HRI as a Tool to Monitor Socio-Emotional Development in Early Childhood Education

Javier R. Movellan
Emotient.com
6440 Lusk Blvd,
San Diego, CA, 92121
javier@emotient.com

Mohsen Malmir
University of California San Diego
9500 Gilman Dr.
La Jolla, CA, 92093
mmalmir@ucsd.edu

Deborah Forester
University of California San Diego
9500 Gilman Dr.
La Jolla, CA, 92037
dforster@ucsd.edu

ABSTRACT

Sociable robots are benefiting from machine perception systems that automatically recognize social behavior (e.g., detect and recognize people, recognize their facial expressions and gestures). These systems can be used to support sophisticated forms of human-robot interaction. In addition the data provided by the perceptual systems can be data-mined to discover the socio-emotional structure of the environments where the robot operates. In this paper we analyze the data collected by a social robot, named RUBI-5, during a field study at an Early Childhood Education Center in which the robot autonomously interacted with 16 toddlers for a period of 28 days. RUBI-5 was equipped with face detection, person identification and automatic recognition of facial expressions of emotion. The data automatically collected by RUBI during the 28-day period revealed the children's preferences for different activities as well as each toddler's preferences to play with or to avoid playing with other specific children. The study illustrates that social robots may become a useful tool in early childhood education to discover socio-emotional patterns over time and to monitor their development. The data provided by the robots could be used by educators and clinicians to discover problems and provide appropriate interventions.

Keywords

Social robotics, face recognition, Sociogram, Facial expression recognition.

1. INTRODUCTION

Previous research shows that relatively simple sociable robots can generate rich forms of socio-emotional interaction with toddlers that are sustained for months [7]. In addition randomized pretest/posttest studies have shown that interaction with these robots can result in measurable gain in vocabulary skills [1]. Recent advances in machine perception are making possible the automatic recognition of emotion-relevant behavior in real time (e.g., detect and recognize faces and facial expressions of emotion). These new systems can be used to support sophisticated forms of HRI. In addition the sensory data used by the robot can be stored and data-mined. In this paper we analyze the data collected by a social robot, named RUBI-5, during a 28 day long field study at the UCSD Early Childhood Education Center. RUBI-5 was equipped with 3 cameras connected to computer vision systems to detect people, recognize them and analyze their facial expressions. The results of the analysis show that the data collected by social robots can indeed be very useful to discover socio-emotional patterns and to monitor their development over time.

The study is part of the RUBI project, which started in 2004 with the goal of studying the potential value of social robot technologies in early childhood education environments [2,3,4]. Figure 1 shows the different robot prototypes used in the project, starting with QRIO and ending with RUBI-5. The diagram organizes the different prototypes by level of mechanical complexity, degree of robot autonomy, and quality of the observed human-robot interactions. The latest prototype so far is RUBI-5, the prototype we used in the field study described here. In this study RUBI-5 functioned autonomously for 28 consecutive days with 16 toddlers in real life conditions. By the time she broke our previous record, RUBI-5 was in full need of repair: a physically broken arm, and several burnt servomotors. Historically we use the pronoun "she" to refer to the RUBI robots, the pronoun most children use to refer to them During the 28 days of operation RUBI-5 collected a wealth of sensory data. Previously, we showed how the data collected by RUBI-5 could be used to predict kids preference over different activities using facial expression recognition [13]. Here, we delve into more details about socio-emotional analysis of the environment in which RUBI-5 operated.

Robots have previously been used in classrooms for educational purposes. In a 2-week field study Movellan et. al. [1] a social robot was used to teach kids English and Finnish vocabulary. It was shown that the kids that had the most persistent interaction time with the robot learned most. In another study, Kanda et. al. [14] used a humanoid robot to interact with elementary school students and teach them English. They proposed if the students have some background and familiarity with English language, education with robot might be more fruitful. Robots have also been used in schools to monitor the social structure of the environment. In a study, Kanda used Robovie to monitor the social structure of an elementary school and discover the pattern of friendship between students [15]. Tanaka and Movellan [16] analyzed behavior of toddlers interacting with the QRIO robot and found evidence for forms of social behavior towards the robot that lasted for long periods of time.

