Home Alone: Social Robots for Digital Ethnography of Toddler Behavior

Mohsen Malmir[†] Deborah Forster[†] Kendall Youngstrom[†] Lydia Morrison[†] Javier R. Movellan[‡]

[†] University of California San Diego, 9500 Gilman Dr, La Jolla, CA, 92093
[‡] Emotient, 6440 Lusk Blvd, San Diego, CA, 92121

Abstract

We describe the results of a field study in which the social robot RUBI-5, was left alone for a 28 day period to interact autonomously with 16 toddlers at an Early Childhood Education Center. The study is part of the RUBI project, which started in 2004 with the goal of exploring the potential of social robotics for research and enrichment of early childhood education environments. As part of the 28 day field study RUBI-5 collected data about the facial expressions, activities, and spatio-temporal proximity of the toddlers. We found that RUBI-5 could use the facial expression data to accurately predict the children's preference for different activities: on average robot agreed with human judges as much (Pearson Correlation =0.67) as human judges agreed with each other (Pearson Correlation = 0.68). In addition RUBI discovered some useful aspects of the social structure of the toddler's group. The study is an important milestone in social robotics, both for the length of time the robot could interact autonomously with children, and for the richness of the data that it provided. The results indicate that social robots have the potential to act as low cost, autonomous "digital ethnographers" in a manner that may revolutionize the science and technology of early childhood education.

1. Introduction

An unprecedented number of children in the US start public school with major deficits in basic academic skills [1]. Scientific evidence shows that children who have early failure experiences in school are those who are most likely to become inattentive, disruptive, or withdrawn later on. Empirical research using longitudinal randomized control studies is now showing that early childhood education programs can effectively prevent academic deficits [1]. However, due to their high costs, such programs may not find widespread use. Thus it is critical to find innovative ways to gather and analyze data on early childhood education so as to better understand toddler's behavior and to conduct rapid, big-data experiments. As part of this overall vision, the RUBI project started back in 2004 with the goal of studying the potential of social robot technologies in early childhood education. Since then 5 different robot prototypes have been developed and immersed on an early childhood education center for sustained periods of time (see Figure 1). The early prototypes (RUBI-1, 2) were remotely operated by humans. RUBI-3 was a transitional design. RUBI-4 was the first prototype to operate autonomously for a period of 15 days. RUBI-4 provided useful data about toddler behavior. In particular it was shown that a 2-week period of interaction with RUBI-4 resulted in improvement of vocabulary skills in 18-24 month olds [2]. However many of the results found with RUBI-4 required for human ethnographers to analyze hundreds of hours of video, a process that was both slow and costly. The latest prototype (RUBI-5) (Figure 1, bottom row) was designed to operate as an autonomous "digital ethnographer" that would embed itself on the daily routine of the toddlers life and enrich their environment while gathering and analyzing the observed behaviors.

After a construction period that lasted several years, including hundreds of hours of short-term field studies, we felt that RUBI-5 was ready for a long-term study. One of our goals was to break the previous record of 15 days of autonomous operation, achieved by RUBI-4. More importantly we wanted to explore whether RUBI-5 could autonomously data-mine the toddler's behavior and provide insights about what activities the children like most. To this end RUBI-5 was equipped with off-the-shelf computer vision tools to recognize the children she interacted with and to analyze their facial behavior. Here we show that these tools, while still not perfect, provided very useful information, including the preference of children for different activities, as well as sociograms indicating which children tend to interact with RUBI and one another on a daily basis.

The rest of the paper is organized as follows. In Section

978-1-4799-3022-7/13 $31.00 © 2013 IEEE
DOI 10.1109/ICCVW.2013.104

Figure 1. Top row: RUBI progression: from left to right RUBI-1, RUBI-2, RUBI-3 & RUBI-4. RUBI-1 and 2 were remotely operated. RUBI-3 was the first fully autonomous design. Bottom row: RUBI-5 playing with kids at ECEC classroom 2A.

2, we describe the design and modification process for RUBI-5 along with the software architecture we used to make this study feasible. In Section 3, we describe the specification of the study in terms of experiments setup, number of toddlers involved and the data collection and processing procedures. Section 4 describes the results for the experiment and is followed by a conclusions section.

