Acoustic-Prosodic, Turn-taking, and Language Cues in Child-Psychologist Interactions for Varying Social Demand

Daniel Bone[1], *Chi-Chun Lee*[1], *Theodora Chaspari*[1], *Matthew P. Black*[1], *Marian E. Williams*[2], *Sungbok Lee*[1], *Pat Levitt*[3], *Shrikanth Narayanan*[1]

[1]Signal Analysis and Interpretation Laboratory (SAIL), USC, Los Angeles, CA, USA
[2]University Center for Excellence in Developmental Disabilities, Keck School of Medicine of USC
[3]Zilka Neurogenic Institute, Keck School of Medicine of USC
http://sail.usc.edu

Abstract

Impaired social communication and social reciprocity are the primary phenotypic distinctions between autism spectrum disorders (ASD) and other developmental disorders. We investigate quantitative conversational cues in child-psychologist interactions using acoustic-prosodic, turn-taking, and language features. Results indicate the conversational quality degraded for children with higher ASD severity, as the child exhibited difficulties conversing and the psychologist varied her speech and language strategies to engage the child. When interacting with children with increasing ASD severity, the psychologist exhibited higher prosodic variability, increased pausing, more speech, atypical voice quality, and less use of conventional conversational cue such as assents and non-fluencies. Children with increasing ASD severity spoke less, spoke slower, responded later, had more variable prosody, and used personal pronouns, affect language, and fillers less often. We also investigated the predictive power of features from interaction subtasks with varying social demands placed on the child. We found that acoustic prosodic and turn-taking features were more predictive during higher social demand tasks, and that the most predictive features vary with context of interaction. We also observed that psychologist language features may be robust to the amount of speech in a subtask, showing significance even when the child is participating in minimal-speech, low social-demand tasks.

Index Terms: autism spectrum disorders, atypical prosody, social reciprocity, turn-taking, language cues

1. Introduction

Autism spectrum disorders (ASD) are developmental disorders characterized by social communication deficits, social impairments, and the presence of restricted, repetitive and/or stereotyped behaviors [1]. The ASD population is large– 1 in 88 people [2]– and heterogeneous. Autism diagnosis is based on clinical judgment, usually including ratings of the child's social behavior during standardized, semi-structured interactions. For those individuals with ASD who are verbally fluent, qualitative aspects of their language and communication skills are often markedly atypical.

Social interaction is a multi-layered process which requires a person to continually transmit information through vocal, visual, and gestural cues, even when not the active speaker. For example, turn-taking trends in conversation rely on rules operating at various granularities: prosody, syntax, semantics, and discourse [3]. They reflect intricate aspects of intent and social emotions unfolding in an interaction; hence models of turn-taking are an important ingredient in behavioral signal processing [4], especially in the study of ASD. Prosody, not only a component of turn-taking but a tool to convey pragmatics and affect, is often atypical in verbal individuals with ASD. An impaired theory of mind may underlie core prosodic deficits; individuals with ASD have been found to have difficulty using prosodic cues to understand complex emotions and mental states in others [5], and challenges in receptive prosody may lead to difficulty recognizing when their own speech patterns are different from others and using appropriate expressive prosody [6].

Although atypical prosody is well-documented in the ASD literature [6], it is not well-defined [7]. Descriptions used for assessment are qualitative and subjective, and the precise cause of a perceived abnormality may be indiscernible. Previous studies on ASD prosody have incorporated structured tasks to target specific aspects of prosody [6], or prosody-voice observation during interview samples [8]. In our previous work [9], we investigated objective acoustic-prosodic features taken from semi-structured conversational samples of the Autism Diagnostic Observation Schedule (ADOS) [10]. We demonstrated prosodic changes in the child's behavior according to the child's rated ASD severity. Furthermore, we found that the psychologist's behavior is reflective of the child's social-communicative behavior– i.e., the psychologist's prosody was more predictive of ASD severity than the child's prosody. In this study, we expand on our previous analysis as outlined below.

