We propose a 1 year continuation of the UC Discovery Project Dig03-10158 on applications of personal Robots to learning and education. The goal of this project is to explore the idea of robots that teach children skills and assess their development while interacting with them in an affective and human-like manner. To pursue this idea for the past 3 years we have been developing robot systems designed to interact with children in a social manner. In year one we immersed researchers and two prototype robots, named RUBI and QRIO, in the daily life of Room 1 at the Early Childhood education center at UCSD. In Year 2 we analyzed the data collected during Year 1 and developed hardware and software tools that incorporated the lessons learned so far. In Year 3 we focused on dataset collection for development of robust perceptual primitives, and on development of robots that operate fully autonomously in the classroom environment for longs periods of time. For the 4th and last year of this project we propose the following activities: (1) Collection of datasets for development of critical perceptual primitives. (3) Refinement of perceptual primitives, including expression recognition, person identification and mood detection. (4) Development and deployment of 3 robots that operate autonomously in the classroom for months at a time. (4) Development of tools for robots to summarize and communicate the information they collect in a useful and compelling manner to teachers and parents. (5) Development of tutoring engines for robots that adapt to the mood in the classroom and the characteristics of individual students. (6) Behavioral studies to evaluate the effectiveness of the systems developed.

A Scientific Background

Tutoring, one of the most effective methods for learning and knowledge transmission in humans, relies on the rich signals and real-time feedback that occurs in face-to-face interaction. Advances in machine learning and machine perception have resulted in perceptual software (e.g., face detection, person identification, expression recognition) that works reliably in everyday-life conditions. Computers have also become powerful enough to run this software in real time. This opens new possibilities for systems that interact with humans in an
affective, human-like manner, and that are sensitive to the sources of information (e.g., facial expressions, eye gaze) that make tutoring such an effective form of teaching.

List of Participants

- **Javier R. Movellan (PI):** Research Scientist. Institute for Neural Computation. UCSD. Research interests focus on machine learning, machine perception, and robotics.

- **Marian Stewart Bartlett (Co-PI):** Assistant Research Professor. Institute for Neural Computation. UCSD. Research interests in computer vision, expression recognition and affective science.

- **Gwen Ford Littlewort (Co-PI):** Research Scientist. Institute for Neural Computation. UCSD. Research interests focus on machine learning and computational neuroscience.

- **Aaron Cicourel (Co-PI):** Professor Emeritus. Department of Cognitive Science. UCSD. Research interests in the study of social change, digital ethnography, and ecological validity in behavioral and social science research.

- **Morana Alac (Co-Investigator):** Assistant Professor. Communication Department. UCSD. Research interests include ethnographic studies of emerging technologies.

- **Marjo Virnes.** Visiting Scholar. Computer Science Department. University of Joensuu, Finland.

- **Kathryn Owen:** Director of UCSD’s Early Childhood Education Center (ECEC) at UCSD.

- **Lydia Morrison:** Head Teacher, Room 1 at the Early Childhood Education Center (ECEC).

B Progress Report

In its three years of life the RUBI project sponsored by the UC Discovery program has generated a wealth of scientific articles spanning behavioral analysis of human-robot interaction [17, 34, 35, 36], new machine perception primitives [4, 5, 6, 20, 31, 32], new machine learning algorithms [7, 23], behavioral studies, and documentation of the process of designing robots by immersion [13, 25]. Below is a description of some of the research conducted under the project.

B.1 Behavioral analysis of human-robot interaction

During Year 1, the humanoid robot QRIO (See Figure ??) was immersed in a classroom of 18 to 24 month old toddlers for 45 sessions spanning 5 months (March 2005 to July 2005). QRIO is a 23 inch tall robot prototype built by SONY as the result of a long research and development effort [? ? ]. The robot displays an impressive array of mechanical and computational skills yet its ability to interact with humans for prolonged periods of time had not been tested. A detailed quantitative analysis of the 45 sessions conducted Year 1 of the RUBI project is currently in press at the *Proceedings of the National Academy of Sciences*. The study revealed the following insights:

2
Figure 1: (A): Independent coders were asked to operate a dial in continuous time to indicate their perception of the quality of the interaction between children and QRIO observed in video tapes. (B): Blue dots plot the average quality of interaction score on a random sample of 15 sessions of the study. The quality of the interaction steadily increased over the first 27 sessions when the robot exhibited its full behavioral repertoire (Phase I). During Sessions 28 to 41 (Phase II) the robot’s repertoire was restricted to dancing. The quality of the interaction plummeted. During the last 3 sessions (Phase III) the robot’s behavioral repertoire was restored at which point the quality of the interaction improved back to the levels observed during Phase I.
The quality of the child-robot interaction, as assessed by 5 independent coders, improved steadily as the project progressed except for a period that lasted 15 sessions in which the robot was programmed to behave in an interesting but repetitive manner (See Figure ?? ContAudRes B).

As the study progressed the children progressively treated QRI0 as a social entity rather than a conventional toy (See Figure ??).

It is possible to predict quite well the quality of the interaction between robots and children based on a signal processing model that uses the touch sensors of the robot (See Figure 2). The correlation coefficient between the model and the human evaluation of the quality of interaction across a total 1,244,224 frames was (0.78), almost as good as the average human-to-human agreement (0.80).

**B.2 Hardware Development**

While QRO was an impressive robot, it was not particularly well suited for teaching toddlers, which is a key component of the RUBI project. Thus one of our goals was to develop robots specifically designed to interact with and teach toddlers. Each year of the project we developed a different robot prototype that incorporated the lessons learned from the previous years. We respectively named them RUBI-1, RUBI-2, and RUBI-3 (See Figure 3). RUBI-1 was a three-foot tall, with a head, two arms, and a touch screen (See Figure 3). The robot was
remote controlled by a human operator. By the end of Year 1 children exhibited a variety of social behaviors towards her including pointing, hugging, imitation, and social referencing (see Figure 3-Left).

In Year 2 we redesigned RUBI’s hardware while keeping her external appearance relatively unchanged. RUBI-2 had more powerful computers than RUBI-1 to implement perceptual primitives like expression recognition, in real time. The main limitations of RUBI-1 and RUBI-2 was that the key aspects of the interaction with toddlers where handled by remote control from a human operator. The focus of the RUBI-3 prototype, was on fully autonomous operation. RUBI-3 is two-feet tall, a significant reduction in height that had a positive effect on the interactions that developed between children and robots. RUBI-3 has arms that allow it to play give-and-take games. So far RUBI-3 has been able to operate completely autonomously in the classroom for weeks at a time.

Figure 3: Left: RUBI-1 teaching materials targeted by the California Results Developmental Profile from the California Department of Education. Center: RUBI-2 goes for a walk with the children. Right: RUBI-3 Playing educational games in autonomous mode. So far her record has been 3 days of continuous autonomous operation.

B.3 Software Architecture: RUBIOS

We developed a software architecture for social robots named RUBIOS. The architecture provides a layer of abstraction on top of a network of computers, micro-controllers, sensors, and actuators. RUBIOS is built using a chat-room model. In its basic mode, it provides facilities for different robot services to register and communicate with each other the way people communicate with each other in Internet chartrooms. In addition RUBIOS provides facilities to implement a market oriented behavior model. Each node of a RUBIOS robot has a measure of success (i.e., its reason d’etre) and learns to schedule actions so as to optimize its long term success, i.e., its wealth. Rather than commanding actions, RUBIOS nodes interact with other nodes by means of temporal bids, i.e., promises to provide “funds” to other nodes provided particular goals are accomplished at specified points in time. Each node has access to a history of bids and makes decisions so as to maximize its own wealth. RUBIOS is on its second revision now and it provides the basic software architecture controlling RUBI-3. It is stable and shall be released to the research community by the end of the project.

