Daily HRI Evaluation at a Classroom Environment
– Reports from Dance Interaction Experiments –

Fumihide Tanaka
Sony Intelligence Dynamics Laboratories, Inc.
Shinagawa-ku, Tokyo, 141-0022
boom@idl.sony.co.jp

Javier R. Movellan
Institute for Neural Computation
University of California, San Diego
La Jolla, CA 92093-0523
movellan@mplab.ucsd.edu

Bret Fortenberry
Institute for Neural Computation
University of California, San Diego
La Jolla, CA 92093-0523
bret@fortenberry.com

Kazuki Aisaka
Sony Corporation
Shinagawa-ku, Tokyo, 141-0001
aisaka@sm.sony.co.jp

ABSTRACT

The design and development of social robots that interact and assist people in daily life requires moving into unconstrained daily-life environments. This presents unexplored methodological challenges to robotic researchers. Is it possible, for example, to perform useful experiments in the uncontrolled conditions of everyday life environments? How long do these studies need to be to provide reliable results? What evaluations methods can be used?

In this paper we present preliminary results on a study designed to evaluate an algorithm for social robots in relatively uncontrolled, daily life conditions. The study was conducted as part of the RUBI project, whose goal is to design and develop social robots by immersion in the environment in which the robots are supposed to operate. First we found that in spite of the relative chaotic conditions and lack of control existing in the daily activities of a child-care center, it is possible to perform experiments in a relatively short period of time and with reliable results. We found that continuous audience response methods borrowed from marketing research provided good inter-observer reliabilities, in the order of 70%, and temporal resolution (the cut-off frequency is in the order of 1 cycle per minute) at low cost (evaluation is performed continuously in real time). We also experimented with objective behavioral descriptions, like tracking children's movement across a room. These approaches complemented each other and provided a useful picture of the temporal dynamics of the child-robot interaction, allowing us to gather baseline data for evaluating future systems. Finally, we also touch the ongoing study of behavior analysis through 3 months long-term child-robot interaction.

HRI'06, March 2–4, 2006, Salt Lake City, Utah, USA.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems; I.2.9 [**Artificial Intelligence**]: Robotics; J.4 [**Social and Behavioral Sciences**]: *Psychology, Sociology*; K.3.1 [**Computers and Education**]: Computer Uses in Education; K.4.2 [**Computers and Society**]: Social Issues

General Terms

Experimentation, Human Factors, Measurement, Performance

Keywords

QRIO, child development, child education, child robot interaction, children, daily HRI evaluation, engaging interaction, human robot interaction, long-term interaction, social interaction

1. INTRODUCTION

More than 10 years have passed since sociable robots became a focus point in robotic research [1, 3], but still very little is known about their real-world capabilities, i.e., very few studies have been conducted in daily non-laboratory conditions for sustained periods of time. Pioneering work was conducted by Kanda [6] where a humanoid robot joined an elementary school society for two weeks, and by Kozima [7], which is monitoring the development of communication between a small interactive robot and autistic children. While long-term field studies are invaluable to advance the field, very little is known about methods appropriate to learn as quickly as possible from conditions that are dramatically less controlled than in typical laboratory studies. The challenge for social robotics is that the controlled conditions typical of laboratory environments do not generalize well to the everyday life environments where these robots need to operate. In fact, research programs that rely on constrained laboratory conditions tend to develop theories and focus on problems that are of little help to understand the organization of behavior in real life environments.

In this paper, we present our first attempts to evaluate HRI (Human-Robot Interaction) in a daily environment for prolonged periods of time. The works are part of the RUBI project [9] whose goal is to design social robots in a continuous manner by immersion in educational environments. As part of this project we introduced a small humanoid robot, QRIO into the Early Childhood Education Center (ECEC) at the University of California, San Diego for a total of 45 daily sessions, of approximately 45-60 minutes each. One of our targets was to investigate the dynamics of the interactions developed between children and QRIO, to develop methods for evaluating this interaction and for testing the success or failure of algorithms designed to improve this interaction.

