Adaptive Human Robot Interaction in Sensorimotor Task Instruction: from Human to Robot Dance Tutors

Raquel Ros\textsuperscript{a,*}, Ilaria Baroni\textsuperscript{b}, Yiannis Demiris\textsuperscript{a}

\textsuperscript{a}Personal Robotics Lab, Dep. of Electrical and Electronic Engineering, Imperial College London, Exhibition Road, SW7 2BT, London, UK
\textsuperscript{b}e-Services for Life \& Health, Research and Development Scientific Institute H San Raffaele, via Olgettina, 60, 20132, Milan, Italy

Abstract

We explore the potential for humanoid robots to interact with children in a dance activity. In this context, the robot plays the role of an instructor to guide the child through several dance moves to learn a dance phrase. We participated in 30 dance sessions in schools to study human-human interaction between children and a human dance teacher, and to identify the applied methodologies. Based on the strategies observed, both social and task-dependent, we implemented a robotic system capable of autonomously instructing dance sequences to children while displaying basic social cues to engage the child in the task. Experiments were performed in a hospital with the Nao robot interacting with 12 children through multiple encounters, when possible (18 sessions, 236 minutes). Observational analysis through video recordings and survey evaluations were used to assess the quality of interaction. Moreover, we introduce an involvement measure based on the aggregation of observed behavioral cues to assess the level of interest in the interaction through time. The analysis revealed high levels of involvement, while highlighting the need for further research into social engagement and adaptation with robots over repeated sessions.

Keywords: child-robot interaction, adaptive behavior, dance, involvement measure

1. Introduction

Human-robot interaction (HRI) has gained attention in the past years, not only to benefit people from the aid of a robotic system in daily activities, but also as a tool to study social development (ability to interact with others). Extensive surveys can be found in [9, 11].

A particular area of HRI research concerns children-robot interaction in therapeutic and edutainment settings. Nowadays children are very close to technol-
ogy in their everyday activities. However, the language, appearance, environment settings, motivational resources, evaluation tools, etc. used with children greatly vary from those used when interacting with adults [21]. Therefore, devoting research in this area has become crucial.

This research explores the potential for robots to interact with children around a task instruction in an edutainment setting. The ultimate goal is to provide an alternative tool to promote physical activity among young people, where the robot plays the role of a tutor in the first stages, and eventually, as a peer. In this paper we pursue the following sub-goals: (i) study the transfer of human behavior to robotic systems, (ii) implementation of a robotic system with a minimum set of capabilities to autonomously perform the task (except for perception); and (iii) evaluation of such an attempt in a real setting.

In order to study child-robot interaction we have focused on a dance activity. Dance is considered a fun activity that allows children to easily engage [35], probably due to the involvement of body movement during the activity which in turn increases enjoyment [4]. But more importantly it (i) stimulates physical movement, which is essential for the development of fundamental motor skills; (ii) enhances the development of social skills, such as cooperation, coordination, sharing ideas, sharing physical space, accepting individual differences; (iii) increases communication skills; and (iv) promotes the development of creativity and spontaneity [28, 22]. Considered an important educational source, dance is also promoted at schools as part of their curricula (e.g. in England it is part of the Curriculum for England in Physical Education [1]). Dance is not only a good practice for children, but for people in any age range, including elderly people where it has demonstrated to have physical and psychological benefits [14].

Moreover, from the HRI point of view, dance is an activity where embodiment is required. While verbal communication is indeed a valuable source of information, it is not crucial. The main means of communication is the body itself making it natural to interact with a robot. Therefore, we believe that sharing the space with a robot is a motivational resource that aids the user to effectively engage in the task as opposed to a virtual dancer in a screen.

The remaining of this paper is organized as follows. Section 2 summarizes research related to the work presented here. Section 3 reports the observations carried out to study dance sessions with children in schools. Based on these, the design and implementation of a dance tutor robot are introduced in Section 4. The experiments procedure and results are described in Section 5 and Section 6 respectively. We conclude the paper and point future directions in Section 7.

2. Related work

Early work on robots as social partners for children can be found in studies conducted by Kanda et al. [15]. A robot was introduced in a school to interact with children during their recess time. Children were free to interact with the robot as long as they wished. Contrary to our work, the robot was provided with several non-goal directed interactive behaviors which could allow for series of brief interactions.
In a similar setting, Tanaka et al. immersed the QRI0 robot in a classroom of toddlers (18- to 24-month old) during 5 months in three phases [33]. In the first and third phases the robot behaved using a full repertoire of random behaviors. In the second phase the robot performed a dancing activity. While high levels of interaction were found during phases 1 and 3, it significantly dropped during phase 2. The results on the dance study are reported in [35]. Two dance behaviors were evaluated: canned, the robot performed a pre-programmed movement while playing music; and interactive dance, the robot imitated the child’s moves using its upper body only. Each condition was evaluated for three days. An additional one is introduced later on, where only music is played in the room, without the robot being present. Unlike our work, children were free to go and interact with the robot at any time during the day. Comparing the different conditions, children went to visit the robot more often in the interactive mode, suggesting that this condition was quite attractive.

Work on developing a social robot with empathetic capabilities (facial expressions only) to play chess with children is presented by Leite et al. [19]. In their work 5 children played over a five-week period. The results show that the social presence perceived by the children decreases through time, when the novelty effect diminishes. They suggest that the low repertoire of actions the robot is provided with could be one of the reasons for the attention decay.

A study on the impact of social interactions on effective learning is introduced by Okita and Ng-Thow-Hing [25]. In their work they compare three learning strategies (lecture, cooperative and self-directed) in a table setting task, where the robot behaves either in a monotone way or a more human-like style (voice and gestures only). Results show that the combination of cooperative learning strategy and human-like behavior achieves higher performance results, suggesting that the learning styles and social features had an impact, especially in young users (4 to 6 year-old).

