Quantitative Metrics of Social Response for Autism Diagnosis*

Brian Scassellati
Department of Computer Science
Yale University
New Haven, CT, 06520
scaz@cs.yale.edu

Abstract - **Social robots recognize and respond to human social cues with appropriate behaviors. The capabilities used to build social robots can be uniquely applied to assist in the diagnosis and treatment of autism, a pervasive developmental disorder which results in selective impairment of social abilities. This paper outlines some of the ways in which social robots can provide unique perspectives to address critical problems in diagnosing autism. We provide preliminary data and observations on how this result can be achieved based on three years of immersion in a clinical research group that performs diagnostic evaluations of more than 130 children per year.**

Index Terms – Social robotics, autism, humanoid robotics

I. Introduction

Robots that recognize and respond to natural human social cues are being designed to fulfill both engineering and scientific goals [1]. Regardless of the reason for building these systems, the development of technology that produces quantitative measurements of social behavior and that can engage in social interactions offers unique tools to the study of autism.

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. In other work, we have identified a range of possible areas in which social robotics technology may help to diagnose, treat, and understand autistic spectrum disorders [2]. This paper concentrates on our attempts to improve the diagnostic standards of autism by using social robots to provide quantitative, objective measurements of social response. While there are currently a handful of projects world-wide that investigate the use of robots as part of a therapeutic regimen for individuals with autism [3,4,5,6], none of these other projects focus on the impact that this technology could have on diagnosis.

Section 2 provides an introduction to autism which highlights some of the difficulties with current diagnostic standards and research techniques. Section 3 describes applications of passive social cue recognition systems and their application as diagnostic tools. Section 4 introduces the use of interactive, social robots which create standardized social presses designed to elicit a particular social response. Section 5 concludes with a discussion on therapeutic and diagnostic possibilities for this work and speculates on how the use of social robots in autism research might lead to a greater understanding of the disorder.

II. Autistic Disorders

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 [7].

* This work is partially supported by grants from the Doris Duke Charitable Foundation and the National Science Foundation under CAREER grant #IIS-0238334.

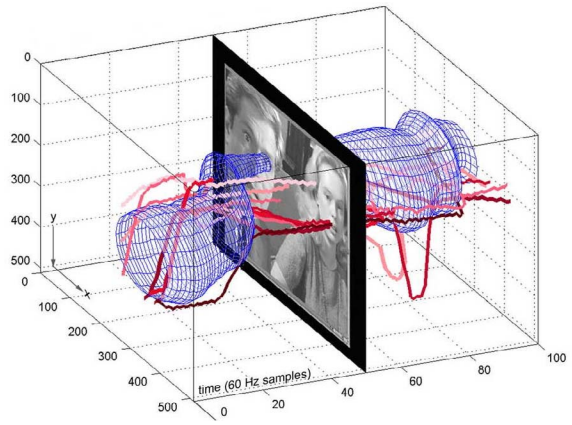

Figure 1: Gaze patterns differ significantly between typical adolescents and adolescents with autism. This spatio-temporal plot shows 10 scan paths of individuals with autism (red lines) and the bounding volume (the blue wireframe) for 10 typical individuals. The typical group shows very similar, structured gaze patterns. The group with autism shows less structure, but is far from random. (Figure reproduced with permission from [15]).

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 [8]; 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, e.g. [9, 10] that are both imprecise and subjective. 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 [11,12].

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 [8,13]. 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 [8]. A more precise characterization and quantification of social dysfunction is required to direct neurobiological research in autism [14,15].

Our goal is to drastically impact the diagnosis of autism by providing technology that can produce quantitative, objective measurements of social response. We believe that this can be accomplished through two methods: (1) through passive observation of the child at play or in interactions with caregivers and clinicians, and (2) through structured interactions with robots that are able to create standardized social "presses" designed to elicit particular social responses.

III. Passive Social Cue Measurement

For three years, we have been developing the technological capabilities to record certain social cues from cameras, microphones, and other sensors that have been installed in parts of our clinic [2]. 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. In this section, we provide three examples of passive social cue measurement that are of particular interest in autism diagnosis: (1) detecting gaze direction, (2) measuring aspects of prosody from human voices, and (3) tracking the position of individuals as they move throughout a room. One of these examples (gaze tracking) is already an integral part of our clinical evaluation and has become a rich source of data on the early social deficits of autism, one has been successfully deployed within the clinic (position tracking) but has not yet been used as a diagnostic tool, and the remaining example (prosody) has yet to be applied directly to the clinic but has been evaluated on typical adults in a laboratory setting. Our current work is on integrating these methods and evaluating their effectiveness as diagnostic tools.

Gaze direction and focus of attention: Most of our current data comes from commercial eye-tracking systems which require subjects to wear a baseball cap with an inertial tracking system and camera/eyepiece combination that 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 1**) [15,16,17]. Fixation time variables predicted level of social competence (e.g., at an average r=.63). [15] 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.

Figure 2: Using a calibrated pair of stereo cameras, we can successfully locate and track individuals as they move throughout one of our clinical evaluation rooms during an interview session. Pairs of stereo images are used to compute a disparity image which is combined with information on color, direction of motion, and background models to identify and track individuals moving throughout a room.