Room-1 hosts around 12 children between 10-24 months old. After each child's 2-year old birthday, they move to one another room for older children. At that time a new child joins in the class, keeping the number of children at maximum of 12. We targeted the age group of Room-1 because it forced us to focus on non-verbal, affective forms of communication. While verbal communication plays an important role in human interaction, current technologies such as speech recognition and natural language processing are not sufficiently advanced to sustain engaging interaction for significant periods of time. On the other hand, children even younger than 2 years old are capable of establishing engaging interactions for long periods of time despite their very limited verbal abilities. One of our goals was to gain a better understanding of the principles and important factors in these early forms of communication and to attempt to replicate them in social robots.

In this paper we report the result of a study designed to evaluate two different robot dancing algorithms: (1) An open-loop motion generation algorithm in which QRIO executes a choreographed (canned) dance sequence played and danced in a canned sequence without any responsiveness to the external world, and (2) A closed-loop algorithms in which QRIO responded to perceived motion such as the movement produced by a dancing partner. We also report the latest data from the ongoing study of long-term behavior analysis through 3 months children-QRIO interaction.

The structure of the paper is as follows: In the next section, we introduce the RUBI project and a humanoid robot platform, QRIO. Section 3 describes the Early Childhood Education Center (ECEC) where the project is being conducted. We then present the design of the dance interaction experiment, followed by an analysis of the results. At last, we glance over a future research direction with the latest data from the ongoing study, and we conclude this paper.

2. THE RUBI PROJECT AND QRIO

The research project presented in this paper is part of the RUBI project at the University of California, San Diego (UCSD) [9]. The goal of the project is to explore the use of interactive robot technologies in educational environments. To this effects two robot platforms, RUBI and QRIO, are being tested on a daily bases for prolonged periods of time.

Figure 1: Sony Entertainment Robot, QRIO (QRIO is a test prototype.)

Figure 2: Early Childhood Education Center

QRIO is a small humanoid robot developed by Sony (Figure 1) [8, 5, 4]. It is a stand-alone autonomous robot with three CPUs: the first is used for audio recognition and text-to-speech synthesis; the second is for visual recognition, short and long term memory, and behavior control architecture; and the third CPU is used for motion control. In addition to the onboard CPUs, remote PCs can be utilized as remote-brains by using QRIO's embedded wireless-LAN system. In our laboratory in Tokyo (Sony Intelligence Dynamics Laboratories, Inc.), a PC-cluster system called IDEA that has 352 CPUs (Opteron 248) providing more than 1 Tflops is utilized for real-time intelligence dynamics. After years of research and developmental effort, QRIO's ability encompasses a very wide range: walking, running, jumping, playing soccer, throwing a ball, swinging a putter, singing songs, recognizing humans by vision and audio, making a conversation (dialogue), learning, imitating human motions, etc [4, 8, 11].

One of QRIO's most striking skills involves motion generation such as dancing. QRIO is endowed with various choreographed dance sequences, and is also capable of mimicking the motion of its human partner in real-time [11].

3. EARLY CHILDHOOD EDUCATION CENTER (ECEC)

The RUBI project, which started in October 2004, is being conducted at the Early Childhood Education Center at the University of California, San Diego (Figure 2). We decided that it was important for the researchers involved in the project to immerse themselves in the environment the robots were going to operate and thus we spent 3 months

Figure 3: At the introduction period of QRIO to Room-1 in ECEC: QRIO was treated like a newcomer to the classroom.

Figure 4: (Left) Canned dance: QRIO is dancing in the playback mode where there is no interactivity with the outside. (Right) Interactive dance: QRIO is mimicking the external motion dynamics.

volunteering 10 hours a week at ECEC. This allowed us to establish personal relationships with the teachers, parents and children and helped us identity the challenges and the situations for robots could be helpful in a classroom environment. After the volunteering period, QRIO was introduced to Room-1 on March 2005 (See Figure 3). The work presented here is based on the first 45 days of interaction between QRIO and the children.

During the first three months of volunteering and observation at Room-1, we were shocked by the power of music and dance on the children's behavior. Children are engaged by music and dancing for a long periods of time and teachers also utilize music skillfully in their daily activities to control the room's atmosphere (e.g., playing slow, relaxing music to prepare the children for sleep time).

We thus decided to introduce QRIO in Room-1 during the time of the day that teachers typically reserved for the children to dance, sing, and move actively around the room. In this paper, we present our first experiment in which we evaluated dance interaction between children and QRIO using two different dance programs which we refer to as *canned dance* and *interactive dance*.