A different approach for learning through social interaction is proposed by Tanaka and Matsuzoe, where a care-receiving robot to promote children’s learning is studied [34]. They propose a scenario where instead of having the robot as tutor/teacher, the child becomes a teacher for the robot. 17 participants between 3 and 6 years-old took part in the experiments. The results suggest that the teaching framework proposed helped children learn English verbs efficiently.

Our work is placed in between these works, where the target age is 7-12 years-old, there is a specific task to be carried out, and the robot is provided with both specific engaging behaviors related to the task and high movement capabilities.

3. Analyzing Human-Human Instruction Sessions

The aim of having observation sessions is to study and to analyze human-human interaction within everyday activities that children take part in. While the global goal is to study interaction strategies used between a tutor and pupils, a more specific one within the dance context is to analyze dance methodologies and techniques to teach dance to children. Some of the questions we would like
<table>
<thead>
<tr>
<th>No.</th>
<th>Strategy</th>
<th>Implemented</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Abstract concept vs. specific motion learning</td>
<td>✓</td>
</tr>
<tr>
<td>2.</td>
<td>Individual vs. collective training</td>
<td>✓</td>
</tr>
<tr>
<td>3.</td>
<td>Non-verbal and verbal feedback</td>
<td>✓</td>
</tr>
<tr>
<td>4.</td>
<td>User specific adaptation</td>
<td>✓</td>
</tr>
<tr>
<td>5.</td>
<td>Incremental sequence learning</td>
<td>✓</td>
</tr>
<tr>
<td>6.</td>
<td>Performance and appreciation</td>
<td>✓</td>
</tr>
<tr>
<td>7.</td>
<td>Motivational feedback</td>
<td>✓</td>
</tr>
<tr>
<td>8.</td>
<td>Learning resources</td>
<td>✗</td>
</tr>
<tr>
<td>9.</td>
<td>Music</td>
<td>✓</td>
</tr>
<tr>
<td>10.</td>
<td>Creativity support</td>
<td>✗</td>
</tr>
<tr>
<td>11.</td>
<td>Problem identification and support</td>
<td>✓</td>
</tr>
<tr>
<td>12.</td>
<td>Activity summary</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 1: Summary of strategies observed during the dance sessions. The third column indicates whether they have been implemented in the current system.

to answer are: How does the teacher prepare a session? How does she/he adapt the teaching methodology based on the group to be taught? How to motivate children? How to keep them engaged?

A researcher attended 30 dance sessions which spanned four months in different schools in London\(^1\). The target age participants were between 7 and 10 years-old divided in 5 groups. Each group was composed of around 20 students. Only journal notes were taken by the researcher through the sessions since video recording instruments were forbidden in schools due to privacy concerns.

3.1. Observations summary

A summary of the observations annotated during the dance sessions we took part in is shown in Table 1. Annotations were focused on the observation of two main aspects: instruction methods and engaging/motivational techniques. While the former covers any type of instructions used by the teachers to convey knowledge, the latter refers to any relevant behavior the teachers performed to draw children’s attention towards the task. Moreover, discussions with the teachers regarding the methodologies used were also included.

1. Abstract concept learning (creative) vs. specific motion learning (guided): in the former methodology the goal is to teach body movement based upon a set of principles or concepts (e.g. shapes, actions) as opposed to teaching systematized exercises or skills (e.g. ballet positions) [32]. This latter approach is usually used in professional dance teaching. In this paper we adopted the specific motion learning due to its specificity property.

\(^1\)The dance sessions were organized by the Language Of Dance Centre.
2. Individual vs. collective training: working in pairs or groups (collective training) is in general appreciated by children when having to develop new ideas. However, other children prefer to work alone (individual training), and forcing them to work with others is not motivating at all. Due to the experiments’ set-up, in this work we only consider individual training.

3. Non-verbal and verbal motion feedback: the former corresponds to performing the motion itself (or examples), while the latter corresponds to providing verbal description of the motion. Both are essential to guide the children’s comprehension and learning. Non-verbal feedback serves as visual input providing most of the information they need to understand the concepts. Verbal feedback on the other hand is useful to help children recall and name the ideas worked. Thus, a combination of both strategies is applied where usually simple descriptive words are used to refer to the different moves while performing them.

4. User-specific adaptation: continuous observation and evaluation of the children’s capabilities to propose new challenges. Only when a movement is successfully learned, its complexity can be increased.

5. Incremental sequence learning: learning a sequence of movements adding one (or very few) movement(s) at a time. Every time a new movement is inserted, the overall sequence is repeated. Children end up learning the whole sequence by repetition. In further sessions, the teacher can go through the sequence faster, only using the associated verbal descriptors.

6. Performance and appreciation: children like showing the knowledge acquired and their own creations to receive approval from the teacher. Their motivation increases and they feel confident to be challenged in turn.

7. Motivational feedback: constant approval of how well children are doing the task encourages them to continue. When negative feedback should be given, always employ positive words to avoid disappointment.

8. Learning resources: additional visual resources are used to better illustrate the concepts taught. For instance: pictures, objects, videos, board, etc. are shown/used during the sessions. As suggested in [7], using different sources of information (visual, verbal, sound, etc.) promotes the use of different learning strategies, which impacts on the learning effectiveness.

9. Music: on the one hand it is used to guide the dynamics of the group and their movements (e.g., faster rhythms during warm-up). On the other hand, it serves as motivational support, especially to start the sessions. If children like it, they easily engage in the task and start moving around along with the teacher. Similar findings were reported in [35].

10. Creativity support: encourage the child to propose his/her own variations of the ideas learned. Children constantly repeating a given sequence risk feeling bored at some point. On the other hand, forcing the use of creativity from an early stage is not helpful either since a minimum guidance is most of the times required. A combination of a guided and a creative approach serves as a motivational technique to keep children engaged.

11. Problem identification and support: when a child has a difficulty performing a dance movement, the teacher identifies the problem and tries to
overcome it (e.g. providing more motion feedback, performing it at different speed, etc.). However, if there is no success after several attempts, the motion is replaced to avoid discouraging the child. In this work, only the last strategy is applied.