Not surprisingly, emotion plays a critical role in the interaction between toddlers and robots. Having something akin to emotional states that the children could understand was critical for surviving the rigors of interacting with toddlers. From the early versions of

RUBI [2, 3], we also pioneered the development and testing of expression recognition technology in daily life environments, including smile detection [4, 12] and the ability to analyze and detect infants crying from sound [5, 6]. This pioneering work was influential on the development of the sophisticated facial expression recognition software, FACET 1.1 that we used in RUBI-5.

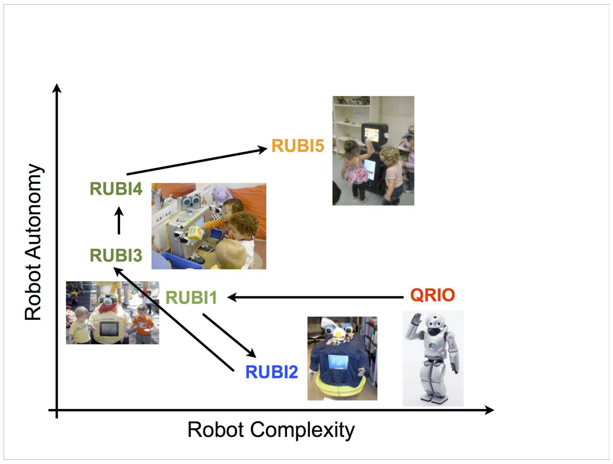

Figure 1. Prototypes used in the RUBI project organized in terms of their complexity (Y axis), degree of autonomy (Y axis) and quality of the HRI (red for high, blue for low).

The paper is organized as follows. In Section 2, we briefly describe the RUBI-5 architecture, including face recognition and facial expression recognition. In Section 3, we describe the field study that is the focus of this document. Section 4 describes the main results of the study and is followed by a discussion section.

2. The RUBI-5 Prototype

2.1 Hardware

RUBI-5, the latest prototype of the RUBI series is shown in figure 2. This was the first prototype developed using modern digital fabrication methods [8]. Each of Rubi's arms has 4 DOF, independently controlled using Robotis' Dynamixel servos EX106+, RX64 and RX28. Each hand has an IR sensor inside the gripper that is used as object proximity sensor. The head has 3 DOF, a webcam for image capture and an iPad2 for the animated face. RUBI's "belly" has a touchscreen tablet PC, which is used to display educational games and popular songs. A MacMini server with 2 GHz Intel Core i7 processor and 8 GB of RAM runs the Robot Operating System (ROS), the machine perception engines (face detection, person recognition, and expression recognition), activity scheduling, and motor control algorithms.

2.2 Software

ROS: RUBI-5's software architecture is based on the Robot Operating System (ROS). The entire system is distributed and works by passing ROS messages between ROS nodes that provide a variety of services. A node called RUBIScheduler is a finite state machine that schedules the activity to play with the children. This is based on the previous activities, the amount of time since the children touched RUBI's belly and the constraints of the ECEC classroom's daily schedule.

Figure 2: RUBI-5.

Games: RUBI performs four types of activities: (1) Sings songs ("Wheels on the Bus", and "Monkeys on the Bed") while playing animations on her belly's tablet, and dancing. (2) Educational Games targeting vocabulary development. For example, in one game 4 images are presented on the screen and RUBI asked to touch one of them (e.g., where is the orange?). These games combine sounds and visuals presented on RUBI's belly, as well as physical actions, like clapping, looking towards the screen when the child touches it, and smiling. (3) "Give and Take" games: children give objects to RUBI. She takes the object, looks at it and gives it back to the child saying "Thank You". (4) Idle. If RUBI's belly is not touched for a period of 10 seconds the RUBIScheduler puts her in "Idle Mode", in which she makes randomly scheduled idle movements, and displays simple visuals on her belly. When children touch RUBI's belly the scheduler chooses a new activity, provided it is consistent with the classroom's daily schedule.

Person Recognition: RUBI-5 captures the ongoing scene using three cameras, located in the head, right and center of the belly's tablet and under the belly. Currently these cameras are used to identify who is playing with RUBI and what facial expressions they are making. We use the following face recognition pipeline: Each image is fed to OpenCV [9] face detector to find the faces in the image. The detected faces are normalized to the same size and converted to 4 layer Gaussian image pyramids with a between layer downscale of 1.2. Daisy features [10] are extracted from overlapping image patches from each layer of this pyramid. PCA is then used to reduce the dimensionality of the features. A Multinomial Logistic Regression classifier is used to recognize the different participants (see figure 3).

