the usual daily routine.

Clearly, a simple "Thanks and goodbye" is not appropriate for terminating an experiment with child participants. It is important, if possible, provide some follow-up activity to remind the child of the experience and allow them to express any subsequent reactions to it. One example of this follow-up procedure, that was used in conjunction with the experiments described here, is the creation of comics with the robot as subject, that can be left with the children and used as a basis for talking about the experience subsequently. Simple certificates showing an image of the robot and some expression of official thanks were also found to be a useful way to demonstrate appreciation for the child's contribution in taking part. Interestingly, some children asked if there is a website where they can follow the development of the robot.

3.4 Planning the interaction

The design of the interaction itself requires additional care if the experiment involves children. We noticed that, while with an adult if the robot suddenly stops participating in the interaction (e.g., through a technical failure, falling over, or some other error), is not a major issue, child participants are quite disappointed by such 'breaks' in the flow of their joint activity. This implies that it is very important, when dealing with children, to have very robust code, that the robot must be mechanically reliable, and that the interaction scenario must be *tested exhaustively.* Furthermore the experimenter, should be aware mind that children are relatively *unpredictable* in their reactions, so it is important to stress test the robot with unexpected actions as much as possible in the lab environment, before going on with field experiments.

Another implication is that the robot (or the experimenter, especially in interactions using the Wizard of Oz) should have a *backup strategy* to finish the experiment earlier than planned in case of a technical failure. Often in these cases the data for that trial is automatically invalidated. However the backup strategy can still form a useful part of the protocol, permitting a degree of graceful degradation which might still generate some qualitative data from observation of the participant's response to the back-up strategy for mechanical or software failure. Strategies can even be varied for different experimental groups. For example when studying child-robot engagement, in the case of low battery participants in the experimental group might hear the robot say:

```
R: 'I'm sorry, I was really having
fun, but now my battery is too
low and I should rest'
```

while participants in the control group might hear:

```
R: 'Out of battery'
```

Clearly it is also important that the interaction should be as natural as possible, and that the robot should at least follow *basic social conventions.* Greetings, a personal introduction and a request for some basic details of the other person (e.g., "Hello, my name is R2D2, what's your name?") are minimal requirements for initiating an interaction. Furthermore, even though the experimenter has already explained what will happen beforehand to the child, it is still important that the robot is very clear in guiding the interaction, telling to the child what to do and when. Also the closing dialogue should be carefully designed, thanking the child for the fun they had together and hoping to meet again soon, etc. Of these requirements must also be adapted to suit the particular characteristics of the experimental protocol under evaluation. The robot should also have a routine available for situations where the child chooses to end the interaction sooner than expected, the child has the option to stop the interaction at any time, and if he decides to do so, the robot should respond appropriately to the situation (or the WoZ provide it) and have a "gentle stop" strategy, again thanking the child, so as to ensure the participant does not feel she has done something wrong in choosing to withdraw.

Another element we found to be useful in making the interaction happen more naturally is the use of additional behaviors, not directly related to the experiment or interaction itself (and that of course will not influence its result), but that can be useful to support the *credibility* of the robot as something which is "living". An example is to implement a fake "breathing" behavior, which is constantly executed by the robot in parallel with other routines. Another example is to avoid the "statue" effect, making the robot move, continuously, albeit very slightly. Additional behaviors can also be inserted into verbal interaction, for example providing variability in the possible answers, or commenting now and then on what's been said, and asking specific questions related to what's happening and so on).

A final remark for experimenters, when dealing with children, regards the understandable desire to *make the child happy*. When an experimenter notices that a participant appears to be bored, especially during Wizarded interactions, there is strong temptation to somehow "fix the situation" (e.g., making the robot say something, etc.). While it is clearly appropriate in some cases to interrupt the actual experiment (e.g., if the child is becoming particularly anxious, or unhappy), it is important to maintain the designed protocols, as additional interventions will adversely influence the study by introducing biases that will invalidate the results.

3.5 Measuring and assessing HRI with children

The importance of conducting survey research with children directly has fairly recently been acknowledged [4]. Markopoulos et al. [9] point out the value of designing products for children with as children as product testers. Both Markopoulos et al and Bell provide guidelines to use when evaluating with children. Several things have to be taken into account when using questionnaires in general and specifically with children; the child has to comprehend the questions, the question itself should be retrievable (memory span of children is shorter than for adults), questions where the social desirable answer can be derived (no right or wrong answers) should be avoided, the scales should be easy to interpret (perhaps 3-point scale instead of 7-point and the use pictures to explain scales or forced choice questions). Research also shows that children have a strong primacy effect when answering questions, so the chances are the child will consistently choose the first or last item on a scale just because he/she did this before. One way to deal with this is to make the scales different for each question and explain this to the child [4]. Thus, while questionnaires are an important part of assessment, a great deal of thought must go into their design.