2. Robot Design

For more than 8 years the RUBI robot series have been built and improved using the experiences achieved from thousands of hours of field studies [3-9]. During these studies, robot prototypes were immersed in an uncontrolled environment to interact with toddlers at the Early Childhood Education Center (ECEC) at the University of California, San Diego. User-friendly appearance, ability to operate autonomously, low cost design and implementation, and safety of interaction with toddlers were among the early design criteria that emerged from these field studies. As a better understanding was gained of the sort of interactions that emerge between toddlers and robots, the emphasis switched from Wizard-of-Oz studies in which RUBI was tele-operated by a human, to sustained, full autonomy. RUBI-4 was the first robot of the series that operated autonomously for 15 consecutive days. To achieve this end a relatively simple design was used (e.g., single degree of freedom head, no facial expressions, 2 degree of freedom arms). While RUBI-4 provided very useful data, confirming that children that interact with the robot show significant increases in vocabulary skills, the analysis of the obtained data was very time consuming and expensive. The main bottleneck was the human coding of 14 days * 8 hours per day * 3 cameras. This coding ended up taking more than a year of time. Thus the design of RUBI-5, the robot prototype presented in this document, focused on two aspects: (1) increase the complexity of the robot so as to support more complex forms of interactions than RUBI-4 could do, and (2) use machine perception software to automatically analyze the sensory data. The goal was for RUBI-5 to become an autonomous "digital ethnographer" that could be left alone with a group of toddlers to analyze their behavior while enriching their educational environment. In the following two subsections we describe RUBI-5 and go into the details of hardware components and software architecture.

2.1. Hardware

Figure 2 shows RUBI-5, the latest prototype of the RUBI series. This was the first prototype developed using modern digital fabrication methods. The robot is 24x20x37 inches. Her drive train consists of two Shayang Ye IG-52GM motors with magnetic encoders, controlled by Roboteq's MDC2250 Motor Controller board. Each of Rubi's arms has 4 DOF, independently controlled using

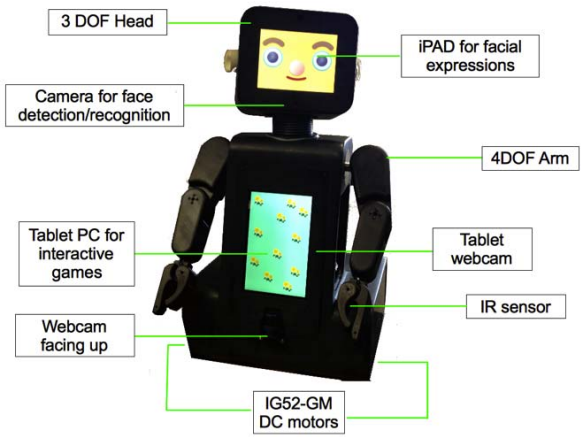

Figure 2: RUBI-5 appearance and hardware list.

Robotis' Dynamixel servos EX106+, RX64 and RX28. Each hand has an IR sensor inside the gripper that can detect an object. The head has 3 DOF, a webcam for image capture and an iPad2 for the 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, expression recognition), activity scheduling, and motor control algorithms.

Because the toddlers can stand in different distances to RUBI-5 while playing the games or dancing with the songs, we installed 3 cameras, each capturing part of the robot's surrounding area: One in the head, another in the torso and the last one under the tablet. Figure 3 shows shots of the same scene from 3 different cameras.

2.2. Software

ROS: 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.

Games: There were two types of material presented on the touch-screen: (1) RUBI sings popular songs ("Wheels on the Bus", and "Monkeys on the Bed") while she physically dances and shows an animation in her belly's touchscreen. (2) Educational Games. These are Flash-based 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. In addition RUBI can play "Give and Take" games (see Figure 1 bottom row). In this game children give objects to RUBI. She takes the object, looks at it and gives it back to the child saying "Thank You".

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.

Facial Expression Production: RUBI's facial expressions are controlled by a ROS node, called RUBIFace, running on an iPad. RUBIFace animates facial expressions using 23 different parameters. The set of facial expressions used at ECEC field studies were selected using a survey conducted on 30 different adults. These people we asked to give scores to a wide range of facial expressions based on multiple questions. We performed MDS on the survey's outputs to represent the different expressions using a 2 dimensional space, which can be roughly interpreted as "valence" and "energy". Figure 4 shows the results for 40 different faces. We picked the top 10 images with the highest positive valence.

