Leveraging Computer Vision and Humanoid Robots to Detect Autism in Toddlers

*

Marie D. Manner
Dept. of Computer Science
and Engineering
University of Minnesota
manner@cs.umn.edu

Jed Elison
Institute of Child Development
University of Minnesota
jtelison@umn.edu

Maria Gini
Dept. of Computer Science
and Engineering
University of Minnesota
gini@cs.umn.edu

Abstract

Autism Spectrum Disorder is a developmental disorder often characterized by limited social skills, repetitive behaviors, obsessions, and/or routines. Using the small humanoid robot NAO, we designed an interactive program to elicit common social cues from toddlers while in the presence of trained psychologists during standard toddler assessments. Our experimental design captures four different videos of the child-robot interaction, and we intend to use captured video data to create a software package that helps clinicians diagnose autism spectrum disorder earlier than the current average age of 5 years. As part of our plan for automated video analysis to flag autistic behavior, we built and tested semi-automated software that logs proxemics information, and tested it on a large group of typically developing children with the robot program.

1 Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder defined by behavioral symptoms that include social communication deficits, and restricted and repetitive behavior patterns. Trained therapists use the Autism Diagnostic Observation Schedule to assess individuals, which is a semi-structured, time-intensive assessment that includes imaginative play, social cues, and communication. Early identification of children with autism allows intensive intervention before atypical patterns of behavior and brain function become firmly established. In 2014, the Center for Disease Control and Prevention released its latest estimate that one in 68 children has an ASD. Early intervention significantly improves long-term outcomes for toddlers identified in the second year of life [Dawson *et al.*, 2012] and is the best approach for affecting lasting positive change for children with ASD. The cause of ASD is unknown, and interventions are primarily designed to treat exceptionally complex, established behaviors, such as learning to interpret social situations.

Robotics research in autism is over a decade old, yet does not currently meet standards of psychology and child development researchers [Diehl *et al.*, 2012; Scassellati *et al.*, 2012], Robotics research in so-called Socially Assistive Robotics stems from the fact that children with autism especially enjoy robots [Dautenhahn and Werry, 2004].While the reason for this is unknown, researchers clearly have the potential to leverage robotics for autism diagnosis or treatment [Scassellati, 2005]. Problems with existing research include lack of robot integration to established treatments, lack of study participant followup, small sample sizes, little scrutiny on the actual therapeutic protocol, and little detailed characterization of participants [2012].

In [Manner, 2015] we asked two questions: First, can we use a small humanoid robot with toddlers to reveal symptoms of autism? Second, can we create video processing software to help clinicians diagnose toddlers with autism? To explore the first question, we designed a protocol in which a humanoid robot interacts with very young children to elicit joint attention, a highly important pro-social behavior, and began testing with children aged 2-3 years old enrolled in a longitudinal study looking at early affect, behavior, attention, and reciprocal relationships. We adhere to Diehl's suggestions by using participants in our study that will be followed over three years, using large sample sizes (with an N of 28 in our ground-truth, typically developing sample), subjecting our protocol to the review and feedback of clinicians and psychologists, and collecting complete participant information from multiple surveys and standard assessments, including the Mullen Scales of Early Learning and the Vineland Adaptive Behavior Scales. Since we are seeking to help diagnose, rather than treat, autism spectrum disorder, we will try to establish our protocol in diagnostic sequences.

To explore the second question, we built configurable software to perform semi-automated location tracking of persons (or robots) of interest in overhead video footage, providing the first quantitative metric for our typically developing group. The software tracks the location of the child, robot, caregiver, and experimenter, allowing us to log distances between the child and the others, showing clear differences in comfort-seeking behavior (seeking out the caregiver), avoidance behavior (hiding from the robot or experimenter), or indifference (ignoring or otherwise not engaging with the robot) on the part of the child. Diehl also noted in [Diehl *et al.*, 2014] the lack of research in using robots for diagnosis, where robots have the potential to control for human differences in

*Work supported in part by the National Science Foundation under grant IIP-1439728

illiciting behaviors and providing quantitative measurements of important diagnostic information like eye contact and gestures. Our work aims to help fill this gap.

Thus far our novel contributions are quantitative proxemics information as well as the distance-metric generating software, tested on a large number of subjects, with a reproduceable and portable robotic program. This is part of our larger goal to produce automated software that can flag suspicious markers of autism or other developmental disorders.

