How social robots will help us to diagnose, treat, and understand autism

Brian Scassellati

Yale University, 51 Prospect Street, New Haven, CT, 06520, `scaz@cs.yale.edu`

Autism is a pervasive developmental disorder that is characterized by social and communicative impairments. Social robots recognize and respond to human social cues with appropriate behaviors. Social robots, and the technology used in their construction, can be unique tools in the study of autism. Based on three years of integration and immersion with a clinical research group, this paper discusses how social robots will make an impact on the ways in which we diagnose, treat, and understand autism.

1 Introduction

For the past three years, our robotics group has been immersed in one of the premiere clinical research groups studying autism, led by Ami Klin and Fred Volkmar at the Yale Child Study Center. This paper outlines our initial attempts to apply technology from social robotics to the unique clinical problems of autism.

Section 2 provides an introduction to autism which highlights some of the difficulties with current diagnostic standards and research techniques. Section 3 describes attempts to use robots as therapeutic aids and discusses the as yet unfulfilled promise of these methods. Section 4 describes how diagnosis can be improved through the use of both passive social cue measurement and interactions with a social robot to provide quantitative, objective measurements of social response. Section 5 speculates on how the use of social robots in autism research might lead to a greater understanding of the disorder.

2 What we know about autism

Autism was first identified in 1943 by Kanner who emphasized that this congenital condition was characterized by an inability to relate to other people from the first days of life. Over the past 6 decades considerable work has been done to refine the concept and identify important aspects of the condition. Current research suggests that 1 in every 300 children will be diagnosed with the broadly-defined autism spectrum disorder (ASD), but studies have found prevalence rates that vary between 1 in every 500 to 1 in every 166. For comparison, 1 in every 800 children is born with Down syndrome, 1 in every 450 will have juvenile diabetes, and 1 in every 333 will develop cancer by the age of 20. Furthermore, the rate of diagnosis increased six-fold between 1994 and 2003. It is unclear how much of this increase is a result of changes in the diagnostic criteria, increases in awareness, or a true increase in prevalence. Early intervention is critical to enabling a positive long-term outcome, but even with early intervention, many individuals will need high levels of support and care throughout their lives [fDCP06].

The social disability in autism is a profound one affecting a person's capacity for understanding other people and their feelings, and for establishing reciprocal relationships. To date, autism remains a behaviorally specified disorder [VLB$^+$04]; there is no blood test, no genetic screening, and no functional imaging test that can diagnose autism. Diagnosis relies on the clinician's intuitive feel for the child's social skills including eye-to-eye gaze, facial expression, body postures, and gestures. These observational judgments are then quantified according to standardized protocols that are both imprecise and subjective (e.g. [SBC84, Mul95]). The broad disagreement of clinicians on individual diagnoses creates difficulties both for selecting appropriate treatment for individuals and for reporting the results of population-based studies [KLCV00, VCK05].

The need for improved characterization of the core social disorder in autism that underlies the broad spectrum of syndrome manifestations has been highlighted by genetic and neuro-functional research [VLB$^+$04, SR05]. It is clear that autism is a brain-based disorder with a strong genetic basis. Approximately 25% of children with autism develop seizures and the recurrence risk for siblings is between 2 and 10% (a 50-100 fold increase over the general population). Genetic studies have underscored the importance of understanding both the broader phenotype of autism and the remarkable heterogeneity in syndrome expression. However, the causes and etiology of the disorder are still unknown [VLB$^+$04]. A more precise characterization and quantification of social dysfunction is required to direct neurobiological research in autism is still lacking [BPR96, KJS$^+$02a].