Person Recognition: RUBI-5 captures the ongoing scene using three cameras, located in the head, right and center of the belly's tablet and bellow the belly. Currently these cameras are used to identify who is playing with RUBI and what facial expressions they are making. The face recognition system was developed by us, using the following pipeline. After an image is captured, we use the OpenCV [10] face detector to find the faces in the image. Each detected face is normalized to a specific size, it is converted to a 4 layer Gaussian image pyramid with a between layer downscale of 1.2. Daisy features [11] 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. 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. Because adults were also present in the field while RUBI-5 was interacting with kids, we added them to our dataset to reduce the number of false positives. Figure 5 shows some examples of this dataset. 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.

Facial Expression Recognition: Facial expression recognition was performed using the FACET R1.1 SDK from Emotient.com. FACET is the commercial version of CERT [12], 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.

Figure 3. Scene shots from RUBI-5's three different webcams. The left image is the picture from camera under the tablet. The middle picture is the tablet webcam's shot and the right one is captured by the webcam inside the head.

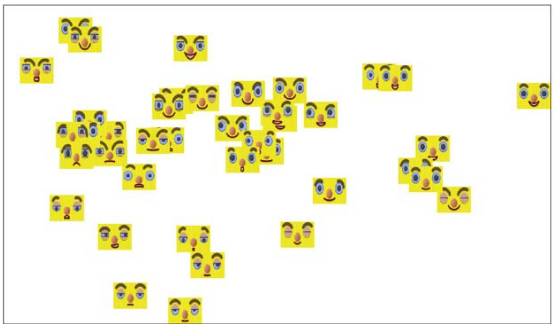

Figure 4: Visualization of the results of RUBI-5 online facial survey. People were asked to give their opinions about each face by grading them in different aspect such as positive-negative, active-passive and like-dislike.

3. Study Design

Participants: 16 toddlers (ages 11 to 23 months) from room 2A 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 2A 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.

Figure 5: example faces from our dataset. Each row represents different views of the same child. The images were collected over a course of 8 months at ECEC.

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

Expression Recognition: One of our goals was to test whether RUBI could automatically detect which kind of activities the children liked most. To this end we asked 2 teachers and the research assistant to rank from last (1) to first (10) how much the children liked the 10 activities they did with RUBI: "Give and Take", "Bus Song",

Figure 6. Faces corresponding to largest and smallest values of the joy channel. Faces in the left side have the top 8 values of joy for the corresponding toddler. Faces in the right side have the lowest value of joy for the same subjects. . The average of joy values for the faces in each row are written on the sides.

"Monkey Song", "Animals Game", "Balloons Game", "Fruits Game", "Transportation Game", "Photos Game", "Triangles Game", "Objects Game". The average Pearson correlation between the three human judges was 0.68. We then computed the correlation between the output of the different emotion recognition channels 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. This means that FACET was at least 95% sure that the face exhibited the expression of the target emotion. There was a statistically significant correlation ($r=0.73$, $p< 0.05$, 2-tails) between the output of the Joy channel and the average human rankings of the different games (see Table 1). The average agreement between the Joy channel and each of the human judges was 0.67, which was very close to the average agreement between the 3 judges (0.68). We were surprised that some of the channels associated with negative emotions (e.g., Fear, Disgust) showed relatively large, though not significant, correlation with game preferences. One explanation provided by the teachers is that some children get upset when some of their favorite games are about to end.

Figures 6 shows rows of 8 faces corresponding to the highest values of joy for 10 toddlers and faces corresponding to the lowest values of joy for the same toddlers. The numbers written on each side of the figure are the averages of joy value calculated from the faces in the corresponding row. 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.

Table 1: Correlation between expert observer ranking of preferred game and FACET output channels.

Channel	Pearson Correlation	Significance: 2-tailed t-test (9 df)
Anger	0.179	$p > 0.05$
Disgust	0.534	$p > 0.05$
Sadness	0.011	$p > 0.05$
Surprise	0.463	$p > 0.05$
Fear	0.534	$p > 0.05$
Joy	0.733*	$p < 0.05$

RUBIGrams: Next we map the frequency of dyadic activities between the different children. The goal was for RUBI to automatically generate a RUBI-centric sociogram [13,14], which here we call a RUBIGram. To this end, RUBI partitioned the time she was interacting with children into intervals defined by the start and end time of a game or a song. For each interval, RUBI identified the toddlers interacting with her during that