2 Related Work

We began our video processing by tracking proxemics, the study of the amount of space individuals need. Studies like [Asada *et al.*, 2016] have shown atypical distance requirements for individuals with ASD, generally that individuals with ASD require less personal space than typically developing individuals. [Feil-Seifer and Matarić, 2011; Feil-Seifer and Mataric, 2012] studied automated classification of negative and positive reactions of children with autism towards a large robot using distance features. This free-play scenario allowed the child to interact with or hide from the robot. [Mead *et al.*, 2013] studied automatic annotation of dyadic interactions of two people with a non-responsive robot as a social cue, using individual, physical, and psychophysical features (e.g. torso pose, distance, and 'social' distance).

For the larger goal of automated flagging software, we look to research on classifying symptoms and behaviors. [Hashemi *et al.*, 2012] successfully classified symptoms of arm asymmetry, visual tracking, and attention disengagement. [Hashemi *et al.*, 2014] more closely studied disengagement of attention and visual tracking abilities of infants, and showed their system agrees with experts in most trials. More recently, [Fasching *et al.*, 2015] has successfully classified repetitive body movements like hand flapping, shrugging, and ear-covering from video footage.

3 Experimental Setup

Robot

We use the small humanoid robot NAO from Aldebaran Robotics; the NAO is about two feet tall, has 25 degrees of freedom, and many sensors and colored LEDs (see Fig. 1). Each participant session includes at least one trained psychologist, a toddler and at least one caregiver, a NAO robot, and a data collector controlling the NAO. The robot is controlled wirelessly from a laptop in the same room, running a static program, and the experimenter is always next to the robot to prevent the child from getting too close to the robot and injuring himself or damaging the robot.

Participants

Subjects are drawn from a larger study, in which parents fill out four behavioral assessments for their children; random parents who have filled out all four surveys are asked to come in with their child for two more standard developmental assessements as well as the robot interaction.

We collected data on 28 typically developing (TD) children in all, ranging from 125 - 144 weeks old (2.40 to 2.76 years) at the time of the assessment, with an average of 136 weeks (2.61 years). These childen averaged 120 for Non-Verbal (Intelligence) Quotient, 121 for Verbal (Intelligence) Quotient, and 123 for Early Learning Composite (a combination of assessments over visual reception, fine motor skills, receptive language, and expressive language). Each value is similar to the intelligence quotient in adults, with 100 being average and a standard deviation of 30, meaning this group of toddlers are in the higher end of the average population (possibly due to parental self-selection of finishing all initial surveys).

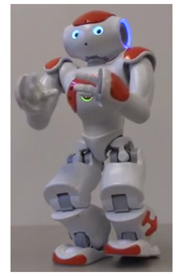

Figure 1: NAO in mid-dance.

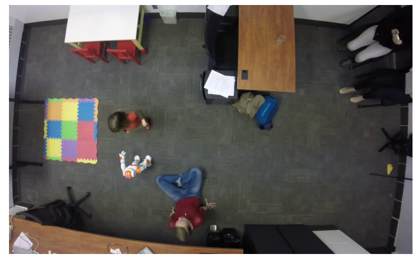

Figure 2: Overhead perspective of child dancing with NAO.

Study Protocol

The child and caregiver are brought in to the study room by a psychologist to take a behavioral assessment, the Mullen Scales of Early Learning (MSEL). The robot is already in the room, off but visible to the child, and the psychologist tells the child the robot is also interested in how children play, and that after the 'games' (the MSEL assessment) the child will play with the robot. After the MSEL assessment is complete, in roughly 40-50 minutes, another experimenter enters the room to introduce the robot, turn it on, and after an initial warm-up period, play the robot interaction program with the child. The program takes 9-16 minutes, depending on if the child is focused and interested in playing with or watching the robot, or is easily distracted and takes frequent breaks; the experiment time is the duration of the robot program.

During the robot interation, four video perspectives are taken: from two cameras at the north and south end of the room, from a camera mounted on the ceiling, and from the robot's point of view (see Fig. 2 for the overhead view).