4. DANCE INTERACTION EXPERIMENTS BETWEEN CHILDREN AND QRIO

4.1 Canned Dance

QRIO can dance in an open-loop (playback) mode (Figure 4, Left). Typically this function is used for stage demonstration purposes. QRIO has a large archive of dance data ranging from simple gestures to more complex choreographed dances which are programmed by using a 3-D motion editing system. From this set, we chose a dance program called "Johnny". For about one minute QRIO dances its preprogrammed movement while playing music of the Johnny B. Goode song from a speaker embedded in its mouth.

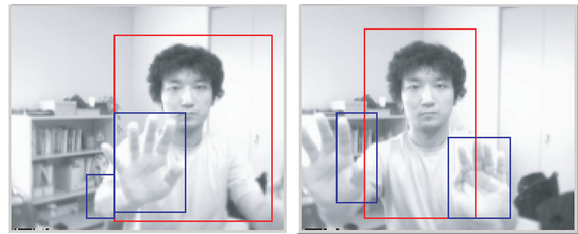

Figure 5: A set of moving-regions obtained by QRIO's stereo cameras. They are clustered by disparity (distance) information.

4.2 Interactive Dance

QRIO can also dance in a closed-loop (interactive) manner, responding to outside motions (Figure 4, Right). The current approach to interactive dancing relies on motion imitation using stereo vision cameras. QRIO has FPGAs-embedded vision system that allows it to obtain frame difference and disparity (distance) images in real-time. By exploiting them, a set of moving-regions clustered by each region's disparity information can be calculated in real-time using standard image segmentation techniques (Figure 5). QRIO can detect its outside partner's rough shape and the motion dynamics they generate. The first author of this paper developed an interactive dance program where QRIO responsively imitates its dance partner's movement by controlling all of its upper-body's joint angle information so that its silhouette fits the moving-regions set in each time cycle [11].

The operating conditions in Room-1 were very challenging: The illumination was very difficult for computer vision programs to operate, and the conditions were very unconstrained, with children moving continuously about the room in an unpredictable manner. Although in our interactive dance algorithm the accuracy in the imitation of the user's shape is not the best when compared with other systems, our approach is robust and proved relatively responsive in the very difficult, but realistic, conditions of Room-1.

Due to safety concerns, an operator controlled the locomotion part of QRIO during this operational mode, and thus the overall system was semi-autonomous, where only the robot's upper-body keeps moving autonomously while reacting to outside motions. We programmed QRIO to produce the same music as in the open loop dance condition.

4.3 Experimental Plan

We had two goals in mind: (1) To observe and evaluate the effects of the two dance algorithms, and (2) To experiment with different methods for evaluating and leaning about the interaction developed between the children and QRIO.

During the experiment, each day QRIO played either the canned dance or the interactive dance (in alternately each day for a total of 3 days for each dance condition). By that

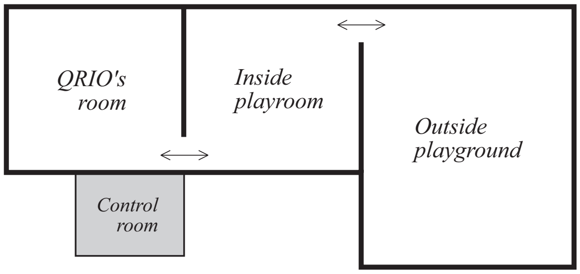

Figure 6: A plane figure of the Room-1. Children can freely move between the two inside rooms and an outside playground. An operator tele-operates QRIO from a control room which is separated by a one-way mirror from QRIO's room.

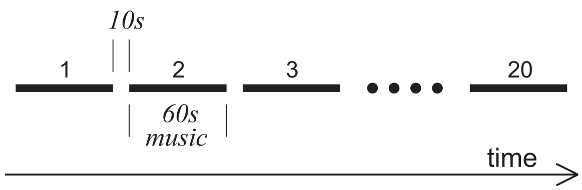

Figure 7: Each day a music song was repeated 20 times with 10 second rest interval, and we tested either of the three conditions: QRIO in a canned dance mode, QRIO in an interactive dance mode, or no QRIO (only music).

time, QRIO had been spent 30 days over 2 months with children in Room-1, and they were already familiar with QRIO. Room-1, consists of two indoor rooms and one outdoors playground (Figure 6). There is also a room with a one-way mirror designed to observe the children's behavior. All the experimental sessions were conducted with a robot operator, which stayed inside the observation room, an experimenter that stayed in the same room as QRIO and was in charge of insuring safety, a teacher that was in the room as long as one of more children were present, and a third person for videotaping the sessions. The operator, which was not visible by the children, kept track of time and sent dance command to QRIO appropriately. The other three adults were instructed to avoid interaction with the children as much as possible. After it became clear that they were not responsive, the children learned to ignore them. All sessions were videotaped using two camcorders; one fixed at the ceiling to provide a wide angle view, and the other one hand-held by one of the experimenters.