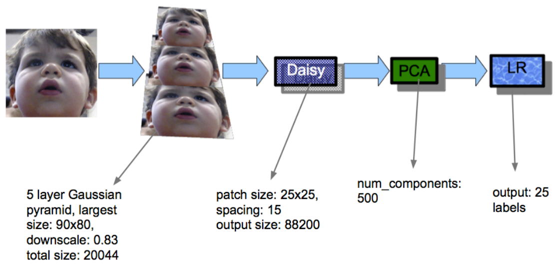

Figure 3. Face recognition pipeline in RUBI-5.

The classifier was trained using 7000 images of 28 subject collected by RUBI. Of these 28 subjects, 16 were toddlers and 12 were adults, including classroom teachers and researchers accompanying RUBI. We divided the entire dataset into two non-overlapping sets: train and test. The test set size was 35% of the entire dataset. After training using the training set we tested the system on the test set using the following procedure: For each image on the test set the system had to choose amongst 28 possible alternatives (16 toddlers, plus 12 adults). The system guessed the correct alternative with 93% accuracy. Figure 4 shows the confusion matrix for our dataset.

Facial Expression Recognition: Facial expression recognition was performed using the FACET R1.1 SDK from Emotient.com.

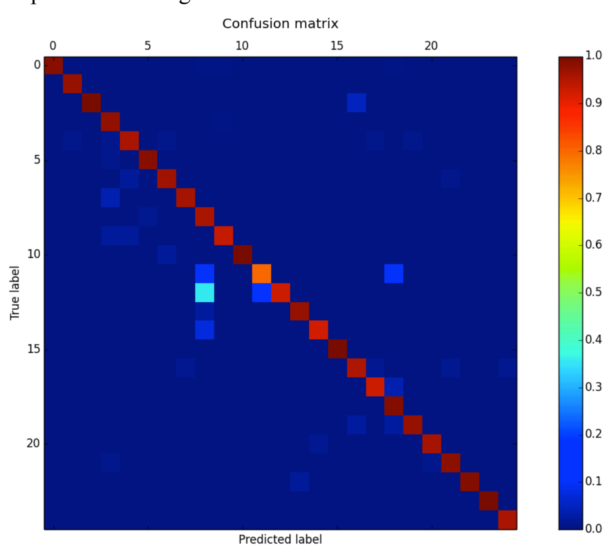

Figure 4. Confusion matrix for face recognition on ECEC faces dataset.

FACET is the commercial version of CERT [11], one of the most popular and accurate facial expression recognition systems. FACET R1.1 recognizes 6 primary expressions of emotion: anger, disgust, fear, joy, sadness and surprise. Since FACET R1.1 was trained to recognize adult facial expressions from faces that deviate no more than 15 degrees from frontal, it was not clear whether the system would prove useful to recognize toddler facial expressions.

3. STUDY DESIGN

Participants: 16 toddlers (ages 11 to 23 months) from room 1B of UCSD's Early Childhood Education Center (ECEC) enrolled during the period of Jan 24 to September 11, 2013. The total number of children at any given time ranged from 9-12. Two teachers informally observed the interaction between the children and RUBI. A research assistant under the supervision of an ethnographer took notes to characterize the observed interactions between children and robot using standard ethnographic methods.

Procedure: RUBI was left alone in Room 1B of ECEC, starting on Jan 24, for increasingly longer periods of field-testing. On Aug 12 we brought RUBI to ECEC with the intention of continuing the study until she stopped operating. This happened on September 11, 2013. During this period RUBI was relatively stationary, making only small rotational movements with the drivetrain, thus allowing to obtain power using a standard electrical outlet. RUBI-5 ran on two types of schedules: continuously while research staff were on location; and an automated schedule designed to coincide with activity periods chosen by the educational staff as curriculum appropriate. During every session, RUBI-5 was on idle state until someone touched her belly. At this time, she chose either a game or a song. The songs always end after a specific pre-determined time, while the games continue until no one touches the belly for 10 seconds. After game or song has finished, RUBI goes back to the idle state, showing the idle game on the belly and looking around while moving slightly her arms and head. This cycle continues until the session finishes.