3 Robots provide motivation and engagement in therapy

A few projects world-wide seek to include robots as part of the therapeutic regimen for individuals with autism [WD99, MTT02, Dau00, KNY]. Each of these studies has demonstrated that robots generate a high degree of motivation and engagement in subjects, including subjects who are unlikely or unwilling to interact socially with human therapists. The great hope of this line of research is the development of a ”social crutch,” a robot that motivates and engages children, teaches them social skills incrementally, and assists in the transfer of this knowledge to interactions with humans. Since the behavior of a robot can be decomposed arbitrarily, turning off some behaviors while leaving others intact, we can selectively construct complex social abilities through layers of social responses, sometimes in combinations that cannot be performed by humans. This layering of response allows the therapist to focus on single behaviors while ignoring all other social factors or maintaining their response at a constant. This type of isolation of cues and responses is difficult to train human therapists to perform. The as yet unfulfilled promise of this line of research is that learning skills with a robot will be simpler because of the ability to isolate particular responses, thus allowing a unique form of incremental therapy. In a different domain, but using a similar principle, we have preliminary data suggesting that computerized face perception training leads to therapeutic benefits for individuals with autism [SK05].

However, the design criteria for what makes individuals with autism likely to respond to these devices are not understood. The robots used in these studies include four-wheeled rovers, anthropomorphic robotic dolls, a spherical robot ball with eyes, and an expressive snowman-like device. These robots show a wide range of anthropomorphic characteristics, behavioral repertoires, aesthetics, and sensory and interactive capabilities. While there are many studies of the effects of these interaction variables on typical adults, very little is known about how individuals with autism respond to these design dimensions. While we have many expectations for why children with autism respond so positively to these robots, we have no direct experimental data that provide an analysis of the design criteria that are important to producing this response. We would expect that one reason that children (both autistic and typically developing) would respond so positively to robots (as seen in the studies mentioned above) is that the robots offer simple, contingent, predictable responses to the child’s actions. The preference for things that interact with us, that respond directly to our actions, is well known for typical adults and children. However, our expectations derived from studies with typical adults and children often do not carry over to adults and children with autism.

As an example of how our expectations regarding these design criteria are misplaced, we conducted a simple pilot study that looked that the effects of social contingency. Using an extremely simple commercial robot called ESRA (see Figure 1) which generates a small set of facial expressions using five servos, we compared children’s attentiveness to the robot in two experimental conditions. In the non-contingent condition, ESRA was programmed to perform a short script which included both a set of actions and an accompanying audio file that was played from speakers hidden near the robot. The robot had no sensory capabilities and did not respond to anything that the child did. In the contingent condition, the robot performed behaviors from this same repertoire, but the initiation of these behaviors was triggered by an experi-

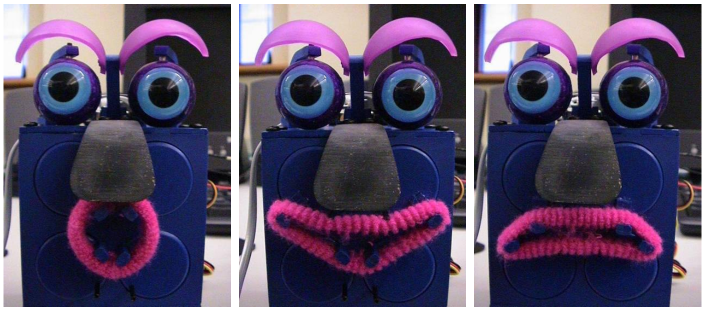

Fig. 1. Three facial expressions from the ESRA robot.

menter sitting behind a one-way mirror. The experimenter triggered behaviors that they deemed to be socially appropriate based on the actions of the child. 13 subjects (mean age 3.4 years) including 7 children with autism spectrum disorders and 6 typically developing children were positioned across a table from ESRA for a period of 3-5 minutes. Even with the extremely limited capabilities of ESRA, the robot was well tolerated by all of the children and many of them (including many of those within the autism spectrum) seemed to thoroughly enjoy the session.

In the contingent condition, typical children were universally engaged with the robot, and often spent the entire session touching the robot, vocalizing at the robot, and smiling at the robot. In the non-contingent condition, typically developing children were initially attracted to the robot but tended to lose interest quickly, preferring instead to attend to other (non-robotic) toys that were in the room. In contrast, the children with autism did not differ significantly in their interactions between these two experimental conditions. They tended to spend almost all of the session attending to the robot, regardless of whether or not it was responding contingently to them. In both conditions, children with autism often generated behavior similar to their typically developing peers, including smiling at the robot, making eye contact, and vocalizing to the robot. For many of the children with autism in this pilot study, these positive proto-social behaviors are rarely seen in a naturalistic context.