In this context, behavioural measures have their advantages (but are subject to subjective interpretation by raters, e.g. what constitutes laughter?, and to differences between users, e.g. expressive vs. non-expressive interaction styles). Some behavioral measures, like shared gaze and movement parameters, can be automatically measured, and such metrics have major advantages over coding by human raters: 1) automated measurement is the same for each child, 2) it saves a lot of time (no multiple raters required to look at videos). However, a drawback to these techniques is that they typically assume laboratory conditions (e.g. relatively controlled subjects, and uniform lighting conditions), which can generally not be guranteed '*in the wild*'. Manual coding of behaviors is subject to subjective interpretation, and to differences between users, but these can generally be at least partially overcome through the application of the relevant coding schemes.

4. PLAYFUL DESIGN

The observations drawn from our experiences with implementing and child-robot interaction in a hospital setting centre on common themes reflecting the necessity for taking account of the special requirements of the participant group and the setting in which the encounter takes place. While some of the lessons learned reflect more or less standard considerations for experimental design some less obvious requirements become apparent through such experimentation "in the wild". One issue concerns the trade off between efforts to foster naturalistic interaction and the need for robust experimental paradigms for testing the efficacy of an approach to implementation.

When we set out to create scenarios for believable robot-child interaction most often we envisage some notion of play as the basis for the encounter. Thus the robot is endowed with routines which allow it to initiate or participate in games which provide a structure for interaction with a child. There is a difficulty with this approach in that play, is in itself ill-defined, there is a clear common-sense notion of what it is to engage in play but the characteristics that define playful activity as such, are less well-specified. Huizinga characterises play as being distinct from the constraints of 'normal' behaviour and of reality [6]. In this view, the fundamental essence of play is its freedom to subvert norms for action and interaction. Thus in play, a child can designate a banana to stand-in for a telephone and can take the role of parent giving her mother the role of child.

However, when we set out to design and experiment with child-robot interactions we construct scenarios which afford the application of metrics for measuring quality of engagement between interactants and to test the effects of various manipulations. In the ALIZ-E project we have used games (such as "Simon Says") as the basis for evaluation scenarios. However, it has become clear from these experiments that the highly structured format of such scenarios can easily inhibit engagement on the part of the child participant. In this sense we found that less really can be more in child robot interaction.

As an illustrative example, we found that during the dance game, due to the mechanical requirements of the dance steps sometimes led to the robot being oriented away from the child part-way through the interaction. As a stop-gap measure, during the piloting phase, the robot was given a routine in which it would call on the experimenter to 'help' it turn back to face the child. If the child was no longer in view the robot's speech generation system would output an utterance of the form "Hey [experimenter's name], could you help me turn around so that I can see [child's name] properly".

This utterance was found to be at least as effective in fostering engagement with the child as the games themselves. Like Tanaka's giggling robot [19], a spontaneous request for help showed the robot as behaving in a socially contingent fashion (acting to regain a shared gaze space with an interaction partner) and as having some capacity to go beyond the constraints of the game scenario i.e. to take some action not directly involved in gameplay. Indeed, a similar effect is seen with children who interact with a spontaneously malfunctioning robotic artefact: engagement levels rise as the children are drawn into an interaction [17].

The 'request for help' example, though a seemingly trivial implementation detail which would never have come to our attention if the robot had been programmed to correct its own position from the outset, illustrates a more subtle point. In play based interactions, externally imposed structure, whether to ease the process of implementation or as a requirement of an experimental protocol can have the counterintuitive effect of inhibiting interaction depth. Simple reactive behaviours which allow a robot to behave in a socially contingent fashion and afford imaginative engagement on the part of the child participant can be considerably more effective in supporting engagement and long-term interaction than structured games and complex behavioural routines.

These observations serve to demonstrate that experimenting with children and robots requires careful thought, regardless of the setting. In the experiments described here a selection of interaction types are implemented with the robot taking varied roles (such as instructor, companion, and playmate) this allows more general observations to be drawn which are of relevance to a broad range of HRI applications.

5. ACKNOWLEDGMENTS

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

6. ADDITIONAL AUTHORS

Tony Belpaeme (University of Plymouth, tony.belpaeme @plymouth.ac.uk); Alessio Giusti (Hospital San Raffaele, giusti.alessio@hsr.it); Clara Pozzi (Hospital San Raffaele, clara.pozzi@fastwebnet.it)

7. REFERENCES

[1] Physical Education, The National Curriculum for England. http://curriculum.qcda.gov.uk.