Data: During each session, RUBI kept the log of the games and songs that were played. She also recorded images from the three RUBI cameras. The head and belly-mounted cameras captured an image when they detect a face and a game is being played. The tablet camera captured images every 2 seconds during the game episodes. These pictures were then processed to extract the identity using the face recognition described in section 2.2. Facial expressions were also extracted using the FACET SDK.

4. RESULTS

4.1 PREDICTING ACTIVITY PREFERENCES

We asked the 2 classroom teachers and a research assistant that observed the children on a daily basis to rank how much the children liked the 10 different activities they played with RUBI: 7 educational games, 2 songs, 1 give and take game. The average Pearson correlation between the three human judges was 0.68. Then we computed the correlation between the output of the different emotion channels (obtained using FACET) and the activity rankings averaged across the 3 human judges. The independent variable was the total number of images greater than 0.95 on the corresponding emotion channel (e.g. FACET was at least 95% confident about the target emotion). Amongst all the facial expression channels we found one statistically significant correlation ($r=0.73$, $p< 0.05$, 2-tails): the Joy channel. The average agreement between the Joy channel and each of the human judges was 0.73, which was a bit larger than the average agreement between the 3 judges (0.68). Thus the facial expressions of Joy, automatically detected by the robot, provide reasonable estimates of how much the children like the different activities. Figures 5 shows rows of 8 faces corresponding to the highest values of joy for 3 toddlers and faces corresponding to the lowest values of joy for the same toddlers. The Figure shows that, while not perfect, on average the images that FACET chosen as being more joyful, do indeed look more joyful than those chosen as being less joyful.

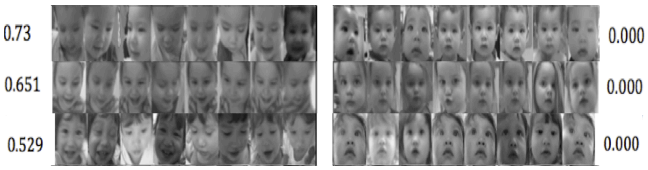

Figure 5. Examples of toddlers faces with maximum and minimum Joy value. The left 8 columns are the faces with top Joy value, while the right 8 columns are the faces with least Joy value. The average of Joy channel for corresponding set of faces is written next to them.

We then retrieved the top pictures that were used in predicting activity preference (Figure 6). We found that some of these faces were actually from some of the adults in the classroom. Basically RUBI detected the general mood of the classroom, including the response of adults, while RUBI was engaging the children on different activities.

Figure 6. Faces with top Joy values. Each row indicates one class according to the face recognition system. Some of the misclassified samples are adults (usually parents of toddlers) that were not in our face training dataset.

4.2 Detailed Temporal Analysis

During the 28 days of field study RUBI played the same activities multiple times. We synchronized the outputs of the Joy detector channel for each activity and averaged it across the 3 cameras and the different times the activity was played. In all the activities except for one, the result was that Joy was approximately constant across the activity. However for the "Wheels on the Bus" song the function had clear peaks and valleys (See Figure 7). The local peak in the Joy channel appeared at the beginning of the song, indicating that they were happy RUBI was playing this song, and at the points in the songs where RUBI said "all to the town".

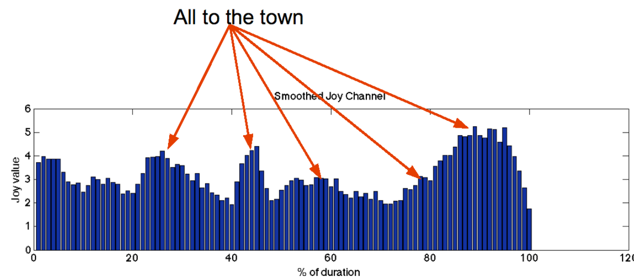

Figure 7. Average of Joy channel for different trials of *"The Wheels on the Bus"* song. A local peak is observed at the start of the verse *"all to the town"*.