Turn-taking and language cues provide an enhanced depiction of communication style and quality beyond prosodic features. Researchers have initiated large-sample, computational study of language and turn-taking in ASD. Studies have considered, for example: pauses, fillers, and discourse markers [11]; semantic and pragmatic errors [12]; and language referencing internal states [13]. In addition to prosody, in this work, we incorporate turn-taking and language cues that may capture aspects of reciprocal social communication in child-psychologist interaction.

Furthermore, we extend our study of speech and language cues to the entire 30-60 minute ADOS session data. In our previous study, we focused only on two similar interview-style subtasks that allowed us to examine expressive prosody in a semi-structured context. Given the varying social demands across the subtasks of ADOS, our analysis also considers their effect on the nature of the communication cues. A complex interplay between the social, affective, and cognitive demands has been hypothesized to influence some of the observed variability in the communication patterns in ASD [10, 14]. The investigation of spontaneous speech acoustic-prosodic, turn-taking, and language cues of both child and psychologist during interactions of varying social demand provides insights into dyadic interactions involving children with ASD.

2. Experimental Design

2.1. The USC CARE Corpus

The USC CARE Corpus [15] consists of spontaneous child-psychologist interactions during administration of the Autism Diagnostic Observation Schedule (ADOS) to youth previously diagnosed with autism. Demographic information for the participants is given in Table 1. All ADOS sessions were audio-video recorded using 2 HD cameras and 2 high-quality far-field microphones. Each session is lexically transcribed based on the SALT transcription manual [16].

The current analysis focuses on interactions from 29 children who were administered the ADOS Module 3, chosen by the psychologist because the children were judged to be verbally fluent. All subjects participated in the ADOS in English. Bilingual (English and Spanish) participants were evaluated by bilingual psychologists. Therefore, small portions of the discourse may be in Spanish; those portions are disregarded. One subject was excluded due to a primarily Spanish discourse.

This study investigates the psychologist's behavior in addition to the child's. Three licensed, research-certified psychologists administered the ADOS for this study. All three women had extensive clinical experience with ASD children. Two of the three psychologists were bilingual in English and Spanish.

Table 1: *Demographic statistics of the 29 recorded children in this study that were administered Module 3 of the ADOS.*

Category	Count/Statistic
Age (years)	mean: 10.0, std. dev.: 2.6, range: 5.8-15.0
Gender	male: 23, female: 6
Native language	Spanish: 9, English: 10, Sp&Eng: 4, unk: 6
Ethnicity	Hispanic: 20, White/+Other: 8, AF-AM: 1
ADOS module	#3: 29
ADOS diagnosis	autism: 18, ASD: 5, below ASD cutoffs: 6

2.2. ASD Severity

The ADOS Module 3 includes 28 codes scored by the examiner in the domains of Social Interaction, Communication, and Restricted, Repetitive Behaviors. A subset of the codes are included in the algorithm. For this analysis, the revised algorithms [17] were used because they incorporate more empirical data. The ADOS severity score is a transformation of the ADOS total for module and age that ranges from 1 to 10, with higher scores indicating higher severity of ASD symptoms [18]. By directly correlating with ASD severity we may gain insight into the way speech and language are related to overall social-communicative abnormality.

2.3. Social Demand

The ADOS consists of 14 activities, or subtasks, (e.g., joint-play, emotions interview, telling a story), each with different social presses and consequently, level of social demand. Seven psychologists with ADOS research training rated, on a 1-to-N scale, each of the subtasks for Social Difficulty (N=5), Cognitive Difficulty (N=5), Naturalness (N=3), and the Amount of Speech Required (N=3). Ratings were z-normalized per rater and averaged.

We concentrate our analysis on social demand (difficulty). The description for this rating is, "How much does the activity require the child to interact socially (verbally or non-verbally) with the psychologist?" Inter-rater agreement for social demand was moderate, *intra-class correlation*$=0.54$. The 14 subtasks were separated into high (5), medium (6), and low (3) social demand based on the distribution. Social demand was correlated with ratings for cognitive demand ($r_s(14)=0.80$), naturalness ($r_s(14)=0.63$), and amount of speech required ($r_s(14)=0.82$).