12. **Activity summary**: brief explanation of the goals and contents of the next stage of the session, linking the ideas with previous sessions if applicable. This process allows children to be aware of their learning progress.

### 3.2. Dance education framework

The dance teachers followed the framework proposed by the Language of Dance Centre based on an educational model. In this framework, the main goal is to allow children to explore and create body movement using building block movement primitives. Common stages in a dance session are the following:

1. **Warm-up**: the teacher starts the session by warming-up so that the body is ready to exercise.
2. **Exploration**: the teacher explains the concepts, one at a time, while practicing them with the children.
3. **Creation**: individually or in groups, the children create dance phrases (a sequence of movements) using the concepts learned in the exploration stage.
4. **Performance and Appreciation**: the children perform the dance phrases created during the session to show their work to the rest of the classmates. This stage enables them to discuss about the different performances identifying and reinforcing the learned concepts.

Stages 2 through 4 are repeated several times during a session. At each cycle a different concept is explained, explored and performed/observed.

### 4. Moving to a Robot Dance Tutor

We next describe the robot dance tutor approach. First we describe the overall robot behavior. We next go through the movement selection mechanism. Finally, we detail the additional engaging behaviors and the architecture.

#### 4.1. From Human to Robot

The goal of the robot’s behavior is twofold. On the one hand, it is to teach the child a set of movements to dance together. On the other hand, it is to keep the child engaged in the task as much as possible, not only to finish the task, but to repeat the encounter in future occasions if possible. An interaction history is maintained to support both, the robot’s task and social performance.

---

2Located in London, The Language of Dance Centre (founded by Dr Ann Hutchinson Guest, an expert in dance notation) offers dance clubs, professional development, certification courses and resources used in the Language of Dance® approach to movement and dance (http://www.lodc.org).
The dance activity was designed to involve three stages, each with a specific functional and social component. The stages were driven from the strategies (strat. x) and dance framework observed in the human-human sessions:

1. *Introduction*: the robot greets the child, introducing itself and shows a sample dance. From the task performance perspective the goal is to exhibit the robot’s motor capabilities to the child and to draw her attention to the task. From the social perspective the primary aim is to encourage 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 child elects to stop).

2. *Exploration*: the robot shows various different moves, one at a time, asking the child to repeat them (strat. 1) to assess the child’s capability to perform them. The robot executes a move along with verbal support (either through guided visioning, e.g. “the zombie step”, or through movement description, e.g. “bend one leg to the side”) and asks the child to perform it in turn (strat. 3). The robot autonomously selects the next movement to perform based on the child’s performance history (strat. 4 and 11). The child is encouraged and supported in continuing to dance through the robot’s adaptation of the activity to suit her abilities. The social aspect requires careful handling of the evaluation process so as not to discourage the child with negative feedback. Thus, all evaluation feedback is designed to be positively valenced (strat. 7).

3. *Performance*: the robot creates a sequence of moves to dance only based on those successfully performed by the child so far. The sequence is performed by both, robot and child as peers, along with music as many times as the child desires (strat. 6 and 9). The child can also opt for introducing more dance moves. In this case, the robot goes back to the exploration stage to try additional moves and the dance sequence is extended accordingly (strat. 5). This process can be repeated as long as the child agrees.

Each stage is briefly described before starting it (strat. 12). When the dance session is over, the robot thanks the child for her participation and says goodbye hoping to meet her once again. A diagram of the overall process is shown in Figure 1. The activity can be concluded for any of the following reasons: the child decides to finish the interaction, the duration of the session reaches a timeout, or failure in the robot system.

### 4.2. Selection Algorithm

Dance moves are composed of different body parts’ movements (e.g. moving the arms and legs at the same time). We represent a dance motion based on the limbs taking part in it. We consider the following limbs: head, arms and legs. Moves which are composed of more than one limb movement are referred to as *hierarchical* moves (an example is depicted in Figure 2). In the instruction activity their decomposition is first considered (i.e., only showing
legs or arms) and then the whole motion (i.e. legs and arms at the same time) allowing the algorithm to gradually increase the complexity of the motion when teaching. Motions may also require pre- or post-motion conditions. For instance, a movement on the floor requires the agent (either robot or human) to sit first. In addition, we also consider whether a motion corresponds to the mirrored version of another one, referred to as symmetric motion. Thus, the following representation corresponds to a dance move:

\[
dance\_move = [id, limbs, sym, precond, postcond]
\]

where \(id\) corresponds to the motion Id; \(limbs\), a list of body parts that compose the overall motion; \(sym\), the Id of the symmetric move; and \(precond\) and \(postcond\), the a priori and a posteriori poses, respectively.

Moves can be classified based on the relation between their occurrence and evaluation outcome as follows: \(new\), the motion has never been shown before; \(retry\), the motion has been performed before with a negative evaluation outcome; and \(repeat\), the motion has been correctly performed in the past.
The selection mechanism consists of a filtering algorithm driven by the evaluation outcome of the last dance move performed. \( \text{move}_t^t \) determines the type of movements (new, retry or repeat) to be used in the search. The initial search criteria is computed as:

\[
\text{move}_t^t = \begin{cases} 
  \text{“new/retry”} & \text{if } \text{eval}(m_{t-1}^t) = 1 \\
  \text{“new”} & \text{if } \text{eval}(m_{t-1}^t) = 0 
\end{cases}
\]

where \( m_{t-1}^t \) is the dance motion performed at time \( t-1 \) and \( \text{eval}(\cdot) \) is the evaluation function. Figure 3 depicts the flow diagram for the selection algorithm. First it removes the dance moves correctly performed in the current session from all available motions to give priority to new or failed past motions. From the resulting set, it selects moves with criteria \( \text{move}_t^t \) (as defined above). Next, it checks whether hierarchical motions are available, meaning that their decomposition have been successfully performed at some point. Otherwise, non-hierarchical motions or the decomposition of hierarchical ones are considered. This strategy allows the algorithm to gradually increase the complexity of the motions to learn. Finally, a motion is randomly selected from the available set. If no motion is available at this point, the algorithm is repeated but with criteria \( \text{move}_t^t = \text{“repeat”} \), i.e. the algorithm considers the motions correctly performed so far to include repetitions.