These results are only preliminary, but they point out that the severe social deficits that accompany this disorder do not respond in accordance with the interaction dynamics that have been observed with typical adults by research in human-robot interaction. These results should also not be interpreted to show that children with autism fail to respond to any form of social contingency. Because of the very wide range of functional capacities of children and adults who receive the diagnosis of autism (see the following section), these types of generalizations are notoriously dangerous. This simple study does demonstrate that further detailed study of these design variables are necessary to begin to delineate the factors that cause this remarkable response from children with autism.

4 Quantitative objective metrics for diagnosis

Many of the diagnostic problems associated with autism would be alleviated by the introduction of quantitative, objective measurements of social response. We believe that this can be accomplished through two methods: through passive observation of the child at play or in interactions with caregivers and clinicians, and through structured interactions with robots that are able to create standardized social ”presses” designed to elicit particular social responses. While the information gathered from both passive and interactive systems will not replace the expert judgment of a trained clinician, providing high-reliability quantitative measurements will provide a unique window into the way in which children with autism attempt to process naturalistic social situations. These metrics provide both an opportunity to compare populations of individuals in a standardized manner and the possibility of tracking the progress of a single individual across time. Because some of the social cues that we measure (gaze direction in particular) are recorded in greater detail and at an earlier age than can occur in typical clinical evaluations, one possible outcome of this work is a performance-based screening technique capable of detecting vulnerability for autism in infants and toddlers.

4.1 Passive sensing

Passive sensors record information on social response without directly engaging in interactions. In many cases, the perceptual systems of a social robot can act as a passive social cue sensor. To evaluate the usefulness of this idea, we have outfitted some of our clinical evaluation rooms with cameras and microphones and software similar to that used on the social robots Nico, Cog, and Kismet [Sca03, Sca01, BEF$^+$00]. Most of these passive sensors record and interpret data while the subjects are actively engaged in standard clinical evaluations and do not require any specific protocol to be employed. Currently, three cue recognition systems have been developed: (1) detecting gaze direction, (2) tracking the position of individuals as they move throughout a room, and (3) measuring aspects of prosody from human voices.

Gaze direction and focus of attention

For several years, we have used commercial eye-tracking systems which require subjects to wear a baseball cap with an inertial tracking system and camera/eyepiece assembly which allows us to record close-up images of one eye. In addition to this commercial system, we have developed computational systems that give much less accurate recordings but do not require the subject to be instrumented. When viewing naturalistic social scenes, adolescents and adults with autism display gaze patterns which differ significantly between control populations (see Figure 2) [KJS$^+$02a, KJSV03, KJS$^+$02b]. Fixation time variables predicted level of social competence (e.g., at an average

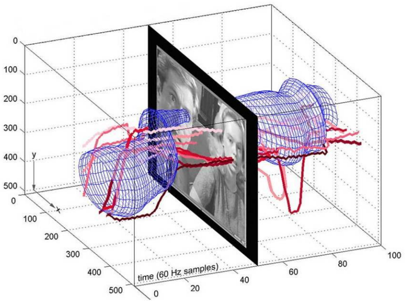

Fig. 2. Gaze patterns differ significantly between typical adolescents and adolescents with autism. This spatio-temporal plot shows 10 scan paths of individuals with autism (red lines) and the bounding volume (in blue) for 10 typical individuals. The typical group shows very similar, structured gaze patterns. The group with autism shows less structure, but is far from random. (Figure adapted from [KJS$^+$02a]).

r=.63). This was the first experimental measure to successfully predict level of social competence in real life for individuals with autism. Visual fixation data related to viewing of naturalistic scenes of caregivers' approaches reveals markedly different patterns. Toddlers with autism fixate more on the mouth region rather than on eye regions of faces. Combined with experiments probing these children's capacity for mentally representing human action, it has been suggested that these children are treating human faces as physical contingencies rather than social objects (they fixate on mouths because of the physical contingency between sounds and lip movements). Although visual fixation on regions of interest are sensitive measures of social dysfunction, moment-by-moment scan-paths are even more sensitive and offer further insight into the underlying dysfunction (see section 5 for an example) [KJS$^+$02a].