4.3 RUBIGrams

We also investigated whether the data collected by RUBI could reveal some aspects of the social structure in the classroom. To this end we collected the frequencies with which RUBI detected two children together during each specific game and song trial. The results are presented in Figure 8, using a Sociogram-like display, which we called RUBIGram: The width of the lines in Figure 8 represents the relative amount of time each pair of children was seen together. The graph shows that some children play much more with RUBI than others, and that some pairs of children are seen together much more than others. However two children may be seen together often for two reasons: (1) they may like playing together with RUBI. (2) They may be playing independently and, just by chance, those children that play more with RUBI are more likely to be seen together. We then compensated for the effects of chance as follows: for each edge between x,y, denote the strength of edge with P(x,y). This is the amount of time x,y were seen together. Denote by P(x) the total amount of time x spent with RUBI. Then we are interested in the quantity P(x,y)-P(x)P(y), which is 0 if the times x,y spent with RUBI are independent of each other. Figure 9 shows the edges corresponding to P(x,y)-P(x)P(y). Positive values are shown by red, while negative values are in blue. Thus positive values indicate that two children are seen together more than it is expected from chance. Blue lines indicate that two children are seen together less than would be expected from chance (i.e., they tend to avoid each other).

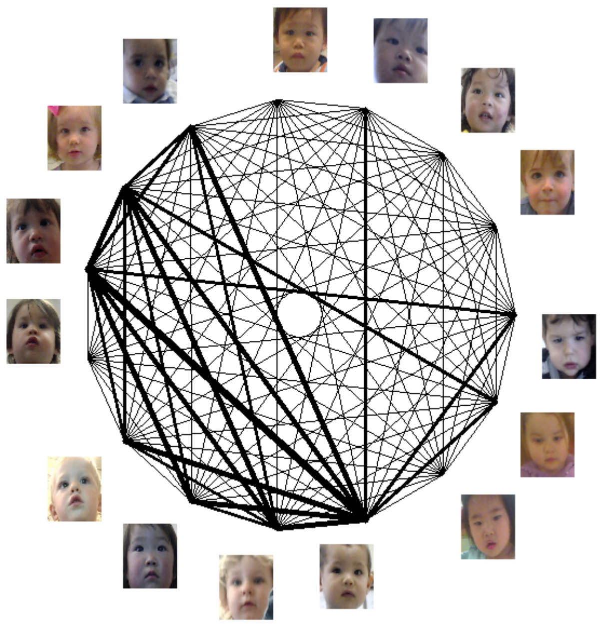

Figure 8. RUBIGram. Each link between two toddlers indicate the amount of time they spent together playing with RUBI. The width of the line represents the time.

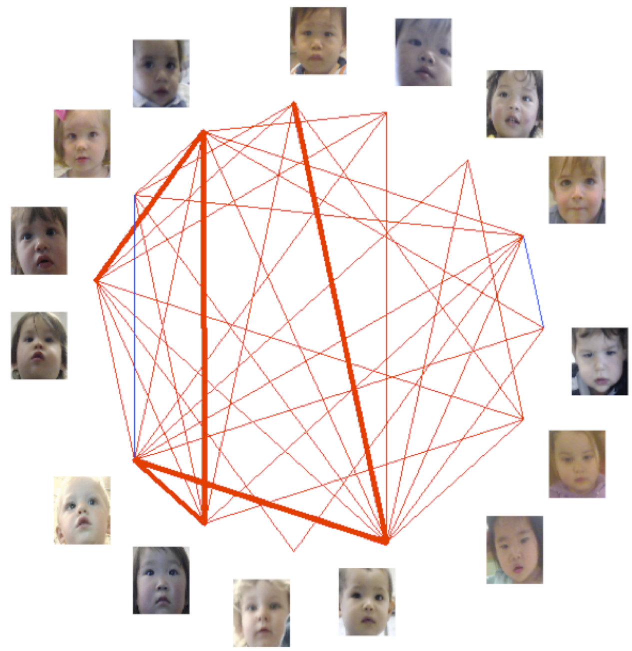

Figure 9. Chance corrected RUBIGram. Red edges indicate pairwise associations larger than expected from pure chance, while blue edges indicate avoidance.