4.3. Engaging Behaviors

Additional cues for enhancing the social interaction were included in the design of the robot behavior. These types of behaviors prevent the robot from
being too static, being more human-like and provide additional feedback about the robot’s internal state [24]. In this work we make use of the following:

- **head movements**: random smooth head movements during idle periods. To emphasize the fact that the robot is expecting a child response [6] the robot faces him/her and remains still until an answer is provided.

- **blinking eyes**: the robot’s eyes are provided with colored LEDs. Turning them on and off simulates natural human eye blinking making the robot appearance more “alive” [37].

- **spatial orientation**: the robot turns or walks towards the child to maintain an acceptable spatial-orientational arrangement [17].

- **name reference**: random references to the child’s name to reinforce the bond between them [16, 34].

- **verbal feedback to provide robot’s internal state**: during idle periods inform the child about its internal state to avoid disengagement or confusion [24].

4.4. System Architecture

We next detail the different components of the system architecture. An overall description is illustrated in Figure 4. The implementation of the system has been done using Urbi³, an open-source platform to control robots.

- **Dance Manager**: robot behavior control in charge of executing the dance behavior described in Section 4.1. To support the decision making it i) receives inputs from the external world through an event-based approach and ii) requests relevant information to the User Model and the Dance Move Manager. Commands are sent to the verbal and non-verbal components for action execution.

- **Vision**: processes visual information centered on human activity (human presence, position and body configuration) through a kinect device. Although not yet integrated, this module should also provide the evaluation outcome of the dance moves performed by the child.

- **Speech**: in charge of processing the speech generated by the child. In the current implementation the automatic speech recognition system is not yet used due to low reliability.

- **WoZ**: provides external input to the system while the perceptual components are completed (vision and speech) based on a Wizard of Oz approach (WoZ [10]). It allows the operator to introduce the child’s both performance evaluation (right/wrong) and answers to questions posed by the

³www.ur biforge.org
Figure 4: Diagram of the system architecture. Arrows represent the flow of information between components. Note that the arrows from the vision and speech components connect to the Dance Manager through the WoZ, representing that in the current version these are wizarded.

robot (yes/no), and to re-position the robot so that it always faces the child during the interaction.

The module triggers the same events that the automatic components would generate, making it transparent for the other modules how these have been generated (i.e., either wizarded or not). This way we ensure a straightaway integration of available components.

• **Dance Move Manager**: in charge of the dance move selection algorithm described in Section 4.2, it provides the next action to perform. It manages the library of dance moves and provides the required information to the **Non-Verbal Behavior** for the execution of the moves.
• **User Model:** handles the users' interaction history to enable adaptation in the decision process. Besides storing general information about the child (such as Id, name, age and gender), it also keeps track of the dance moves performed through the different sessions based on their evaluation outcome. A dictionary-based representation is used:

\[
\text{user\_history}[\text{id}] = [\text{success}, \text{failure}]
\]

where \(\text{id}\) corresponds to the user Id and,

\[
\text{success} = [[\text{move}_1, \text{limb}_1], \ldots, [\text{move}_n, \text{limb}_n]]
\]

\[
\text{failure} = [[\text{move}_1, \text{limb}_1], \ldots, [\text{move}_m, \text{limb}_m]]
\]

where \(\text{move}_i\) is the Id of the dance move, and \(\text{limb}_i\) is the body part of the move under evaluation.

• **Utterance Generator:** generates the utterance to be verbalized by the robot. The following classes of messages are available: (i) greetings and farewell (e.g. “Hello, happy to see you”, “Thanks for dancing with me, it has been great!”), (ii) instructions (e.g. “Ok, first we’ll go through some steps...”), (iii) inquiries (e.g. “Did you like it? We can try it again!”), (iv) performance feedback (e.g. “Well done!”, “Almost, but let’s try it again”), (v) dance move description (e.g. “Open and close the arms like me”), and (vi) robot’s internal state (e.g. “Just a second, I’ll show you a motion soon”).

In the current version of the system the generated utterances are based on canned text which have been manually created. Variations of each message have been produced to avoid repeated verbalizations. In total 87 types of messages were generated, where 51 correspond to dance move descriptions (with up to two variations for each) and the remaining 36 correspond to the other types of messages (with two to three variations, except for performance feedback which may have up to 12 variations). Besides, utterances are also modified in two additional ways: the child’s name is included in a pseudo-random way, and references to past interactions are considered when available (e.g. “Hello again Tom”).

• **Text-To-Speech (tts):** two components are available for speech synthesis: the commercial Acapella TTS [12], which is provided with the robot, and the open source MARY TTS platform\(^4\)[31]. Both systems are available in English and Italian. The Italian voice for MARY TTS was constructed using the MARY TTS voice creation toolkit [30]. The advantage of MARY TTS is that it allows control of prosody through symbolic markup [26].

• **Body Motion:** manages the robot’s body motion. The output is sent to the robot motors through the Naoqi\(^5\) framework. Sources of motion are:

\(^4\)http://mary.dfki.de
\(^5\)http://www.aldebaran-robotics.com/documentation/dev/tools/naoqi.html
– dance moves: choreographic robot motion. Dance moves are manually created and exported using the Choreographe\textsuperscript{6} software. Each individual dance move is represented by a set of joints along with sequences of control points at given time-steps:

\[
dance\_move = \begin{bmatrix}
  j_1 & t_{11} & k_{11} & t_{12} & k_{12} & \ldots & t_{1n} & k_{1n} \\
  \vdots \\
  j_m & t_{m1} & k_{m1} & t_{m2} & k_{m2} & \ldots & t_{mn} & k_{mn}
\end{bmatrix}
\]

where $j_i$ corresponds to a joint, $t_{ih}$, the time-step $h$ for joint $i$ and $k_{ih}$, the control point $h$ for joint $i$. A smooth motion is obtained through a Bezier interpolation of these available in Naoqi. Partial move performance is available through a filtering process where only the joints corresponding to the body parts requested by the Dance Manager are executed by the robot.