Position tracking

Some of the most basic information on social response can be derived from the relative positioning of individuals. How close a child stands in relation to an adult, how often the child approaches an adult, how much time is spent near an adult, and whether or not the child responds when an adult approaches are a few of the relatively simple statistics that can be derived from positional information. These social cues, especially the concept of "personal space," are often deficient in individuals with autism and are part of the diagnostic criteria [VLB$^+$04].

Using a pair of calibrated stereo cameras and a computational vision system developed in part by our team, we have been able to successfully track the position of individuals as they move about in our clinical space. Computed disparity information is used in conjunction with information on color, direction of motion, and background pixels to segment the moving objects in the scene. A multi-target tracking system (similar to the system developed in [Sca02])

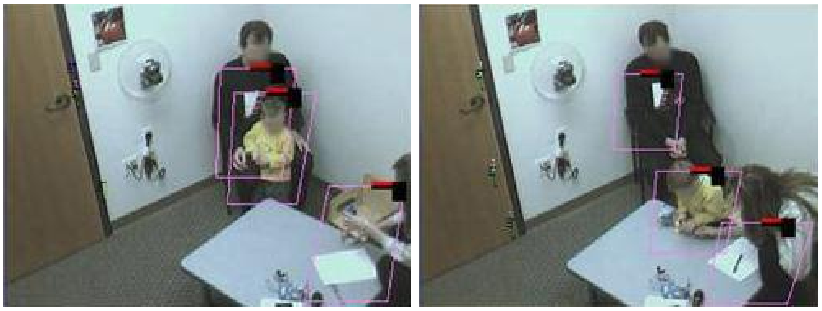

Fig. 3. Two tracking images from the left camera of a calibrated stereo cameras rig. Individuals are tracked as they move throughout one of our clinical evaluation rooms during an interview session.

is then used to predict relative motion and identify motion trajectories of individuals. Figure 3 shows two images obtained during a standard diagnostic interview. Note that the recording and computation performed by this system impact the diagnostic interview no more than other video recording devices would.

Our initial experiments with this technique were able to successfully track the positions of toddlers during a standard behavioral assessment. This included instances when individuals left the field of view, were occluded completely by objects or other individuals, and changed postures dramatically (moving from a standing position to crouched in a corner to lying down horizontally). However, the range of motion of these children during the assessment is limited; in order to allow the completion of the evaluation, both the parent and the experimenter act to try to keep the child on-task at the table. We are currently deploying this system in a larger space that is used for social skills training sessions for adolescents with autism. We anticipate that the data obtained in this environment will be more indicative of natural social response.

Vocal prosody

Individuals with autism often have difficulty both generating and recognizing vocal prosody and intonation [SPM$^+$01]. (Simply put, prosody refers to not what is said, but how it is said.) There are no standardized measures of prosody in the clinical literature [Pau05], and the only research instrument available [SKR90] is very laborious and thus seldom used in diagnostic evaluation or experimental studies.

We recently constructed a multi-stage Bayesian classifier capable of distinguishing between five categories of prosodic speech (prohibition, approval, soothing, attentional bids, and neutral utterances) with an accuracy of more than 75% on a difficult set of vocal samples taken from typical adults (both male and female). In comparison, human judges were able to correctly classify utterances 90% of the time within this data set [RMS04]. To develop this technique to the point where it can be used as a diagnostic tool in the clinic will require us to develop two different forms of classifier based on our initial system design. First, we would like to have a very selective system with a low false-positive rate that can be used continuously on microphone arrays in our clinical evaluation rooms. This system would mark portions of the recorded audio/video streams when extreme prosodic utterances occurred. Second, a system that can be used under more controlled conditions (during experimental protocols) would be developed that was more sensitive to prosodic cues but would suffer from higher rates of both false positives and false negatives. Both of these systems can be obtained by altering a small set of parameters in our initial multi-stage classifier design, but these systems have yet to be evaluated in the clinic.