B.4 Perceptual Primitives

As part of this project we found that many of the perceptual primitives (e.g., expression recognition) that worked well in laboratory conditions did not perform well in the relatively
uncontrolled conditions of the classroom. To improve this state of affairs we invested a great deal of effort on the collection and labeling of large datasets of faces and facial expressions from the Web. Images from the Web were targeted because they provide have high variability in lighting, background, and camera parameters, thus providing a diverse dataset for training robust computer vision systems using machine learning methods. These datasets have been used to develop robust face detectors and smile detectors that operate well in daily-life conditions. These systems were the foundation for algorithms used by SONY in their new generation of digital cameras.

B.5 Learning Primitives

The perceptual primitives we have developed so far are based on supervised machine learning methods, i.e., methods that require hand-labeling large collections of images. As part of the RUBI project we developed a mathematical framework for a new generation of machine perception algorithms capable of discovering perceptual categories in an unsupervised manner. We demonstrated the potential of this approach on a robot that discovered, in an autonomous manner the visual appearance of humans [26] without the need for labeled datasets. As explained later this year we propose to employ this approach to develop new perceptual primitives that recognize the general mood (the atmosphere) of classroom environments.

B.6 Interactive Educational Software

We developed a variety of computer games designed for RUBI to teach skills targeted by the California Desired Results Developmental Profile from the California Department of Education. In years 1 and 2, the games were presented on RUBI’s computer screen, and were decoupled from the rest of the robot behavior with the presentation and termination of each game was controlled by a human operator. In Year 3 the games were integrated with the rest of the robot’s behavior. For example, RUBI-3 gives and takes physical objects and, dances when the children perform well, and shakes its head when the children make mistakes. In addition RUBI-3 decides on its own when to start and terminate the educational games.

B.7 Impact of the Project

Below is a list of publications related to the RUBI project. Results from Year 1 of the project are currently in press for the Proceedings of the National Academy of Sciences. To our knowledge this will be the first paper on human-robot interaction ever published at a major scientific journal. Work from the RUBI project also received the Best Paper Award at the 2005 IEEE International Workshop on Robot and Human Interactive Communication. In addition to the academic publications, the RUBI project was featured in the APA monitor and Wired Magazine. Stories have also appeared in a wide variety of media outlets, including Good Morning America, AP, Univision, and the major local television stations in San Diego. Our experience with the RUBI project was a center-piece for the NSF Science of Learning Center recently awarded to UCSD. This may be a good indication of the positive effect that the UC Discovery program has by promoting research which gives UC institutions an edge in the very competitive federal research market.
B.8 List of Publications

2006-2007

- B. N. J. and M. J. R. Learning to learn. IEEE International Conference on Development and Learning (ICDL), 2007
- B. N., F. I., and M. J.R. Learning about humans during the first six minutes of life. Submitted

2005-2006

• F. Tanaka, B. Fortenberry, K. Aisaka, and J. R. Movellan. Plans for developing real-time dance interaction between QRIO and toddlers in a classroom environment. In Proceedings of the International Conference on Development and Learning (ICDL05), Osaka, Japan, 2005

2004-2005

• J. Susskind, J. Hershey, and J. Movellan. Exact inference in robots using topographical uncertainty maps. In Proceedings of the second international conference on development and learning (ICDL04), The Salk Institute, San Diego, October 20, 2004
• J. R. Movellan, J. Hershey, and J. Susskind. Large scale convolutional HMMs for real time video tracking. Computer Vision and Pattern Recognition, 2004
• T. K. Marks, J. Hershey, J. C. Roddey, and J. R. Movellan. 3D tracking of morphable objects using conditionally gaussian nonlinear filters. CVPR, 2004

C Proposed Research: Dataset Collection

Modern machine perception systems, of the type required for social robots to become a reality, are typically developed using machine learning methods applied to large datasets of images, sounds, or other sensory information. As such a critical component of the RUBI project is the collection of datasets to enable the development of practical systems. One such dataset, which we named GENKI\(^1\), consists of 63,000 images downloaded from the Web (see Figure ??). The images were manually labeled for the presence of smiles, head orientation, and location of

\(^1\)The meaning of GENKI in Japanese is similar to the Greek eudaimonia, signifying “robustness”, happiness and “good spirits.”
facial features. GENKI proved an invaluable tool to develop a robust smile detector. However it showed limitations that we propose to address during Year 4 [16].