– poses: static body poses represented by the joint angles.

– repositioning: walking commands that enable repositioning of the robot to its initial dance location and turning to accurately face the child. In the current setting, the operator sends the commands through the WoZ.

– Choreograph moves: built-in behaviors from the Choreographe software (restricted to standing and sitting in this work).

  • Blinking Eyes: emulates eye blinking with the eye LEDs, fading them out for a short duration and turned on back again at random intervals.

  • Head Motion: gradually moves the head in a random way within a given configuration space during periods where no other motion takes place.

We next describe the experiment procedure employed to evaluate the approach presented in this section.

5. Experimental methods

The experiments reported in this work are part of a broader experiment developed within the EU FP7 ALIZ-E\textsuperscript{7} project, where additional activities (also referred to as games) were presented to the children. Next, we present the outcomes related to the dance activity. Reports on the general experiment and other game activities part of the ALIZ-E framework can be found in [3, 23].

\textsuperscript{6}http://www.aldebaran-robotics.com/documentation/software/choreographe/index.html

\textsuperscript{7}www.aliz-e.org
5.1. Participants

A total of 12 children (7 male, 5 female) ranging in age from 7 to 12-years-old ($\mu = 8, \sigma = 1.91$) took part in the dance experiments. The experiments were ran at San Raffaele Hospital, in Milan, from March to May 2012, one day a week. The invitation to the experiments was sent through a diabetic patients' association and brochures displayed at the hospital. However most of the children were healthy and received the invitation through diabetic relatives or classmates/friends. Parents made an appointment by e-mail or phone to participate, if possible, for three sessions.

5.2. Material

The robot selected for this study was Nao, a humanoid robot from Aldebaran Robotics. The Nao is almost 58cm tall, weighs 5.2kg and has a cartoon-like appearance which was considered especially suitable for use with children. Its 25 degrees of freedom allows smooth motions especially required for dance movements. LEDs are located in the head, eyes, ears, chest and feet. Two loudspeakers are located in the ears, used in this work to generate speech and play music. It is also equipped with two cameras and a microphone (not used here since a wizard replaces automatic perception). Its processor is an x86 AMD 500MHz and it is equipped with Wi-Fi to allow remote access. The robot’s behavior is completely autonomous though requires the intervention of a Wizard of Oz for perception (visual and speech inputs). However, the robot was presented to the children as fully automatic, i.e. operating without human intervention.

The setup of the rooms used in the experiment is depicted in Figure 5a. The interaction takes place in the experiment room and the operator remains in the wizard room to deploy the overall system, control the cameras and operate the robot only when required. The TV screen in the wizard room provides visual access to the experiment room. Two cameras are used to record the sessions: one front camera facing the user, and a back camera to record a broader view of the room. The robot is place on top of a table to avoid placing it too low from the the child’s view. A kinect is used to collect data for future work.

5.3. Procedure

Each participant was invited to attend up to three sessions (one-hour session per week at most) to interact with the robot. Upon arrival on the first day the goals of the experiment and the experimental setup were introduced to the children. They were told that they were there to help us to improve a robot that will support hospitalized children [13]. Parents filled demographic and medical background surveys, and signed consent forms for participation and use of video/sound recordings. Additional questionnaires regarding hobbies and attitudes towards technology surveys were completed by children.

In each session the child selected one game (among the three available in the ALIZ-E framework) to play with the robot for at most 30 minutes.
was free to decide when to finish it. During the first activity in the first session the robot introduced itself and asked the child’s name. Every time a new game was selected, the robot explained the rules for that game. At the end of the game, the robot thanked the child looking forward to meeting again soon.

Once the activity was finished, a survey regarding the interaction was completed. If time was still available, the children could choose to play another game, and again, complete the corresponding questionnaire. Finally, an end-of-session questionnaire was filled.

At the end of the whole the experiment, i.e. after the child participated in all agreed sessions, we gave the child a participation certificate and a small gift (a toy) to thank the time invested in the experiment.

5.4. Evaluation measures

The goal of this work is to study the interaction between the robot and the child based on the dance activity. To assess the interaction we are interested in looking at the level of engagement of the child, their interest and involvement in the task itself and the attention to the robot during the performance, as well as the impressions of the children after the sessions. In this work we refer to all these as the quality of the interaction.

In order to assess it we analyze two types of measures: observation scores, based on video analysis; and self-reported measures, based on the survey presented to the user after each session. Therefore we can rely not only on the subjective responses of the children, but also on objective measures from the recordings. To ease the interpretation of the results, we describe the measures in detail in next section along with the outcomes obtained.

6. Results

12 children played the dance activity (a snapshot of a child playing the game is depicted in Figure 5b). From these, 7 played dance one day, 4, two days and 1, all three days, summing up to 18 interaction sessions. Henceforward we refer
Table 2: Summary of dance sessions per participant and interaction day: dance session duration in minutes ($t$); number of exploration/performance rounds ($r$); number of dance steps performed ($s$); and reason for ending the session ($e$, where $T$ corresponds to timeout, $F$, to failure and $V$, to voluntary).

<table>
<thead>
<tr>
<th>Id</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$</td>
<td>$r$</td>
<td>$s$</td>
</tr>
<tr>
<td>032</td>
<td>15:04</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>020</td>
<td>10:37</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>024</td>
<td>11:43</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>026</td>
<td>07:50</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>028</td>
<td>09:09</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>022</td>
<td>12:33</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>025</td>
<td>08:02</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>027</td>
<td>14:06</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>030</td>
<td>09:07</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>035</td>
<td>15:05</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>037</td>
<td>13:06</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>038</td>
<td>19:30</td>
<td>3</td>
<td>18</td>
</tr>
</tbody>
</table>

to the sessions where the same participant took part of as ‘multiple sessions’. To distinguish between these sessions we use the letters ‘a’, ‘b’ and ‘c’ corresponding to the first, second or third interaction day respectively.