4.2 Interactive Social Cue Measurement

While there is a vast array of information that can be obtained by passive sensing technologies, the use of interactive robots provides unique opportunities for examining social responses in a level of detail that has not previously been available. These advantages include the following:

1. By generating a social press designed to elicit a particular social response from the subject, the interactive system can selectively probe for information on low-occurrence social behaviors or on behaviors that may not easily emerge in diagnostic sessions in the clinic.
2. The robot provides a repeatable, standardized stimulus and recording methodology. Because both the production and recognition are free from subjective bias, the process of comparing data on social responses between individuals or for a single individual across time will be greatly simplified. As a result, the interactive system may prove to be a useful evaluation tool in measuring the success of therapeutic programs and may provide a standard for reporting social abilities within the autism literature.
3. Because a robotic system can generate social cues and record measurements autonomously, simple interactive toys can be designed to collect data outside of the clinic, effectively increasing both the quantity and quality of data that a clinician can obtain without extensive field work.

We have developed one simple device, called Playtest (see Figure 4), for determining auditory preferences that can be used in the clinic or in the home. When a button is pressed, the device plays one of two audio clips, produces a series of flashing lights to entice attention, and records the time, date, button pressed and audio clip played to non-volatile memory. This device can be sent home with a family to collect information on typical play patterns. This method has been shown to have important diagnostic value [Kli91] since it

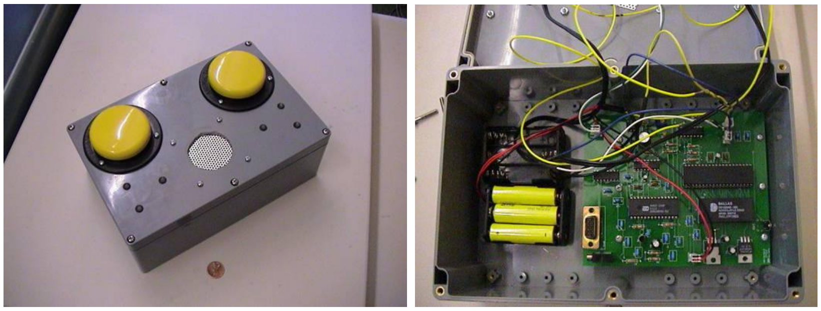

Fig. 4. External view (left) and internal view (right) of the Playtest device for measuring auditory preferences in the home. See section 4.2 for a description.

can measure listening preferences to speech sounds, abnormalities of which are among the most robust predictors of subsequent diagnosis of autism [Lor95].

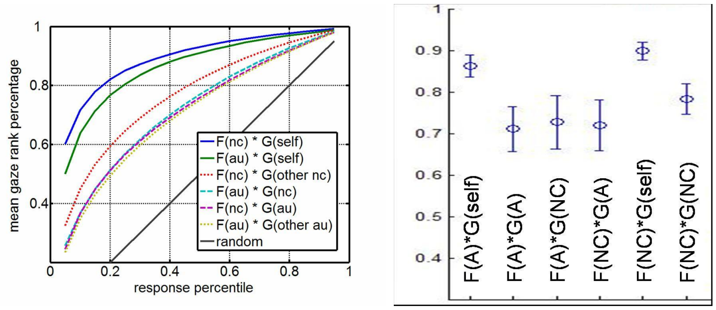

Fig. 5. Results of linear discriminant analysis of autistic (au) and normal (nc) gaze patterns. Linear filters F(x) are trained to reproduce the gaze pattern G(x) of each individual x. Filters can then be applied to predict the gaze patterns of any other individual. For example, F(A)*G(self) indicates a filter trained on an individual with autism is tested on that same individual while F(NC)*G(A) indicates a filter trained on a control individual is tested on data from an individual with autism. At left, the mean performance of this data (y-axis) is a function of the response percentile of individual pairings. At right, significant differences (all p¡0.01 for a two-tailed t-test) are seen between the following classes: (1) F(NC)*G(self), (2) F(A)*G(self), (3) F(NC)*G(NC), and (4) the three other conditions. See section 5 for a discussion.