During the experimental sessions, QRIO stayed in the room adjacent to the observation chamber. At all times children could move freely between the room in which QRIO was dancing, the other indoor room, or the outdoor playground. QRIO stayed in the same room at all times playing the music and dancing even all of the children had left.

Every day the same song was repeated 20 times with a 10 second mute interval (Figure 7). After the first 6 days we decided to add a third experimental condition where the same song was played but without QRIO being present. The music was played with the same timing as in the other conditions from a pair of compact speakers set at the same height position as QRIO inside the room. The reason why

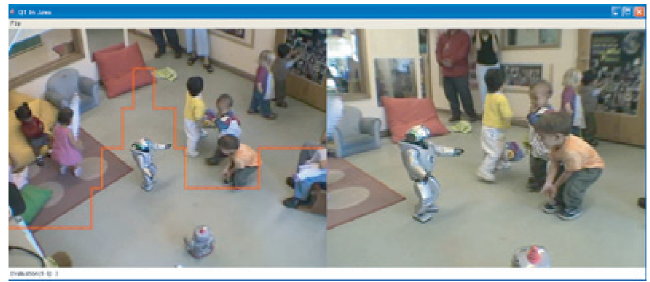

Figure 8: Judges used the keyboard to continuously assign labels to the observed videos representing the goodness of interaction between children and QRIO. Judges can simultaneously see two synchronized movies taken by two cameras are combined into one. The judge can also see the recent history of his/her evaluation which is superimposed on the movie as a red graph.

Instruction to labelers

Estimate the probability that a randomly selected adult that sees the video would agree with the following statement :

"Currently there are examples of good child-robot interactions."

1: Less than 20% of the people would think there are good interactions
2: Between 20% and 40%
3: Between 40% and 60%
4: Between 60% and 80%
5: More than 80%

Figure 9: Coding instruction given to judges.

we introduced the third condition was that our 3 months observation (Section 3) suggested that the power of music was so strong that it was unclear whether QRIO had an effect of attracting children compared with the music itself.

5. EVALUATING THE DAILY DANCE INTERACTION

5.1 Subjective (Qualitative) Method

Videos were coded independently by 5 paid undergraduate students from UCSD. The students were not informed about the goals of the study. After trying various methods, we found continuous audience response methods [2, 10] borrowed from marketing research to be reliable and efficient. Judges were asked to evaluate the goodness of the interaction between QRIO and the children on a 1-5 scale continuously as they viewed the video. Overlaid on the video the judges could see a curve displaying their recent evaluation history (See Figure 8). The judges evaluated the 9 video sessions independently and in random order.

This continuous evaluation method provided good interobserver reliability even though the evaluation instructions were quite vague. Figure 9 shows the exact instructions given to the judges. Figure 10 shows the inter-observer reliability, averaged across all possible pairs of judges, as a

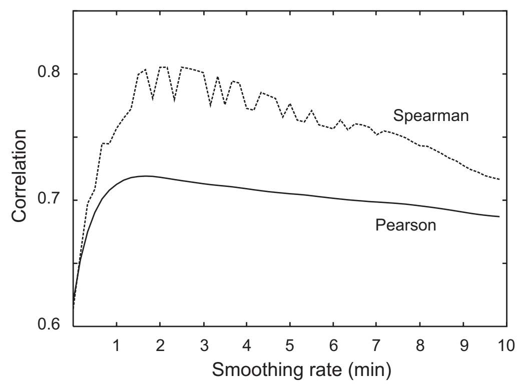

Figure 10: Transition of inter-observer reliability calculated by two correlation functions by changing the window size of data-smoothing procedure.