Sessions lasted for around 13 minutes in average, where the longest corresponds to 26 minutes, and the shortest to around 8 minutes (Table 2). We must remark that the dance activity was the child’s second game choice in the day sessions (except for child 032 who always picked it first). Thus, the available time slot to finish the activity was generally less than 30 minutes, having in mind that two additional questionnaires had to be completed before the one hour slot for each session was over.

A dance session is divided into three main stages: introduction, exploration and performance, and an additional farewell stage (Section 4.1). The exploration/performance stages can be repeated as many times as the child wishes before ending the session. Table 2 summarizes the number of exploration/performance rounds per session and the number of dance moves (steps) each child performed during the exploration stage. In 5 sessions children performed only one round of exploration/dance, in 11 sessions, they repeated twice, and in 2 sessions, they repeated up to three rounds attaining 14.3 dance moves per session in average ($\sigma = 5.52$). Considering that the minimum number of steps per session is 5 (unless the session is interrupted before the end of the first round), it is easy to see that the children were looking forward to repeating the exploration stage to try more steps when time was available.

Table 2 also describes the reasons for ending the sessions. It is interesting to notice that only 3 out of 18 sessions were voluntary ended by the children (‘V’),


<table>
<thead>
<tr>
<th>category</th>
<th>classification</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gaze</td>
<td>around</td>
<td>looking at the sides</td>
</tr>
<tr>
<td></td>
<td>floor</td>
<td>looking at the floor</td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>different from looking at the robot (ceiling, door, camera, self)</td>
</tr>
<tr>
<td>face</td>
<td>amused</td>
<td>smiles, laughter, excitement</td>
</tr>
<tr>
<td></td>
<td>bored</td>
<td>yawns, upset</td>
</tr>
<tr>
<td></td>
<td>mocking</td>
<td>making funny faces at the robot</td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>puzzled, surprised, thinking</td>
</tr>
<tr>
<td>body</td>
<td>distracted</td>
<td>playing with the body, moving around, stretching, scratching, waiting, exaggerating movements, self-arranging</td>
</tr>
<tr>
<td></td>
<td>proactive</td>
<td>dancing or moving by own initiative</td>
</tr>
<tr>
<td></td>
<td>communicative</td>
<td>body gestures conveying a message</td>
</tr>
</tbody>
</table>

Table 3: Summary of features observed in the video coding.

While the remaining ones were forced to end either due to time limitations (12 sessions, ‘T’) or system failure (3 sessions, ‘F’). This is a good indicator that most of the times the children would have extended the interaction if possible.

We next describe the results, both quantitative and qualitative, obtained in the experiments reported here. Henceforth we will use the term “it” as personal pronoun while referring to the participant to avoid referring to the child’s gender.

6.1. Video coding

Overall, 236 minutes (3h 56min) have been annotated by one coder using ELAN\(^9\). Three independent coders validated random segments extracted from all the videos (12% of the data which corresponds to 30 min) achieving 96% of inter-observer reliability with the main coder. The aim at coding the videos is to label the different behavioral cues produced by the children in the sessions (similar coding schemes have been adopted in [13, 24, 18]). Table 3 summarizes the features observed in the video coding. Moreover, we also annotated the speech occurrences classifying them as expected (answers to questions posed by the robot) or unexpected (children’s free speech acts).

We present the results making a distinction between children who participated in the first session (we also include here those in multiple sessions since in the first day they are all in the same condition) and those in multiple sessions.

Figure 6 (left column) shows the results of the classified observations during the first session of the experiments, while Table 4 summarizes the corresponding statistics. In general, children were engaged with the robot and motivated to finish the task in their first session. In average, they only looked away from the robot 5% of the time and expressed amusement 16.12% of the time with little

signs of boredom. They revealed proactive behaviors and tried to communicate with the robot both verbally and non-verbally (communicative acts). Such queues are significant since they correspond to situations where the child opts to show a certain behavior on its own initiative, suggesting its willingness to take part in the interaction. More precisely, proactive behavior was observed in 11 out of 12 sessions, non-verbal communicative behavior, in half of the sessions, and out of 121 utterances, 87 corresponded to unexpected verbal acts (Table 5).

Most of them also were distracted at some point during the interaction. Special attention has to be paid to child 027, who was distracted for almost 5 minutes. Looking at the video recordings we could observe that the child exposed a fidgeting behavior, constantly playing with its hands and fingers (15 out of 19 occurrences within the body category), suggesting that the child was engaged in the session (more than 15\% of the time it expresses amusement) despite the high frequency of distracted observations.

Moving to the results corresponding to multiple sessions (Figure 6, right column), we can observe that in subsequent sessions children looked away from the robot more often, increased their distracted behavior, and expressed less amusement and more boredom. These signs show that the novelty effect diminishes across sessions (suggesting a habituation behavior previously observed in [20, 19, 15, 29]), making it more challenging to keep the children engaged. On the other hand, two children addressed the robot by its name evidencing a gain of familiarity with the robot compared to first sessions, where no child referred to the robot by its name at all. In addition, most of them finished the sessions due to external reasons and not because they opted to and child 032 even came back for a third session.
Figure 6: Left column: time percentage observed in the different categories per child. Right column: distribution of time percentage observed in each category per participants taking part in multiple sessions. Bars limited with vertical dashed lines belong to the same participant, where the left most bar corresponds to the first session and the right most, to the last session. We refer the reader to the Web version of the paper to view the color version of the images.
It is worth pointing out that technical problems occurred during sessions 020b and 028b, which had significant negative impact in the interactions\textsuperscript{10}. However, child 028b played for 26 min, evidencing its interest in interacting with the robot nevertheless.