- **Labeling:** We propose to continue coding of GENKI for different aspect relevant to social interaction. This shall include coding for gender, age, ethnicity, and expression beyond the presence or absence of smiles.

- **Ethnic Diversity:** Current face detection and expression recognition systems tend to perform worse on individuals with dark skin. This is in part due to the fact that they are typically underrepresented in the datasets used to train these systems. In Year 4 we propose to augment GENKI with particular emphasis on achieving a sufficiently large numbers of images from multiple ethnicities, particularly those for which the current systems do not perform as well.

In addition to continuing the work on GENKI, we propose collection of two new datasets:

- **ECEC-FACE-08:** Our work last year revealed an important limitation of Web based datasets, like GENKI. While such datasets contain a rich variety of image rendering conditions, they are typically limited in range of facial expressions: People tend to post images of posed, positive expressions, rather than more spontaneous expressions. During Year 4 we propose to complement GENKI with a new dataset of images automatically collected by the robots to be deployed at ECEC. The target is 30,000 labeled images of faces collected from the different classrooms at ECEC, with children from 6 months to 5 years of age. This new dataset will allow the development of perceptual primitives tailored to the actual operational conditions of educational robots.

- **ECEC-MOOD-08:** As part of the RUBI project we learned of the critical importance or the classroom mood or atmosphere. Robots that assist teachers in early education must choose different behaviors depending on whether the children are crying, laughing, sleeping, or singing songs. In order to develop mood recognition systems we propose to collect and label and audio-visual dataset of prototypical classroom moods. This dataset will be used to train and evaluate the mood detection systems proposed in later sections. This dataset will be collected in close collaboration with the teachers at ECEC. They will help identify the different mood categories relevant to their teaching activities. The
target size of the dataset will be 50 hours of annotated audio and video samples from the different classrooms at ECEC with children from 6 months to 5 years of age.

D Proposed Research: Perceptual Primitives

D.1 Smile Detection

Facial expressions are one of the most powerful and immediate means for humans to communicate their emotions, cognitive states, intentions, and opinions to each other [? ? ? ? ?]. Facial expressions played a critical role in the evolution of complex societies, helping coordinate social interaction, promoting group cohesion, and maintaining social affiliations. Among all facial expressions, smiles, are arguably one of the most basic, recognizable, and
useful ones [? ? ? ]. They may engage up to 7 different pairs of facial muscles but in practice the prototypical smile only needs to involve the Zygomatic Major [? ]. Last year we focused on detection of such prototypical smiles, sometimes known as “Zygomatic Smiles.” The requirement was for the system to work reliably in unconstrained daily life conditions. The system turned out to work very reliably in adults, 98 % correct on the GENKI dataset [16]. However, when tested on images captured by RUBI-3 while playing with children, the performance declined to 88 % correct. This was in part due to the fact that spontaneous smiles are more subtle than the posed smiles presents on the GENKI dataset, and also due to the fact that the appearance of toddler smiles is quite distinct from adult smiles (e.g., many toddler have very few teeth). For year 4 we propose to refine the current smile detector and tailor it so that it works as reliably on children as it does on adults. The methods for development of the detector will be similar to those used for development of the adult smile detector. However here we will use the new ECEC-FACE-08 dataset.

**Person Identification:** An important limitation of RUBI-3 is that it does not recognize the identity of the children it interacts with. Thus, it can only track the learning progress of the classroom as a whole, rather than individual children. In year 4 we propose to develop algorithms for RUBI-4 to recognize the children it interacts with using video. The procedure for developing person identification will be similar to that we used in the past to develop expression recognition systems. First images of faces automatically collected by the robot’s face detection system will be labeled for individual identity. Second pairwise classifiers will be developed using Boosting methods [? ? ] to recognize the difference between pairs of individuals. Thus, for a classroom of 10 children, 45 classifiers will be automatically developed. The output of the pairwise classifiers will then be combined using ridge multinomial logistic regression [? ] to make a decision about the identity of the detected faces. We have experience developing perceptual systems using this architecture with very good results [4, 5, 6, 16, 20? ].