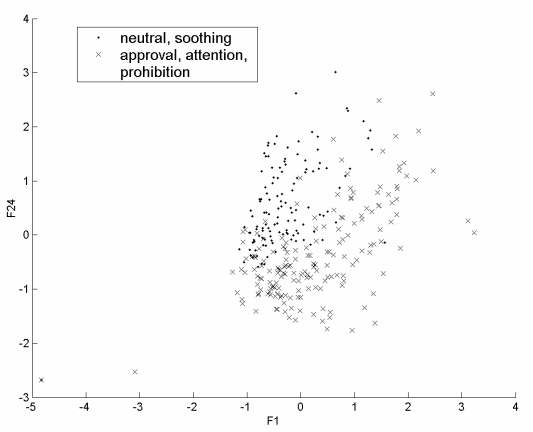

Figure 3: Separation of two features used in the multi-stage prosody classifier (feature F1[mean pitch] vs. feature F24 [the product of mean pitch with energy]) distinguishes between low-energy prosodic categories (neutral and soothing) from high-energy categories (approval, attention, and prohibition).

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 [15].

Gaze detection is our most developed computational technique, but additional refinements are necessary to evaluate the applicability of the technique to diagnosis. A more complex method for analysis, including predictive modeling of gaze patterns, is necessary to automate the currently labor-intensive evaluation of this data (see the Discussion for an example). Furthermore, we need to develop selective stimuli that probe particular abilities rather than relying upon the mixed set of necessary capabilities that occur in our current stimuli.

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 [8].

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 [18]) is then used to predict relative motion and identify motion trajectories of individuals. **Figure 2** 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. (Video recording is routinely used for documentation of diagnostic procedure and reliability assessment).

Our initial experiments with this technique were able to successfully track the positions of toddlers during a standard behavioural 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 [19]. (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 [20], and the only research instrument available [21] 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

Figure 4: ESRA (left) and Playtest (right), described in section IV.

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 (Robinson-Mosher and Scassellati, 2004). **Figure 3** shows one of the feature pairings used in an initial stage of this classifier.

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 very 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.

IV. 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. 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.
2. 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.
3. Since the behavior of the robot can be decomposed arbitrarily, turning off some behaviors while leaving others intact, we can selectively probe responses to individual interaction variables, sometimes in combinations that cannot be performed by humans.
4. 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.
5. Robots generate a high degree of motivation and engagement in subjects, including subjects who are unlikely or unwilling to interact socially with human experimenters. This result is confirmed both by our research findings (below) and by other research groups [4,5,6].
6. Interactive systems may provide a type of social "crutch," that is, the social behavior of the system can be slowly incremented by adding layers of complexity at a rate that matches the (hopefully) increasing capabilities of an individual with autism. Because these robots are often good motivators, the ability to continually update the challenge posed by the system allows the clinician to selectively target individual social skills. This may lead to a form of incremental therapy and training that is unique. 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 [22].

Most robotic systems designed in research settings are not directly suitable for use in the clinic because they were not designed with appropriate safety systems (mechanical, electrical, and computational) and because they are too expensive and too fragile to use realistically with developmentally disabled children. However, they do provide valuable technology that can be used in simpler systems. We report here on two devices that have been successfully used within our clinical environment: a take-home device for measuring auditory preferences and a simple facial robot designed to test the impact of contingency (**Figure 4**).

Playtest is a device 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 [23] since it can measure listening preferences to speech sounds, abnormalities of which are among the most robust predictors of subsequent diagnosis of autism [24].

We have conducted pilot experiments with the simple ESRA robot to demonstrate the feasibility of using anthropomorphic robots in a clinical setting. ESRA is an inexpensive commercial product which generates a small set of facial expressions using five servos. ESRA was programmed to perform a short "script" that was roughly 2 minutes long which included both a set of actions and an accompanying audio file that was played from speakers hidden near the robot. 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. The robot was then activated by an experimenter who was observing the interaction through a one-way mirror. The script started with the robot "waking up", asking a few questions of the child, and then falling back "asleep". The robot had no sensory capabilities and did not respond to anything that the child did. Even with the extremely limited capabilities of ESRA, the robot was well tolerated by all of the children and many of them (including some of those within the autism spectrum) seemed to thoroughly enjoy the session. Children were universally engaged with the robot, and often spent the majority of the session touching the robot, vocalizing at the robot, and smiling at the robot. It is worth noting that for many of the ASD children in this pilot study, these positive proto-social behaviors are rarely seen in a naturalistic context.