The robot runs a structured play series consisting of simplified versions of children's games "I spy," "Simon Says," and various dances. After introducing itself, the robot performs two games of "Simon Says," two games of "I spy," and three song and dance routines in a static order. In each "Simon Says" round, the NAO demonstrates actions that can be done

Figure 3: Child initially used the experimenter for comfort. Participant face blurred for privacy.

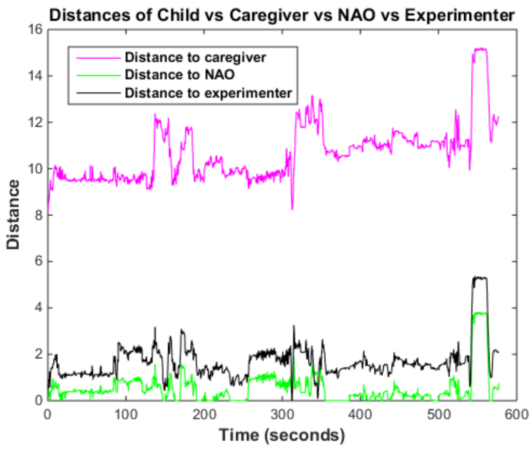

Figure 4: Child that complied with almost all of NAO's requests in looking for objects, simulating actions, and dancing.

with gross motor skills, like arm flapping or hand clapping. In each "I spy" round, NAO looks around and states "I see a..." before stating an object. Before each dance, NAO invites the child to dance with it, and plays a different song and dance. The NAO sequences do not depend on the child interacting with it, but waits for 90 seconds or for the experimenter to start the next sequence, which allows the experimenter to draw the child's attention back to the robot or the child to take a brief break. Each sequence begins with encouragement or ends with positive words like "that was fun!"

Results

We have finished collecting data on 28 typically developing children. We plan to schedule children at high risk for autism to come in for robot assessments in the next two months; many high risk children are already scheduled for an Autism Diagnostic Observation Schedule (ADOS) assessment, and we are working to include or subsequently schedule the robot interaction with these children. We used our software to log the distances between the child and caregiver, robot, and experimenter, which gives us more information about social distance required with a stranger (the experimenter) for TD children vs. children with ASD. See e.g. Fig. 3, in which a child used the experimenter, rather than caregiver, for reassurance.

Some exceptional proxemics examples from TD participants may be found in Fig. 4 and Fig. 5. Fig. 4 shows a child who did not request his caregiver throughout the interaction; he played every "I Spy" and "Simon Says" game

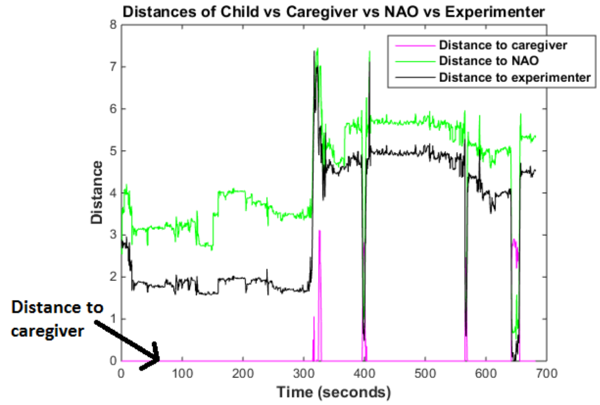

Figure 5: Child that sat with caregiver throughout most of robot interaction, with several instances of approaching robot to deposit or collect toys.

and danced with the robot on all three dances. When the experimenter told him the robot had played all of its games, he hugged the robot before it sat down. Fig. 5 shows a child that stayed with his caregiver almost the entire time; he watched the robot closely but rarely followed prompts to look around, do actions, or dance. He approached the robot twice to put his toys in front of it, and twice to take his toys back.

Other data collected included a child who became easily distracted, wrestling with his caregiver for much of the robot interaction, some children who did not want to play with the robot but were willing to watch it, and four children who became distressed or exhausted enough that the experimenter ended the experiment early. Other responses included becoming distracted or disinterested when the robot repeated games, (e.g., the second round of "I Spy"), becoming more interested when the robot stood up to dance, and seeking caregiver assurance when the robot stood up to dance. Other proxemics show varied and rapidly changing distances between the child and the NAO, experimenter, caregiver, after the initial novelty of the robot wears off and the child moves around the room more. Of the 28 experiments, 5 children did not finish (4 mentioned earlier, 1 that ended due to equipment failure); of the 23 completed experiments, the minimum run was 9 min. 2 sec., the maximum 15 min. 17 sec.; the mean time was 9 min. 59 sec. with a SD of 1 min. 58 sec.