5. Discussion

Advances in machine perception technologies are providing social robots with perceptual primitives that can support sophisticated forms of HRI. Because of the active, real time experience that sociable robots can provide, they are ideal tools to harvest and datamine behavioral data from daily-life environments. Here we showed some analysis of data harvested by a social robot, RUBI-5, that interacted with 16 toddlers for a period of 28 days. In particular we focused on the analysis of the facial expressions the children made while engaging on different activities with the robot, and on the analysis of which toddlers the robot saw playing together. We found that automatic expression recognition (in particular joy detection) was an effective metric for detecting activity preferences. Using expression recognition RUBI achieved a 0.73 average correlation with the preference rankings provided by human observers. This is slightly larger than the human inter-observer correlation (0.68) for preference rankings. RUBI could also provide precise temporal information about which parts of an activity the children liked most. In addition RUBI discovered which children preferred to play alone, play with other specific children, or avoided specific children. The study illustrates that social robots could become a useful tool in early childhood education to discover socio-emotional patterns over time and to monitor their development. The data harvested by these robots could be mined to develop norms for typical socio-emotional development and to help on early detection of developmental disorders.

6. ACKNOWLEDGMENTS

The research presented here was funded by NSF IIS 0968573 SoCS, IIS INT2-Large 0808767, and NSF SBE-0542013.

7. REFERENCES

[1] J. R. Movellan, M. Eckhardt, M. Virnes, & A. Rodriguez. Sociable robot improves toddler vocabulary skills. In Proceedings of the 4th ACM/IEEE international conference on Human robot interaction:307-308, 2009.

[2] B. Fortenberry, J. Chenu & J. R. Movellan. Rubi: A robotic platform for real-time social interaction. In Proceedings of the International Conference on Development and Learning (ICDL04), The Salk Institute, San Diego, 2004.

[3] J. R. Movellan, F. Tanaka, B. Fortenberry & K. Aisaka. The RUBI/QRIO project: origins, principles, and first steps. In Development and Learning, Proceedings. The 4th International Conference on:80-86, 2005.

[4] J. R. Movellan, F. Tanaka, I. R. Fasel, C.Taylor, P. Ruvolo & M. Eckhardt. The RUBI project: a progress report. In Proceedings of the ACM/IEEE international conference on Human-robot interaction: 333-339, 2007.

[5] P. Ruvolo & J. R. Movellan. Automatic cry detection in early childhood education settings. In Development and Learning, ICDL, 2008.

[6] P. Ruvolo, J. Whitehill, M. Virnes & J. R. Movellan. Building a more effective teaching robot using apprenticeship learning. In Development and Learning, 7th IEEE International Conference on: 209-214, 2008.

[7] F. Tanaka, A. Cicourel & J. R. Movellan. Socialization between toddlers and robots at an early childhood education center. Proceedings of the National Academy of Sciences, 104(46):17954-17958, 2007.

[8] D. Johnson, M. Malmir, D. Forster, M. Alac & J. R. Movellan. Design and early evaluation of the RUBI-5 sociable robots. In Development and Learning and Epigenetic Robotics (ICDL), IEEE International Conference on:1-2, 2012.

[9] G. Bradski. PROGRAMMER'S TOOLCHEST-THE OPENCV LIBRARY-OpenCV is an open-source, computer-vision library for extracting and processing meaningful data from images. Dr Dobb's Journal-Software Tools for the Professional Programmer 25.11: 120-126, 2000.

[10] E. Tola, V. Lepetit & P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32(5): 815-830, 2010.

[11] G. Littlewort, J. Whitehill, T. Wu, M. Frank, J. Movellan, and M. Bartlett. The computer expression recognition toolbox (cert). In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 2011.

[12] Whitehill, M. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan. Towards practical smile detection. Pattern Analysis and Machine Intelligence, (11):2106–2111, 2009.

[13] M. Malmir, D. Forster, K. Youngstrom, L. Morrison, J. Movellan. Home Alone: Social Robots for Digital Ethnography of Toddler Behavior. ICCV Workshop on Decoding Subtle Cues from Social Interactions, Sydney, Australia, 2013.

[14] Kanda, Takayuki, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. "Interactive robots as social partners and peer tutors for children: A field trial." Human-computer interaction 19, no. 1 (2004): 61-84.

[15] Kanda, Takayuki, Rumi Sato, Naoki Saiwaki, and Hiroshi Ishiguro. "A two-month field trial in an elementary school for long-term human–robot interaction." Robotics, IEEE Transactions on 23, no. 5 (2007): 962-971.

[16] Tanaka, Fumihide, and Javier R. Movellan. "Behavior Analysis of Children's Touch on a Small Humanoid Robot: Long-term Observation at a Daily Classroom over Three Months." In Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on, pp. 753-756. IEEE, 2006.