V. Discussion

The primary goal of this work though is to substantively enhance methods of diagnosis and treatment of autism. While this 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. Objective measurements of social behavior that can be taken using the same set of stimuli (either static or interaction-based) across individuals and across time would allow for a standardized measurements for comparison of populations of individuals and to track the change in social skill performance of a single individual. Because some of

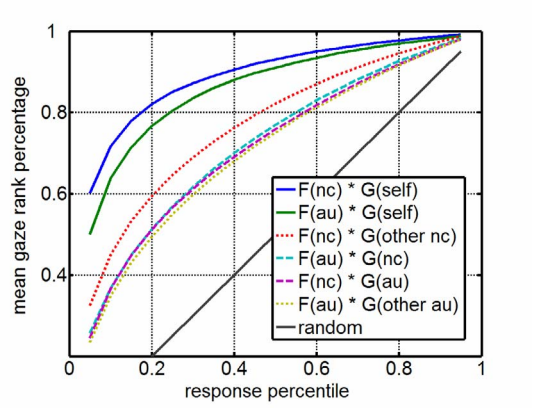

Figure 5: Results of linear discriminant analysis of autistic (au) and typical (nc) gaze patterns. Linear filters F(x) are trained to reproduce the gaze pattern G(x) of each individual x. Filters can then be applied to predict the gaze patterns of any other individual. For example, F(A)*G(self) indicates 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. The mean performance of this data (y-axis) is a function of the response percentile of individual pairings. 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.

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. A secondary impact of this work within the autism research community is the exploration of a novel therapeutic regime using social robots that can scale to accommodate the (hopefully growing) capabilities of an individual. The idea of the robot as a "social crutch" is a fundamentally different methodology than the behavioral training for a static objective performed by other robotics groups.

It is also possible that the fine-grained analysis of social capabilities will 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 very high 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. While our focus is on the applications of these technologies to diagnosis, this example demonstrates that this fine-grained analysis of social behavior may uncover interesting observations about the nature of autism.

ACKNOWLEDGMENTS

This work would not be possible without the patience and innovation of Ami Klin, Fred Volkmar, 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.

REFERENCES

[1] Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003) "A survey of socially interactive robots." *Robotics and Autonomous Systems*, 42:143-166.

[2] Scassellati, B. "How social robots will help us to diagnose, treat, and understand autism," submitted to *ISRR-05*.

[3] Werry, I.P., Dautenhahn, K. (1999). "Applying Mobile Robot Technology to the Rehabilitation of Autistic Children." *Proceedings of SIRS '99, Symposium on Intelligent Robotics Systems*, 20-23 July 1999.

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

[5] Dautenhahn, K. (2000) "Design Issues on Interactive Environments for Children with Autism." *Proceedings International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT),* p. 153-161.

[6] Kozima, H., Nakagawa, C., and Yano, H. "Designing a Robot for Spatio-Temporal Contingency-Detection Game," *International Workshop on Robotic and Virtual Agents in Autism Therapy (Hospital la Salpetriere, Paris, France)*, 2002.

[7] Centers for Disease Control and Prevention, National Center on Birth Defects and Developmental Disabilities (NCBDDD), http://www.cdc.gov/ncbddd/dd/ddautism.htm, May 3, 2005.

[8] Volkmar, F.R., Lord, C., Bailey, A., Schultz, R.T., & Klin, A. (2004). Autism and pervasive developmental disorders. Journal of Child Psychology and Psychiatry, 45(1), 1-36.

[9] Sparrow, S.S., Balla, D., & Cicchetti, D. (1984). *Vineland Adaptive Behavior Scales, Expanded Edition.* Circle Pines, MN: American Guidance Service.

[10] Mullen, E.M. (1995) *Mullen Scales of Early Learning: AGS Edition.* Circle Pines, MN: AGS.Mullen, 1995;

[11] Klin, A., Lang, J., Cicchetti, D.V., & Volkmar, F.R. (2000). "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.

[12] Volkmar, F.R., Chawarska, K., & Klin, A. (2005). "Autism in infancy and early childhood." In A. Kazdin (Ed.), *Annual Review of Psychology*, 56, 315-36.

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

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

[15] Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002a). "Defining and quantifying the social phenotype in autism." *American Journal of Psychiatry*, 159(6), 895-908.

[16] Klin, A., Jones, W., Schultz, R., Volkmar, F.R., Cohen, D.J. (2002b). "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.

[17] Klin, A., Jones, W., Schultz, R.T., & Volkmar, F.R. (2003). "The Enactive Mind – from actions to cognition: Lessons from autism." *Philosophical Transactions of the Royal Society, Biological Sciences*, 358, 345-360.

[18] Scassellati, B. (2002). "Theory of mind for a humanoid robot." *Autonomous Robots*, 12, 13-24.

[19] Shriberg, L.D., Paul, R., McSweeny, J.L., Klin, A., Cohen, D.J., & Volkmar, F.R. (2001). "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.

[20] Paul, R. (2005). "Communicative competence in individuals with autism". In F.R. Volkmar, R. Paul, A. Klin, & D.J. Cohen (Eds.), *Handbook of Autism and Pervasive Developmental Disorders*, *3rd edition*. New York: Wiley.

[21] Shriberg, L. Kwiatkowski, J., & Rasmussen, C. (1990). *Prosody-Voice Screening Profile*. Tuscon, AZ: Communication Skillbuilders.

[22] Schultz, R.T., & Kleinman, C. (2005). Personal communication.

[23] Klin, A. (1991). "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.

[24] Lord, C. (1995). "Follow-up of two-year olds referred for possible autism." *Journal of Child Psychology and Psychiatry*, 36(8), 1365-1382.