**Auditory Mood Detection:** Social robots of the type explored in this project need to detect and adapt their behavior to the current social mood. For example, RUBI shall behave differently depending on whether the children are crying, laughing, sleeping, or singing songs. We propose to develop a prototype mood detection system that uses both Audio and Video to make decisions about the current classroom atmosphere. First we will develop an Audio based system. The system will extend and refine a novel approach explored last year as part of the RUBI project [? ]. The approach is based on a new set of low-level spectral contrast features similar to the ones we used in the visual domain for development of face detection and expression recognition systems. These features are selected and combined using Boosting methods [? ? ] so as to make decisions about the ongoing auditory mood. We have already demonstrated excellent performance of the system on two standard emotional speech databases (the Berlin Emotional Speech [? ], and the ORATOR dataset [? ]). During Year 4 we propose to collect ECEC-MOODS-08, a dataset of classroom moods and use it in conjunction with the aforementioned approach to develop auditory mood detection for toddler classroom environments.
E  Proposed Research: Autonomous Blogging

Social robots like RUBI operate in a social ecology that includes toddlers, teachers, parents and researchers. Thus, for RUBI to be effective it is important that it interacts with all the members of this ecology, not just the children. RUBI-3 currently communicates with researchers via periodic emails that provide information about its internal state, a log of the performance of the children on the different teaching activities, and the image with largest score from the smile detector obtained since the last email. In Year 4 we propose to refine and extend this method of communication to involve other adults, including parents and teachers. We do not have experience developing systems of this kind but believe the following approach is a good start and has a good chance of success. First a simple automatic blogging program will be developed using simple heuristic rules. For example, teachers will be sent information about the performance of the children on the different educational games, thus given them an indication of what particular skills the children may be weaker or stronger at. They will also be sent sounds and images representing of life at the classroom during that day. These images will be chosen based on hand coded indexes that combines the different perceptual primitives available to the robot: frequency of touching, number of faces detected, mood detected, people detected.

This heuristic method will then be automatically adapted to the needs and preferences of individual parents and teachers. To this effect users will be asked to rate their preference for the images automatically blogged. Each image has a tag characterizing the conditions under which it is captured, e.g., the output of the smile detector, the number of faces detected, the output of the proximity sensors, the children recognized by the robot, the number of times the robot was recently touched, the ongoing activity with the children, the day of the week, the time of the day. These variables can be combined using standard statistical methods, like multinomial logistic regression [?] to predict the observed user preferences. For example, this approach may make apparent that a particular likes images that display a particular child, or images that depict a specific mood of the classroom. Time permitting, collaborative filtering methods that capitalize on patterns of common preference between clusters or users will also be explored [? ?].

F  Proposed Research: Learning To Teach

One goal of the RUBI project is to explore the use of social robots as tools for teaching. We have developed a series of educational games designed to teach toddlers materials targeted by the California Desired Results Developmental Profile from the California Department of Education. These skills include recognition of geometric shapes, colors, numbers, facial expressions, and names of familiar people. The structure of these games is as follows: RUBI presents 4 items on the screen (e.g., pictures of 4 geometric shapes), and says the name of a target shape (e.g., “Triangle”). If the children touch the wrong item, RUBI indicates that the answer is incorrect via sounds (e.g., “Nope”, “Try again”) and motor behaviors (e.g., shaking the head). When the children choose the correct item, RUBI dances, claps, and provides auditory rewards (e.g., a cheering sound). RUBI keeps a log of the performance observed on the different games.

Currently RUBI-3 terminates a game if the children do not touch the screen for a period of time, at which point it goes into idle mode. RUBI-3 exits idle and starts a randomly selected gain if the children touch the screen. In Year 4 we propose to explore machine learning methods for RUBI to adaptively improve the way it schedules these educational games. This
will be done within the mathematical framework of stochastic optimal control [? ? ?]. The goal is to develop control policies that map currently available information (e.g., person identification, classroom mood) into actions that optimize a measure of success. Reinforcement learning methods [33] designed to learn approximately optimal control policies will be applied. These methods require for the learning system to be provided with a quantitative measure of success (a reinforcement signal). We propose to utilize a signal that will combine multiple educational and behavioral goals: (1) Percent correct on the different target items. (2) Lack of undesirable moods, like “crying”. (3) Amount of time interacting with the children. (4) An external signal provided by teachers using a remote controller for the desirability of the behaviors chosen by RUBI.