4 Future Work

Our over-all goal is to give a toolbox to clinicians that includes off-line video processing and behavior flagging, and we will continue to work with psychologists to add automated tools to enable easier diagnosis. Our next step is to run the robot program with high risk children, compare approach and avoidance behaviors with our typically developing base group, and try to identify movement patterns for TD and high-risk children. We are also looking at the robot's point of view footage to study how frequently the child is meeting the robot's gaze and to detect expressions.

References

[Asada *et al.*, 2016] Kosuke Asada, Yoshikuni Tojo, Hiroo Osanai, Atsuko Saito, Toshikazu Hasegawa, and Shinichiro Kumagaya. Reduced personal space in individuals with autism spectrum disorder. *Plos One*, 11(1), 2016.

[Dautenhahn and Werry, 2004] Kerstin Dautenhahn and Iain Werry. Towards interactive robots in autism therapy: background, motivation and challenges. *Pragmatics & Cognition*, 12(1):1–35, 2004.

[Dawson *et al.*, 2012] Geraldine Dawson, Emily J H Jones, Kristen Merkle, Kaitlin Venema, Rachel Lowy, Susan Faja, Dana Kamara, Michael Murias, Jessica Greenson, and Jamie Winter. Early behavioral intervention is associated with normalized brain activity in young children with autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(11):1150–1159, 2012.

[Diehl *et al.*, 2012] Joshua J Diehl, Lauren M Schmitt, Michael Villano, and Charles R Crowell. The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Research in autism spectrum disorders*, 6(1):249–262, 2012.

[Diehl *et al.*, 2014] Joshua J Diehl, Charles R Crowell, Michael Villano, Kristin Wier, Karen Tang, and Laurel D Riek. Clinical applications of robots in autism spectrum disorder diagnosis and treatment. *Comprehensive Guide to Autism*, pages 411–422, 2014.

[Fasching *et al.*, 2015] Joshua Fasching, Nicholas Walczak, Vassilios Morellas, and Nikolaos Papanikolopoulos. Classification of motor stereotypies in video. *IEEE International Conference on Intelligent Robots and Systems*, pages 4894–4900, 2015.

[Feil-Seifer and Matarić, 2011] David Feil-Seifer and Maja J Matarić. Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features. *Proceedings of the 6th international conference on Human-robot interaction - HRI '11*, pages 323–330, 2011.

[Feil-Seifer and Mataric, 2012] David Feil-Seifer and Maja Mataric. Distance-based computational models for facilitating robot interaction with children. *Journal of Human-Robot Interaction*, 1(1):55–77, 2012.

[Hashemi *et al.*, 2012] Jordan Hashemi, Thiago Vallin Spina, Mariano Tepper, Amy Esler, Vassilios Morellas, Nikolaos Papanikolopoulos, and Guillermo Sapiro. Computer vision tools for the non-invasive assessment of autism-related behavioral markers. *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics, ICDL 2012*, pages 1–33, 2012.

[Hashemi *et al.*, 2014] Jordan Hashemi, Mariano Tepper, Thiago Vallin Spina, Amy Esler, Vassilios Morellas, Nikolaos Papanikolopoulos, Helen Egger, Geraldine Dawson, and Guillermo Sapiro. Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants. *Autism research and treatment*, 2014.

[Manner, 2015] Marie D Manner. Using small humanoid robots to detect autism in toddlers. *IJCAI International Joint Conference on Artificial Intelligence*, pages 4383–4384, 2015.

[Mead *et al.*, 2013] Ross Mead, Amin Atrash, and Maja J Matarić. Automated proxemic feature extraction and behavior recognition: applications in human-robot interaction. *International Journal of Social Robotics*, 5(3):367–378, 2013.

[Scassellati *et al.*, 2012] Brian Scassellati, Maja J Matarić, and Henny Admoni. Robots for use in autism research. *Annual Review of Biomedical Engineering*, 14(1):275–294, 2012.

[Scassellati, 2005] Brian Scassellati. Quantitative metrics of social response for autism diagnosis. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 2005:585–590, 2005.