Child 026 was quite engaged with the robot itself in both sessions, though not as much in the task during the second session, where it frequently looked around the room (1 min), expressed boredom (40 sec), and behaved distractedly (3 min), especially during idle periods. On the other hand, the child was keen on interacting with the robot in other ways (freely addressing the robot up to 20 times in the second session). Furthermore, the child was sad to finish the session and caressed the robot wishing good night. Thus, while the novelty effect with respect to the task diminishes, the child’s willingness to communicate with the robot does not, suggesting high interest in interacting with it.

**Involvement**

We propose a measure to integrate the observed behaviors to assess the state of the interaction. We define the state of user $i$ at a given time $t$ as

$$state_i^t = [l_1, l_2, \ldots, l_n]$$

where $l_j$ corresponds to a label in any of the categories defined above (gaze, face, body, speech). Additionally, we introduced a category to indicate the intervals when the child is dancing with the robot while copying the movements. We define the involvement metric as a measure that indicates the level of responsiveness of the child at every time step based on the ratio of label types at that time. Thus, the involvement level of child $i$ at time $t$ corresponds to

$$involvement_i^t = \begin{cases} \frac{|pos_i^t|}{|pos_i^t| + |neg_i^t|} & \text{if } |pos_i^t| + |neg_i^t| > 0 \\ -1 & \text{otherwise} \end{cases}$$

where $pos$ and $neg$ correspond to a set of labels classified in the following way:

- $pos$: smile, laugh, excited, thinking, puzzled, surprised, communicative (verbally and non-verbally), proactive, spatial-orientational arrangement, dancing along with the robot.

- $neg$: bored, yawning, upset, mocking, playing, stretching, scratching, self-arrange, waiting, looking around (around, floor, other).

The measure is defined within the interval $[0, 1]$ if labels are available, where 0 corresponds to no involvement in the activity (rather behaving distractedly) and 1 corresponds to fully involved. If no labels are available (meaning that the only action the child is doing is observing the robot), the state of the child cannot be assessed and therefore -1 is assigned. These situations take place\textsuperscript{10}

\textsuperscript{10}We do not remove these participants from this study since their first session had no issues.
Figure 7 shows the average involvement per child/session. The involvement assessment for the first session is mostly over 0.75 (high involvement). However, as previously discussed, the involvement decreases across multiple sessions. Figure 8 depicts a map of the involvement measure per session (time is normalized to allow direct comparisons between sessions). This map allows us to identify the intervals of the sessions where children felt more/less involved, providing evidence for further improvement of the robot behavior and activity design.

We can observe that the highest involvement values correspond to the performance stage, where the child and the robot dance together along with music. No evaluation takes place in this period, which probably leads the child to feel more relaxed and free to enjoy the activity. On the contrary, the exploration stage suffers the drawback of limited computational resources on board the robot, which cause small but significant delays before proposing a new step to the child. During these short periods, the risk of being distracted while waiting for the robot’s next proposal increases. Thus the tendency to increase the frequency of purple areas (low involvement).

6.2. User’s self-assessment

It has been argued that using classic questionnaires (question-and-answer style) with children does not always provide reliable feedback [21]. Thus, alternative tools as the ones proposed in [27, 13] were designed to explore the feelings of the child during the interactions, how they perceived the robot, what they thought about the game played as well as their mood during the interaction.

Bond assessment

A pie diagram (Figure 9a) was used to evaluate the bond the child felt with the robot compared to relationships in its environment. Figure 9b illustrates the frequency of categories selected by the children. Most of them reported a
Figure 8: Evolution of involvement for each session. Black rectangles indicate the intervals where greetings (first rectangle, beginning of the session) and the performance stages (dancing with music) occur. Note that in most sessions multiple rectangles are drawn which correspond to different rounds of performance stages played in the activity. Colors correspond to low (purple), medium (light orange), high (orange) level of involvement. Yellow represents an unknown state, meaning that no labels were coded and the child is only observing the robot.
close bond with the robot, mainly perceiving the robot as a friend or sibling. They were encouraged to justify their response where we can relate the bond reported with the robot and their daily lives. For instance, “I participated in a project called ‘Orma’ and we dance with friends”, “My friends like to dance”, “It spoke to me as a brother”, “It taught me new things” (teacher relationship).

**Emotional assessment**

Visual Analogue Scale (VAS, a pictorial representation of a 1-5 Likert scale through which respondents identify their answers, feelings or opinions [21]. Figure 10 depicts an example.) tools were used to evaluate the following features: satisfaction, amusement, anxiety and game enjoyment.

Figure 11 summarizes the self-assessment for these four dimensions per child and per interaction day. In general most children felt very happy (top vertex of the nets), very amused (left vertex), found the sessions very entertained and relaxed while interacting with the robot.

Child 020 slightly dropped from feeling very happy to quite happy in its second session and as expected, due to the technical problems in this session, it
Figure 11: Net representation for the feelings reported by the children. Each net corresponds to one participant describing the evolution at each day’s session (first day in red, second day in blue, and third day in yellow). The vertices correspond to categories ranked from 1 to 5 (where 0 corresponds to the center of the diagram and 5 to the vertices) in the following way: satisfaction (S), from very sad to very happy; amusement (A), from very bored to very amused; anxiety (X), from very relaxed to very anxious; and enjoyment (J), from very boring to very fun.
Figure 12: Answers reported by children after each session (first session in red color, second session in yellow and third session in cyan). (a) Leadership, where the answers ranged from ‘it played’ (the robot played with me), ‘it led’ (we played together but it led) and ‘same’ (we played together at the same level). (b) Expectations, where values correspond to ‘below’ my expectations, ‘in line’ with my expectations or ‘over’ my expectations.

reported being quite bored and finding the game boring, correlating the observations previously reported. Only one child (035) reported feeling quite agitated, but since it only took part in one session we could not assess whether it would feel more relaxed during a second encounter with the robot.