5 Robots as tools of understanding

The fine-grained analysis of social capabilities that result from work on therapeutic and diagnostic applications have the potential to enhance our understanding of autistic disorders. We have already encountered one example of this potential in our pilot studies of gaze detection. Based on our earlier observations on the differences in gaze direction between typically developing individuals and individuals with autism and in response to our need to characterize potential looking patterns for a robot, we have begun to generate predictive models that show not only the focus of an individual's gaze but also provides an explanation of why they choose to look at particular locations. A simple classifier (a linear discriminant) was trained to replicate the gaze patterns of a particular individual (see Figure 5). The performance of this predictor for a single frame is evaluated by having the filter rank-order each location in the image and selecting the rank of the location actually chosen by a particular individual. Thus, random performance across a sequence of images results in a median rank score of 50th percentile, while perfect performance would result in a median rank score of 1.0 (100th percentile). Trained filters predict the gaze location of the individual they were trained upon with good accuracy (median rank scores of 90th -92nd percentile). By applying a filter trained on one individual to predict the data of a second individual, we can evaluate the similarity of the underlying visual search methods used by each individual. In a pilot experiment with this technique, typically developing individuals were found to all use similar strategies (median rank score in the 86th percentile). Significantly, autistic individuals failed to show similar visual search strategies both among other individuals with autism (73rd percentile) and among the typically developing population (72nd percentile). Filters trained on our control population were similarly unsuccessful at predicting the gaze patterns of individuals with autism (71st percentile). These preliminary results suggest that while our control population all used some of the same visual search strategies, individuals with autism were both not consistently using the same strategies as the control population nor were they using the strategies that other individuals with autism used.

6 Acknowledgments

This work would not be possible without the patience and innovation of Ami Klin, Fred Volkmar, Kasia Chawarska, Warren Jones, and their entire team at the Child Study Center. Ganghua Sun, Fred Shic, Liz Darbie, Avi Robinson-Mosher, Jim Logan, and Reuben Grinberg contributed to some of the preliminary findings reported here. Support for this work was provided by a National Science Foundation CAREER award (#0238334). Some parts of the architecture used in this work was constructed under NSF grants #0205542 (ITR: A Framework for Rapid Development of Reliable Robotics Software) and #0209122 (ITR: Dance, a Programming Language for the Control of Humanoid Robots) and from the DARPA CALO project (BAA 02-21). Support for development of interactive devices for autism diagnosis was provided by the Doris Duke Charitable Foundation.

References

[BEF⁺00] Cynthia Breazeal, Aaron Edsinger, Paul Fitzpatrick, Brian Scassellati, and Paulina Varchavskaia. Social constraints on animate vision. *IEEE Intelligent Systems*, July/August 2000.

[BPR96] A. Bailey, W. Phillips, and W. Rutter. Autism: Towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. *Journal of Child Psychology and Psychiatry*, 37:89–126, 1996.

[Dau00] K. Dautenhahn. Design issues on interactive environments for children with autism. In *Proceedings International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT)*, pages 153–161, 2000.

[fDCP06] Centers for Disease Control and Prevention. http://www.cdc.gov/ncbddd/dd/ddautism.htm, 2006.

[KJS⁺02a] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen. Defining and quantifying the social phenotype in autism. *American Journal of Psychiatry*, 159(6):895–908, 2002.

[KJS⁺02b] A. Klin, W. Jones, R. Schultz, F.R. Volkmar, and D.J. Cohen. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, 59(9):809–816, 2002.