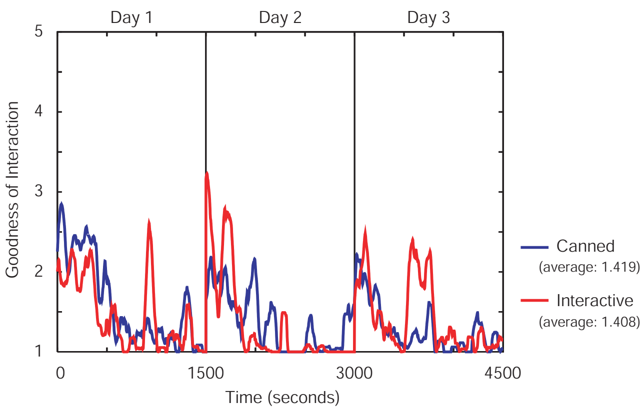

Figure 11: Subjective evaluation of dance interaction experiments: Each graph plots the average data of 5 judges' evaluation using continuous audience response methods.

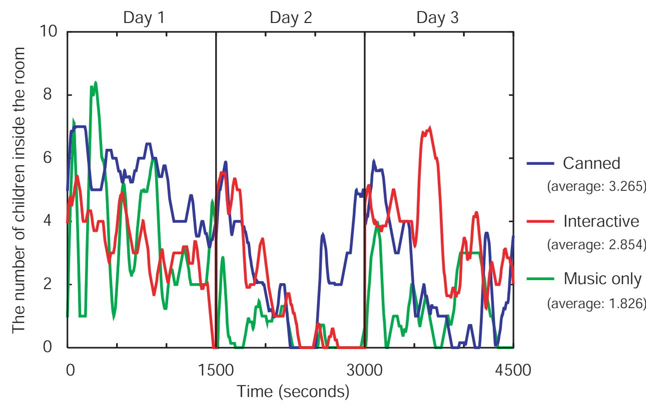

Figure 12: Objective evaluation of dance interaction experiments: Counting the number of children inside the room through time.

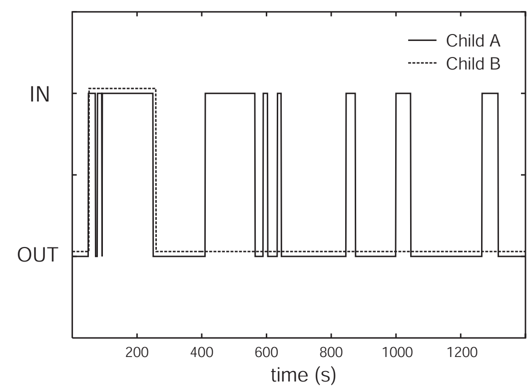

Figure 13: Tracking two children's room-in/out history. Child-A frequently comes into the room whereas Child-B never comes back after leaving the room.

function of the amount of temporal smoothing in the scores provided by the judges. As expected from an additive signal plus noise model, the inter-observer reliability shows an inverted U-curve. Initially, as the smoothing window is increased the inter-observer reliability increases due to the fact that the high frequency noise in the observer's signals is being filtered out. However if the smoothing window is increased, it starts filtering out the actual evaluation signal from the observers, not just the noise, thus decreasing the inter-observer reliability. The graph peaks with temporal smoothing of about 1.5 minutes, suggesting that judges were implicitly averaging about 1 to 2 minutes of the past interaction when making their continuous evaluation.

Figure 11 shows the evaluation curves as a function of time. Each line displays the temporally smoothed evaluation averaged across the 5 judges. The graphs show consistent decays in the goodness of interaction reflecting the fact that the children lost interest on QRIO's dance as time progressed.

5.2 Objective (Quantitative) Methods

We also experimented with objective behavioral descriptions. In particular we counted the number of children in the experimental room throughout time. As described in Section 4.3, children can freely go into or out of the room whenever they want. If there are more attractive things in the next room, they can leave the room. Therefore it is expected that by tracking the number of children we can measure the dynamics of popularity.

Figure 12 shows the results. This time the third condition where only the music played (without QRIO) at the same timing was also plotted. The average values show a small difference between the canned dance condition and the interactive one, but again, the latter presents a more fluctuating graph shape than the former. The music-only condition lags behind the others, which means a dancing QRIO has factors that attract children's interests.