2.4. Data Pre-processing

Before features are extracted, automatic forced-alignment of the entire session audio to transcription is performed using adult and child acoustic models that were trained on a sample of the corpus using the IBM Attila toolkit [19]. The five-minute samples of interaction used in our previous study [9] were re-transcribed and are of a higher quality, and thus were used for training. Additionally, subtask start and end times were manually marked. Segments of interaction involving Spanish were disregarded. Prior to deriving language measures, inaudible speech and Spanish dialogue were removed. This did not affect our analysis since the language features used are percentages.

2.5. Acoustic-Prosodic and Turn-taking Features

We computed 16 prosodic descriptors of social interaction based on our previous work [9]. The features relate to speech intonation, volume, rate, and voice quality.

Intonation and volume contours (pitch and intensity) are extracted on turn-end words using Praat [20]. Log-pitch and intensity are mean subtracted per session, then second-order polynomial parameterization is computed. Median and inter-quartile ratio (IQR) of slope and curvature, as well as raw mean are included as features (10 features).

For the remaining features, median values over all words in each subtask are calculated for robustness. Voice quality is computed through jitter, shimmer, and harmonics-to-noise using Praat with a 40 ms window and 10 ms shift (3 features). Our previous study found that jitter– a breathy/rough/hoarse voice quality correlate [21]– increased for both interactants when ASD severity increased [9]. Speaking rate (SR) is separated into 3 features: SR (#words-per-utterance/time-for-utterance), articulatory SR (syl/s), and intra-turn silence duration.

Four turn-taking features describe the conversation style of each participant: speech %, silence %, overlap %, and median latency. Overlap % is the proportion of time a participant interrupts the interlocutor. Median latency is time taken to speak after the previous speaker ends their turn. Silence % is the same for both speakers.

2.6. Language Features

Language usage potentially related to ASD is quantified using the Linguistic Inquiry and Word Count (LIWC) toolbox [22]. LIWC software has previously been used to study language in ASD [13, 23]. The features are: (1) words per sentence (WPS)– a rough approximation of mean-length-of-utterance (MLU); (2) first-person, singular pronouns (I, me, mine); (3-5) positive emotion, negative emotion, and affect (positive or negative) language; (6-8) assents (OK, yes), non-fluencies (hm, umm), and fillers (I mean, you know). Language features are percentages normalized by the total number of words spoken (besides WPS).

Table 2: *Correlations of session-level features with ADOS severity. Note: [†,*,**] ≡ α=[0.10,0.05,0.01].*

Trend with Severity	Psych Feature	Sp. ρ
more positive	pitch slope	$+0.32^{\dagger}$
more variable	pitch curvature	$+0.33^{\dagger}$
more positive	intensity curvature	$+0.31^{\dagger}$
more variable	intensity curvature	$+0.51^{**}$
increased	articulatory SR	$+0.38^{*}$
increased	intra-turn silence	$+0.32^{\dagger}$
decreased	harmonics-to-noise	-0.47^{**}
increased	speech %	$+0.54^{**}$
increased	personal pronouns	$+0.38^{*}$
decreased	assent lang.	-0.48^{**}
decreased	non-fluent lang.	-0.48^{**}

Trend with Severity	Child Feature	Sp. ρ
more negative	pitch curvature	-0.56^{**}
more variable	pitch curvature	$+0.45^{*}$
more variable	intensity curvature	$+0.43^{*}$
decreased	articulatory SR	$-0.34^{\dagger}$
increased	latency	$+0.34^{\dagger}$
decreased	speech %	$-0.36^{\dagger}$
decreased	words per sentence	-0.42^{*}
decreased	personal pronouns	-0.40^{*}
decreased	affect lang.	-0.40^{*}
decreased	fillers	-0.41^{*}

3. Analysis of Acoustic-Prosodic, Turn-taking, and Language Features

The speech and language features of child and psychologist vary throughout the interaction; recall that each ADOS session comprises several subtasks. In this section, we study the general communicative behavior of both participants across the session. Features are first calculated per subtask, but these may have high variability if the sample of communication is too small. We quantify session-level behavior as the median of all subtask-level values of each feature. Additionally, in this analysis we exclude three subtasks for which "Amount of Speech Required" was rated lowest, so as to reduce variance in subtask-level speech and language features. The excluded subtasks are Construction, Make-Believe Play, and Break.