[2] P. Baxter, T. Belpaeme, L. Canamero, P. Cosi, Y. Demiris, V. Enescu, A. Hiolle, I. Kruijff-Korbayova, R. Looije, M. Nalin, M. Neerincx, G. Sommavilla, F. Tesser, and R. Wood. Long-Term Human-Robot Interaction with Young Users. In *IEEE/ACM Human-Robot Interaction 2011 Conference (Robots with Children Workshop)*, 2011.

[3] P. Baxter, R. Wood, T. Belpaeme, and M. Nalin. On memory systems for companion robots: Implementation methodologies and legal implications. In *AISB Symposium on Human Memory for Artificial Agents*. AISB, 2011.

[4] A. Bell. Designing and testing questionnaires for children. *Journal of Research in Nursing*, 12(5):461, 2007.

[5] H. Cuayahuitl and I. Kruijff-Korbayova. Towards Learning Human-Robot Dialogue Policies Combining Speech and Visual Beliefs. In *To appear in proceedings of the Workshop on Paralinguistic Information and its Integration in Spoken Dialogue Systems, Granada, Spain, Sep 2011.*, 2011.

[6] J. Huizinga. *Homo Ludens*. Beacon Press, 1971.

[7] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro. A two-month field trial in an elementary school for long-term human-robot interaction. *IEEE Transactions on Robotics (Special Issue on Human-Robot Interaction)*, 23(5):962–971, 2007.

[8] R. Looije, M. Neerincx, and V. de Lange. Children's responses and opinion on three bots that motivate, educate and play. *Journal of Physical Agents*, 2(2):13–20, 2008.

[9] P. Markopoulos, J. Read, and S. MacFarlane. *Evaluating children's interactive products: principles and practices for interaction designers*. Morgan Kaufmann Pub, 2008.

[10] M. P. Michalowski, R. Simmons, and H. Kozima. Rhythmic attention in child-robot dance play. In *18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 816–821, 2009.

[11] M. Nalin, L. Bergamini, A. Giusti, I. Baroni, and A. Sanna. Children's perception of a robotic companion in a mildly constrained setting. In *IEEE/ACM Human-Robot Interaction 2011 Conference (Robots with Children Workshop) Proceedings*, 2011.

[12] A. Okely and J. Wright. The gender bias in fundamental motor skills tests - a new agenda needed! In *Proceeding of AARE International Education Research Conference*, 1997.

[13] B. Robins, K. Dautenhahn, R. T. Boekhorst, and A. Billard. Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society*, 4(2):105–120, 2005.

[14] S. Rollnick, W. Miller, and C. Butler. *Motivational interviewing in health care: helping patients change behavior*. The Guilford Press, 2007.

[15] R. Ros, Y. Demiris, I. Baroni, and M. Nalin. Adapting Robot Behavior to User's Capabilities: a Dance Instruction Study. In *HRI (Human-Robot Interaction) Conference 2011*, 2011.

[16] I. E. V. Rossberg-Gempton, J. Dickinson, and G. Poole. Creative dance: Potentiality for enhancing social functioning in frail seniors and young children. *The Arts in Psychotherapy*, 26:313–327, 1999.

[17] T. Salter, I. Werry, and F. Michaud. Going into the wild with child-robot interaction studies. *Intelligent Service Robots*, 1(2):93–108, 2007.

[18] K. Shinozaki, Y. Oda, S. Tsuda, R. Nakatsu, and A. Iwatani. Study of dance entertainment using robots. In Z. P. et al. et al., editor, *Edutainment*, volume 3942 of *Lecture Notes in Computer Science*, pages 473–483. Springer-Verlag Berlin Heidelberg, 2006.

[19] F. Tanaka, A. Cicourel, and J. Movellan. Socialization between toddlers and robots at an early childhood education center. *PNAS*, 104(46):17954–17958, 2007.

[20] F. Tanaka, J. R. Movellan, B. Fortenberry, and K. Aisaka. Daily hri evaluation at a classroom environment: reports from dance interaction experiments. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 3–9. ACM, 2006.

[21] S. Turkle, C. Breazeal, O. Dasté, and B. Scassellati. Encounters with Kismet and Cog : Children Respond to Relational Artifacts. In P. Messaris and L. Humphreys, editors, *Digital Media: Transformations in Human Communication*, pages 1–20. Peter Lang Publishing, New York, 2006.