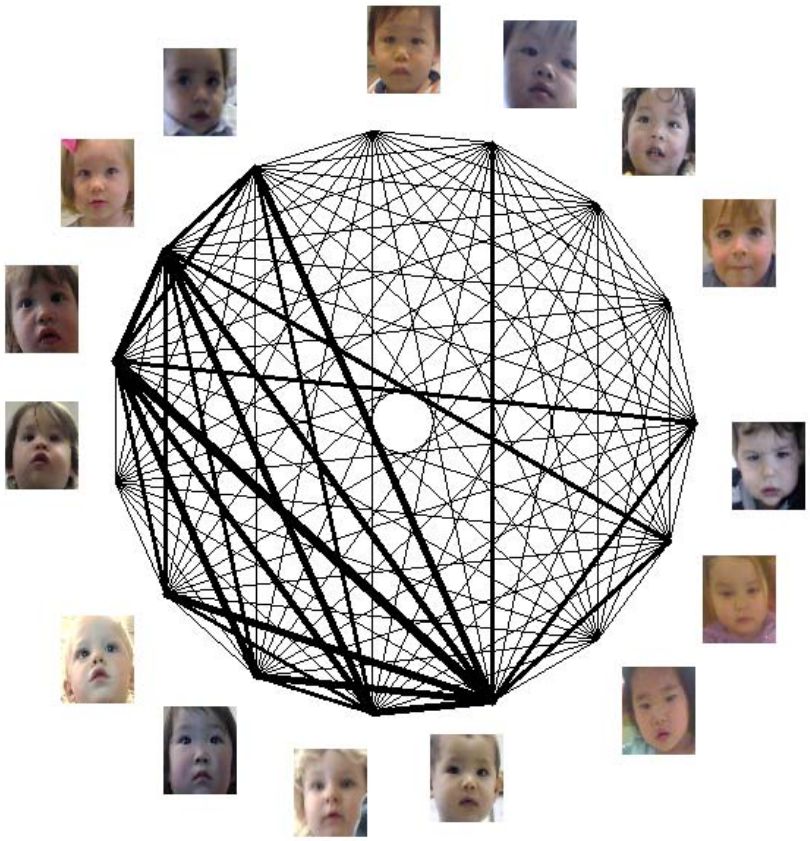

Figure 7. RUBIGram representing RUBI's observed dyadic interactions between the different children. Thickness of lines proportional to the density of time any two children were found in RUBI's view within the same session.

activity. For each pair of children RUBI computed the number of activities in which the two children were seen together. A "RUBIGram" was then constructed as follows: Children that are seen together often were connected with wide lines, and children that are seldom seen together were connected with thin lines. Figure 7 shows the resulting RUBIGram. The graph shows clear patterns in the playing styles of the different children. Some children play with RUBI while other children are present. Other children tend to play with RUBI alone.

5. Discussion

We presented results of a field study in which a social robot (RUBI-5) was left alone at UCSD's Early Childhood Education Center for a period of 28 days in a classroom of 16 toddlers. During this time RUBI interacted with the children in a fully autonomous manner playing "Give and Take" games, singing, dancing, and playing educational games that have been previously shown to improve children's vocabulary skills [2]. More importantly during this time RUBI collected data from 3 cameras, two touchscreens and, proximity sensors. Off-the-shelf computer vision tools allowed RUBI to recognize who she was playing with, what games they were playing, and what facial expressions they were making. Using this information RUBI could detect which activities the children enjoyed most, with as much accuracy as experienced teachers and ethnographic observers could do. In addition RUBI discovered some aspects of the social structure of the group of toddlers, including cliques of children that tend to play together with RUBI.

While the results are preliminary they show that low cost social robots like RUBI could be used as autonomous digital ethnographers, to run experiments, and datamine the children's social, affective, and cognitive behaviors. Based on the experience with the previous robot prototype, RUBI-4, it would have taken thousands of hours to manually code and analyze the data from the 3 video cameras. Instead RUBI-5 could automatically analyze the data in real time, at little or no cost. The results of the data analyses, e.g., sociograms, preferred games, typical facial expressions, scores on the different games, could be provided to teachers on a periodic base, to keep track of the children's social, affective and cognitive state. More importantly a network of such robots may open opportunities to run high volume, low cost experiments, to better understand, monitor, and improve the learning and development in our early childhood

education centers.

Acknowledgments

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

References

[1] J. Sparling, S. L. Ramey, C. T. Ramey, and R. G. Lanzi. The transition to school: Building upon preschool foundations and preparing for lifelong learning. In E. Zigler and S. Styfco, editors, The Head Start Debates. Yale University Press, Connecticut, 2004.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] S. Wasserman and K. Faust. Social Network Analysis: Methods and Applications. Cambridge University Press, 1994.

[14] L.C. Freeman. Visualizing social networks. Journal of Social Structure 1:1, 2000.