In sections 3.1 and 3.2, the significant pair-wise correlations between the 28 child and psychologist speech and language features and ASD severity are examined and interpreted (Table 2). A very inclusive significance level is chosen (α=0.10). Although many correlations reach higher levels of significance, certain results should be interpreted carefully.

3.1. Acoustic-Prosodic Variation

We first consider pitch and intensity contour functionals of either participant. As a child's ASD severity increases, the psychologist and the child tend to speak with more variable prosody. In particular, the psychologist's pitch and intensity curvature, and the child's pitch curvature have increased inter-quartile ratio (IQR). Additionally, the psychologist has increased pitch slope. The psychologist may be exaggerating intonation in order to convey more affect in an attempt to engage the child and elicit a desired response. The child's pitch curvature variability is less intuitive, but it may result from poor control of pitch dynamics or increased arousal. Furthermore, children with higher ASD severity showed a strong tendency to exhibit reduced pitch slope; this behavior may relate to the common perceptions of monotonous intonation in children with ASD, as we suggested in our previous study [9].

The psychologist's harmonics-to-noise ratio (HNR) tends to decrease when interacting with a child who has higher ASD severity. HNR can relate to perceptions of breathy, rough, or hoarse voice [21]– generally an atypical voice quality. In our previous work, we found both participants to have increased atypical voice quality for sessions with children of higher symptom severity [9].

3.2. Conversational Depiction

An image of conversational style and quality as a function of ASD severity emerges from the correlations in Table 2. As ASD severity increases, the psychologist talks more while the child talks less. This may suggest that the child with more severe symptoms is more difficult to engage and less comfortable talking, and the psychologist is trying various speech strategies to elicit engagement. The psychologist articulates quicker, while the child articulates slower. Furthermore, the psychologist has increased intra-turn silence, while the child has increased latency to respond. The psychologist may wait for the child to respond (intra-turn silence) given the child's bias toward delayed response; but, not seeing the desired reaction from the child, the psychologist again prompts or proceeds.

We also consider the language features. Children with more severe ASD tended to use personal pronouns less often. Differences have previously been observed in the production of 'me' and 'you' between autistic and typically developing children [24]. In the ADOS, the children are often asked personal questions which can make the children uncomfortable. We may suspect that children who have an aversion to making personal assertions, especially those with higher ASD severity, will use personal pronouns less often. This finding corroborates a study that found children with autism responding to personal or emotional questions tended to give de-personalized responses and avoid use of the word 'I' [25]. Furthermore, children with higher symptom severity produce less affect language (positive or negative emotional words). Children with higher ASD severity may be freely offering personal information less often; the ADOS code "Offers Information" scores such socially atypical behavior. Additionally, since the psychologist tends to use personal pronouns more frequently, the psychologist may be making overt attempts to engage the child in reciprocal social communication about the psychologist's experiences.

The semi-structured interaction appears less conversational in the children who exhibit more severe social-communicative difficulties. The psychologist produces less assent (OK, yes) language to back-channel the child's comments. The psychologist also uses less non-fluent language (hm, umm), which could indicate the psychologist is making an effort to be very direct and clear in her communication attempts. Additionally, children with higher ASD severity tend to use fewer fillers/discourse-markers (I mean, you know). Fillers can serve turn-taking functions such as pacing and stalling. Previous results from Heeman et al. (2010) also found that children with ASD used fewer fillers [11], although they were surveying use of 'uh' and 'um' which are classified as non-fluencies in our study. Lastly, the child with higher ASD severity exhibits decreased words-per-sentence. This indicates that children with higher ASD severity speak fewer words at a time, in addition to speaking a smaller percentage of the time.