[KJSV03] A. Klin, W. Jones, R.T. Schultz, and F.R. Volkmar. The enactive mind - from actions to cognition: Lessons from autism. *Philosophical Transactions of the Royal Society, Biological Sciences*, 358:345–360, 2003.

[KLCV00] A. Klin, J. Lang, D.V. Cicchetti, and F.R. Volkmar. Interrater reliability of clinical diagnosis and dsm-iv criteria for autistic disorder: Results of the dsm-iv autism field trial. *Journal of Autism and Developmental Disorders*, 30(2):163–167, 2000.

[Kli91] A. Klin. Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders*, 21(1):29–42, 1991.

[KNY] H. Kozima, C. Nakagawa, and H. Yano. Designing a robot for spatio-temporal contingency-detection game. In *International Workshop on Robotic and Virtual Agents in Autism Therapy YEAR = 2000,*.

[Lor95] C. Lord. Follow-up of two-year olds referred for possible autism. *Journal of Child Psychology and Psychiatry*, 46(8):1365–1382, 1995.

[MTT02] F. Michaud and C. Thberge-Turmel. Mobile robotic toys and autism. In Kerstin Dautenhahn, Alan Bond, Lola Canamero, and Bruce Edmonds, editors, *Socially Intelligent Agents - Creating Relationships with Computers and Robots*, pages 125–132. Kluwer Academic Publishers, 2002.

[Mul95] E.M. Mullen. *Mullen Scales of Early Learning: AGS Edition.* American Guidance Service, Circle Pines, MN, 1995.

[Pau05] R. Paul. Communicative competence in individuals with autism. In F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen, editors, *Handbook of Autism and Pervasive Developmental Disorders.* Wiley, New York, 3rd edition edition, 2005.

[RMS04] A. Robinson-Mosher and B. Scassellati. Prosody recognition in male infant-directed speech. In *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2004.

[SBC84] S.S. Sparrow, D. Balla, and D. Cicchetti. *Vineland Adaptive Behavior Scales, Expanded Edition*. American Guidance Service, Circle Pines, MN, 1984.

[Sca01] Brian Scassellati. *Foundations for a Theory of Mind for a Humanoid Robot*. PhD thesis, Massachusetts Institute of Technology, 2001.

[Sca02] Brian Scassellati. Theory of mind for a humanoid robot. *Autonomous Robots*, 12(1):13–24, 2002.

[Sca03] B. Scassellati. Investigating modesl of social development using a humanoid robot. In *Proceedings of the International Joint Conference on Neural Networks*, 2003. CD-ROM.

[SK05] R.T. Schultz and C. Kleinman. Personal communication. 2005.

[SKR90] L. Shriberg, J. Kwiatkowski, and C. Rasmussen. *Prosody-Voice Screening Profile*. Communication Skillbuilders, Tuscon, AZ, 1990.

[SPM$^+$01] L.D. Shriberg, R. Paul, J.L. McSweeny, A. Klin, D.J. Cohen, and F.R. Volkmar. Speech and prosody characteristics of adolescents and adults with high functioning autism and asperger syndrome. *Journal of Speech, Language, and Hearing Research*, 44:1097–1115, 2001.

[SR05] R.T. Schultz and D. Robins. Functional neuroimaging studies of autism. In F.R. Volkmar, R. Paul, A. Klin, and D.J. Cohen, editors, *Handbook of Autism and Pervasive Developmental Disorders*. Wiley, New York, 3rd edition edition, 2005.

[VCK05] F.R. Volkmar, K. Chawarska, and A. Klin. Autism in infancy and early childhood. *Annual Review of Psychology*, 56:315–36, 2005.

[VLB$^+$04] F.R. Volkmar, C. Lord, A. Bailey, R.T. Schultz, and A. Klin. Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*, 45(1):1–36, 2004.

[WD99] I.P. Werry and K. Dautenhahn. Applying mobile robot technology to the rehabilitation of autistic children. In *Proceedings of SIRS '99, Symposium on Intelligent Robotics Systems*, 1999.