Figure 13 shows the history of two specific children's moving in and out of the room. We can see that one child frequently moves in and out whereas the other does not. If QRIO is attractive, it is expected that children remain inside the room, therefore the time when the graph indicates the value, IN (the child is inside the room) becomes longer. Another important feature which the graph can offer to us is the frequency of each child's coming into (out of) the room. If a child frequently comes into the room even though the c : canned dance condition   i : interactive dance condition   m : music condition

	mean			std			P (df=10)		
	c	i	m	c	i	m	c-i	c-m	i-m
In/Out number	3.00	4.06	2.85	1.86	1.76	0.90	0.167	0.800	0.024
Total stay per day (sec.)	653	560	330	421	339	142	0.468	0.029	0.027
Average single stay time	374	146	116	438	93.8	46.6	0.124	0.085	0.328
Average time to come back	173	176	210	136	136	220	0.950	0.641	0.696

Figure 14: Results statistics in multiple time-scales evaluation.

duration of each single stay is not so long, the child is also said to be attracted by QRIO. This example suggests that the evaluation of HRI needs to be done in multiple time-scales.

We analyzed 9 days videos from this motivation, and obtained results statistics shown in Figure 14. The first row shows the statistics regarding the number of times children came in/out the room where QRIO was, for each of the experimental conditions. The second row shows the statistics of total stay time (sec.) per child as well. The total stay is the sum total of small stays, and the third row focuses on the average time (sec.) of the small stays. Similarly, average times for each child's coming back to the room (after he/she left the room) are extracted, and their statistics are shown in the fourth row.

The results of the experiment were very informative. First, as we can see from the second row, there were statistically significant differences between the canned dance condition and the music condition (c-m), and the interactive dance condition and the music condition (i-m) ($P < 0.05$). One reason of the difference can be understood from the first row, which shows significant difference ($P < 0.05$) in the number of In/Out between the interactive dance condition and the music condition (i-m). Children came in the room more often in the interactive dance condition than the music condition, which led to the difference in the total stay time between them (i-m). It is interesting that regarding the difference between the canned dance condition and the music condition (c-m), there was almost no statistical difference in the number of In/Out. It is considered that the difference in the average single stay time which is shown in the third row caused the difference in the total stay time between them (c-m), in turn.

Secondly, the results also suggested that regarding the average time to come back, there was little statistical difference between all conditions. Although it is too early to say, the results showed an example of a time constant factor in child-robot interaction. If it exists, we can utilize the statistics into the design process of social robots. For example, if we find that average single stay time is also constant, (which is still ambiguous judging from the third row), the constant time factor indicates an appropriate standard time length for a single interaction process happens between a child and a robot. It is also very useful information when we develop appropriate contents of entertainment robots such as the dance and music of QRIO.

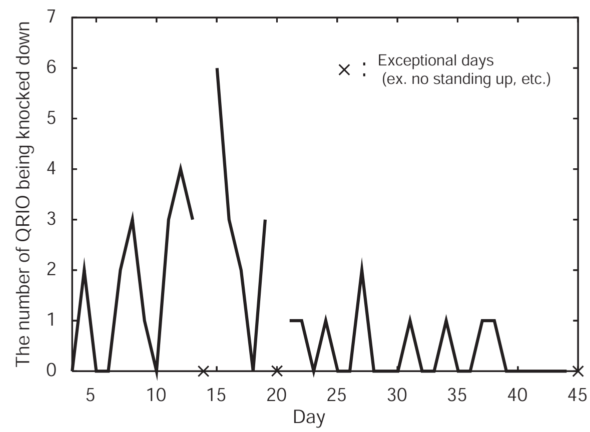

Figure 15: The number of QRIO falls down through 45 days.

6. FUTURE WORK

To date, QRIO has spent more than 45 days over 3 months in Room-1 at ECEC. Currently, we are analyzing the long-term interaction by focusing on children's behavior such as touching QRIO. By investigating the touch behavior in detail, we are expecting to obtain not only another way of evaluating children-QRIO interaction but also data from which we can infer children's attitude against QRIO, which will help the appropriate design process of social interactive robots.

Figure 15 shows an interesting preliminary result of that analysis. It plots the transition of the number of times QRIO falls down each day. One important characteristic of QRIO as a humanoid platform is that it is capable of protecting itself when it falls and then standing up by itself. So far most of 90% of the times that QRIO fell, were caused by the children, e.g., QRIO was knocked by children moving around.