G Proposed Research: Robot Development

For year 4 we propose design and construction of 3 RUBI-4 robots. The RUBI-4 prototype will be similar in appearance to RUBI-4 but will include a more advanced head, with pan and tilt (as opposed to the pan only in RUBI-3) and with simple facial expressions (2 degree of freedom mouth and 2 degree of freedom brows). These robots will operate continuously and fully autonomously in Rooms 1, Room 2a and Room 2b at ECEC. The robots will be sending data to a central server allowing the gathering of data about the progress of the children in the different classroom. The long term vision is a living educational laboratory consisting of a web of classrooms with robots that continuously send information to a repository for datamining, discovery of patterns, and the formulation and testing of educational theories and intervention methods.

G.1 Proposed Research: Behavioral Studies

By now we have collected a great deal of anecdotal evidence from teachers and parents suggesting that the robots are effective teachers and help enrich the classroom activities. For example, anecdotally parents have told us that when we brought RUBI to school, their child started pointing to objects at home and saying their names. Most of these objects, but not all, were being taught by RUBI at school. Should the robots prove effective teachers, we shall also attempt to clarify what components are responsible for their success, e.g., does the social aspect of the robot help or would a computer kiosk perform equally well?

The first study would be conducted on February 2007, using the current RUBI-3 prototype. The experimental design will be as follows. First the set of items currently taught by RUBI-3 will be partitioned into a experimental set and a control set. There will be a 2 week training period in which RUBI-3 will operate in a classroom fully autonomously 24 hours a day, 7 days a week. RUBI will present items from the training set with high probability, but periodically will also probe the children with items from the test set, so as to track the timecourse of spontaneous learning during a 2 week period. In addition there will be a pre-training and post-training test procedure. These tests will be performed as follows: A collection of physical objects representing items from the training and control sets will be located in a box far enough from the experimenter so he/she can't see what is inside. On command the experimenter will ask the child for an object from the training and test set (e.g., “can I get a banana”?). If the child retrieves the correct object it would count as success, otherwise as failure. The tests will be conducted one child at a time. The dependent measure will be the

---

2 RUBI autonomously goes into idle mode when no children are present
degree of improvement on retrieving items from the training set, when compared to items from the control set. In addition we shall correlate the degree of learning as a function of the amount of time each child spends playing with RUBI during the 2 week period.

Parents and teachers will also be given diaries with specific questions selected to track the potential effect of the RUBI on the behavior of the children at home.

By the end of Year 4 we will perform a second study using the new RUBI-4 prototypes. This shall allow to perform the study on 3 classrooms simultaneously. The goal of this study shall be to clarify which aspects of the robot are responsible for its teaching performance. To this effects the robot shall be programmed in different ways on the different classrooms: In one classroom the robot will exhibit its full behavior repertoire, including its adaptive teaching system. In another classroom the robot will use a non-adaptive teaching engine. Finally on the third classroom robot will behave as a computer kiosk, i.e., its sensory and motor systems will be turned off. Variables of interest will include the amount of time spent by the children playing educational games, and the improvement on the target skills.

Ethnographic methods will be employed to code and compare the behaviors of the children towards the different versions of RUBI and towards human teachers. Professor Aaron Cicourel will investigate the impact of the introduction of RUBI in the classroom. What aspects of it were liked/disliked by the teachers? What are the parents reactions? Do they perceive RUBI as a valuable teaching tool? In what ways does it affect the daily routines of children, teachers and parents? Do children mention RUBI to their parents? Do they talk to the parents about the materials she is teaching?

Figure 6: **Left:** RUBI teaching materials targeted by the California Results Developmental Profile from the California Department of Education. **Right:** Screen capture of a favorite game. Pointing to the face of a classmate. RUBI says the name of a classmate and the children point to it on RUBI’s touch-screen.
References