**Interaction assessment**

Two features were assessed to evaluate the perception of the robot behavior: leadership, to evaluate the equality of roles in the task; and expectancy, whether the sessions were over, in line or below the child’s expectations.

In general, children felt playing together either at the same level or having the robot leading the interaction (Figure 12a). Only one child felt that the robot was playing with it. Regarding expectations (Figure 12b), in most cases the first interaction with the dance robot went beyond the children’s expectations, for both single or multiple encounters. In the latter case, children either maintained or lowered their expectancy, but keeping it in line with their expectations at least, proving consistency in the robot behavior across sessions. Only child 020 reported it being it under its expectations as a consequence of the system failure.

6.3. Discussion

The current study is valuable to identify and verify the usefulness of the different strategies considered in this work. Moreover, the evaluation on repeated encounters provided insights on additional essential capabilities that should be included straightaway to sustain long-term interaction (where long-term refers to repeated encounters in a period of time).

The emergence of proactive behaviors, both verbal and non-verbal, evidence the fact that children were willing to interact with the robot beyond performing
the task. Children naturally talked to the robot, though no answer was given (except when the robot had posed a question). Moreover, they would start to move and dance while waiting for the robot to proceed, or even when the robot explained what to do next. Besides, they also reproduced movements that were not meant for copying (e.g. repositioning and robot postures). This suggests that providing creativity support (strat. 10 in Section 3), where children are able to contribute to the task and not only to follow instructions, would increase involvement and overcome boredom as observed in the second sessions.

The performance stage, where the child dances with the robot the learned sequence along with music, had the highest involvement levels, probably due to a combination of favorable factors: (i) active contribution to the task (strat. 6); (ii) peers’ role, as opposed to tutor-pupil role; and use of music, which proved to be a good motivational tool (strat. 8). Children reported liking and having fun with music “[I liked] the dance with music, [I] dance was funnier with music”, “We were able to dance together and we were synchronized (I also loved the music)”. We could also observe that quite often the children smiled when the music started. One child even suggested the robot a couple of times to play the music to continue dancing (revealing proactive behavior once again).

Based on the post-sessions questionnaires we verified that the simple engaging behaviors (Section 4.3) conveyed social behavior and were appreciated. Children were surprised and felt happy when the robot used their names: “I liked that it recognize me and it’s very kind”, “it said that my name is nice”. They liked talking to the robot, the courtesy and the dance motions, reinforcing the argument that body movement supports involvement [4]. When children were asked what they liked most, some answered: “to see him speaking”, “it answers to the greetings”, “to see it dancing and speaking”, “it danced really good”. Some children reported on the robot’s friendliness and envisaged a close relationship with it: “[I liked that] it makes me laugh, but also that it seems human”, “The robot’s friendliness”, “I like to play with him because we could have fun together at the park. I had a lot of fun with him and we became friends!”. Moreover, expectations were in line across sessions, which is relevant to avoid decreasing the child’s motivations to interact with the robot in the future occasions.

Limitations on the feedback about the robot’s decisions were detected. A child’s response to what it liked least was “[I] when I asked for new steps it also showed old steps” when the robot proposed a past failed motion to try it again. This shows that the robot failed in transmitting the need of going through failed past motions when necessary to improve the child’s capabilities through time. On the contrary, providing the right feedback at the right time may result in a good tactic to overcome frustration as suggested in [8]: “[I liked that] it apologized when it said something wrong”.

The case studies on repeated encounters suggest that the task around the interaction should avoid being too repetitive. Children enjoy a minimum degree of novelty. Otherwise their motivation decreases. In our work they appreciated the novelty on the dance movements across sessions. However, the task structure remained the same. We believe that slight modifications of the task framework would sustain the children’s motivation without modifying the goal of task.
Finally, despite dance being the child’s second choice of game to play (probably due to shyness to move the body) it turned out to be appreciated and well accepted among participants. In 9 occasions (out of 18) they reported that what they liked most in that day was playing dance, while the next preferred game was Quiz (4 times), a more static game.

7. Conclusion and Future Work

We have addressed the use of a robot dance tutor around a dance activity in single and multiple interactions, when possible. We have first took part in observation sessions where children performed dance sessions in schools to extract instruction methods and engagement strategies to guide our work. Next, we designed and developed a full robotic system, where the vision and speech components are the only wizarded ones. The decision making and the overall behavior management of the robot is completely autonomous. Experiments in the wild were performed to evaluate the engaging capabilities of the system. Observational analysis through video recordings and survey evaluations were carried out to assess the quality of interaction. Moreover, we presented an involvement measure based on the combination of behavioral cues expressed by children to assess the level of interest of the child in the sessions through time. In general we can conclude that children were quite engaged with the robot and the task. However, the involvement decreased along multiple sessions. Although a decrease of the novelty effect had an impact on such outcome, external factors also played an important role (mainly system failure).

We believe that the combination of simple engaging behaviors introduced in this work along with the strategies from the observation sessions effectively supported the interaction in single encounters. Children responded in natural ways to the robot, and more importantly, they showed interest in contributing to the task. Thus, to promote their spontaneity, and as a consequence their motivation in the task over multiple sessions, we intend to move from the current guided approach to a creative approach so that children can introduce their own ideas. The aim is to also promote learning through social interaction. Dance can be used as an instrument to convey concepts that children work in their everyday school activities (some theme examples observed at the schools were the states of water, creation of volcanoes, painting styles). Given that our work is part of the ALIZ-E project, where the targeted children suffer from diabetes or obesity, a candidate knowledge area under investigation is healthy habits. The amount of learned knowledge will serve as an additional measure in the search for suitable interaction evaluation metrics. Moreover, we intend to further evaluate the long-term aspect in more depth considering the observations in our study and the challenges pointed out in [2, 5].

Acknowledgments

We acknowledge Marco Nalin for organizing the experiments at San Raffaele Hospital. This work was funded by the EU FP7 ALIZ-E project (no. 248116).