At the beginning, the number was relatively small because children were cautious about QRIO as an unknown quantity. As they accustomed themselves to QRIO, the number of knockdowns increased, over a period of about 10 days. It is interesting that after that the number of knockdowns decreased. During the last week of the 2005 part of the study, he almost never fell down. This is in spite of the fact that by the end of the study the quality of the interaction improved, including behaviors such as hugging, and pulling QRIO's hands to go for a walk with them. Our hypothesis is that children's attitude toward QRIO is developing (e.g. they are getting to know that QRIO is very weak and easily falls down, and treat QRIO softly), not just they are getting to be bored with QRIO, and it appears especially in their touching behavior.

7. CONCLUSION

The design and development of social robots that interact and assist people in daily life requires for researchers to move outside constrained laboratory conditions and to evaluate their systems on realistic, daily life environments.

In this paper we presented preliminary results on a study designed to evaluate and learn about an algorithm for social robots. First we found that in spite of the relative chaotic conditions and lack of control existing in the daily activities of a child-care center, it is possible to perform experiments in a relatively short period of time and with informative results. In the study presented here we evaluated an interactive dance algorithm in a period of 9 days, with 45-60 minutes session each day. In the process we found that continuous audience response methods provide good inter-observer reliabilities, in the order of 70%, and temporal resolution, in the order of 1.5 minutes, at low cost. They also provided a useful picture of the temporal dynamics of the child-robot interaction, allowing us to gather baseline data, for example, about how long it takes for children to loose interest on QRIO, how often they come back into QRIO's room, and how long average stay with QRIO is. Finally, we also touched our ongoing study of behavior analysis between children and QRIO through 3 months long-term interaction.

8. ACKNOWLEDGMENTS

This study is ongoing at Room-1 of UCSD's Early Childhood Education Center. We thank the director of ECEC Kathryn Owen, the head teacher of Room-1 Lydia Morrison, and the parents and children of Room-1 for their support. We also thank Dr. Ronald C. Arkin for giving us comments on the study and proofreading the manuscript of this paper. The study is funded by UC Discovery Grant dig03-10158 and by a grant from the National Alliance for Autism Research (NAAR).

9. REFERENCES

[1] C. L. Breazeal. *Designing Sociable Robots*. The MIT Press, 2002.

[2] I. Fenwick and M. D. Rice. Reliability of Continuous Measurement Copy Testing Methods. *Journal of Advertising Research*, 31(1):23–29, 1991.

[3] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A Survey of Socially Interactive Robots. *Robotics and Autonomous Systems*, 42(3-4):143–166, 2003.

[4] M. Fujita, Y. Kuroki, T. Ishida, and T. T. Doi. SDR-4X II: A Small Humanoid as an Entertainer in Home Environment. In *Proceedings of the 11th International Symposium of Robotics Research*, 2003.

[5] T. Ishida, Y. Kuroki, and J. Yamaguchi. Development of Mechanical System for a Small Biped Entertainment Robot. In *Proceedings of the 2003 IEEE International Workshop on Robot and Human Interactive Communication*, pages 297–302, 2003.

[6] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro. Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. *Human-Computer Interaction*, 19(1-2):61–84, 2004.

[7] H. Kozima, C. Nakagawa, and Y. Yasuda. Interactive Robots for Communication-Care: A Case-Study in Autism Therapy. In *Proceedings of the 2005 IEEE International Workshop on Robot and Human Interactive Communication*, pages 341–346, 2005.

[8] Y. Kuroki, T. Fukushima, K. Nagasaka, T. Moridaira, T. T. Doi, and J. Yamaguchi. A Small Biped Entertainment Robot Exploring Human-Robot Interactive Applications. In *Proceedings of the 2003 IEEE International Workshop on Robot and Human Interactive Communication*, pages 303–308, 2003.

[9] J. R. Movellan, F. Tanaka, B. Fortenberry, and K. Aisaka. The RUBI/QRIO Project: Origins, Principles, and First Steps. In *Proceedings of 2005 4th IEEE International Conference on Development and Learning*, pages 80–86, 2005.

[10] M. Polsfuss and M. Hess. 'Liking' through Moment-To-Moment Evaluation: Identifying Key Selling Segments in Advertising. *Advances in Consumer Research*, 18(1):540–544, 1991.

[11] F. Tanaka and H. Suzuki. Dance Interaction with QRIO: A Case Study for Non-boring Interaction by Using an Entrainment Ensemble Model. In *Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication*, pages 419–424, 2004.