Table 3: *Correlations of prosodic and language feature sets predictions with ADOS severity over varying social demands. Note: [†,*,**,***] ≡ α=[0.10,0.05,0.01,0.001]. P≡Psychologist's features, C≡Child's features.*

	High Social Demand		Medium Social Demand		Low Social Demand	
	$Sp.\ \rho$	Chosen Features	$Sp.\ \rho$	Chosen Features	$Sp.\ \rho$	Chosen Features
Psych Acoustics	**+0.43***	1. P engy curvature IQR 2. P harmonics-to-noise 3. P silence %	**+0.46***	1. P intra-turn sil DUR 2. P speech %	+0.20	1. P engy curvature MED 2. P intra-turn sil DUR 3. P speech %
Child Acoustics	**+0.76*****	1. C pitch curvature MED 2. C pitch curvature IQR 3. C articulation rate	+0.17	1. C speaking rate 2. C overlap % 3. C intensity MED	+0.02	1. C pitch curvature IQR 2. C intra-turn sil DUR
Both Acoustics	**+0.50*****	1. C pitch curvature IQR 2. P engy curvature IQR 3. C intra-turn sil DUR	+0.29	1. P intra-turn sil DUR 2. P speech %	−0.11	1. C pitch curvature IQR 2. P intra-turn sil DUR 3. P engy curvature MED
Psych Lang.	**+0.49****	1. P affect 2. P non-fluencies	+0.10	1. P non-fluencies 2. P affect	**+0.39***	1. P non-fluencies
Child Lang.	−0.01	1. C filler	+0.35†	1. C filler	−0.97	*none*
Both Lang.	+0.27	1. P affect 2. P non-fluencies	**+0.46***	1. C. filler 2. P non-fluencies	+0.27	1. P non-fluencies

4. Prediction over Varying Social Demand

The severity of social-communicative deficits for children with ASD may not fully emerge until a certain level of social demand is reached. In this section, we examine speech and language features from the child and psychologist in subtasks with high, medium, and low social demand through a predictive modeling task. The high social demand tasks are: Emotions, Loneliness, Social Difficulties and Annoyance, Joint Interactive Play, and Cartoons; medium: Friends and Marriage, Description of a Picture, Creating a Story, Demonstration Task, Telling a Story from a Book, Conversation and Reporting; low: Construction, Make-Believe Play, and Break. Social demand group-level features are the medians of those from corresponding subtasks.

Multiple linear regression with forward-backward feature-selection is performed using a leave-one-session-out framework with two layers, one for prediction and another for parameter tuning. Spearman's rank correlation coefficient is used for analysis and tuning. Acoustic prosody and turn-taking cues were used separately from language descriptors. These were divided in this analysis because we found over-fitting often occurred due to the large size of our initial feature set. Features chosen in at least 50% of the cross-folds are presented in Table 3.

4.1. Acoustic-Prosodic and Turn-taking Features

Interestingly, both psychologist and child acoustics are significantly predictive of ASD severity in the subtasks with high social demand. Generally, prosodic curvature variability, voice quality, articulatory speaking rate, and pausing are selected for prediction. We find that constraining the range of subtasks leads to diverse modeling of appropriate social functioning– the most-often chosen features for prediction vary between conditions.

We also consider medium and low social demand tasks. The psychologist's features are only predictive in the medium social demand subtasks. The psychologist's intra-turn silence (pausing) and speaking percentage are selected for prediction. Low social demand subtasks are sparse in expected speech content from children, explaining why no significant prediction occurs.

4.2. Language Features

The selected language features show some success in prediction of ASD severity, but generally less so than the acoustic features. Only three features are chosen for prediction in any group: psychologist affect and non-fluencies, and child fillers. We note that after separating high and medium social demand, psychologist use of affect language becomes informative, whereas it was not a significant correlate with ASD severity before.

The psychologist's language features are significantly predictive of ASD severity in high social demand activities, but the child's are not. The child's features are significantly predictive in medium social demand, but the psychologist's are not; however, the addition of the psychologist's features leads to statistically significant prediction. The trend of increased prediction with increased social demand does not seem as pronounced for language cues. Interestingly, the psychologist's features are significantly predictive in low social demand, although no features of the child are selected for prediction– underscoring the salience of the psychologist's language features even when the child receives few presses for interaction.

5. Conclusion and Future Work

In this work, we demonstrated that prosodic, turn-taking, and language features taken during child-psychologist interactions are indicative of the degradation in conversational quality for children with greater severity of ASD symptoms. We modeled not only the child's behavior, but also the psychologist's, since the psychologist is both interlocutor and evaluator. In particular, we found that as ASD severity increases, the psychologist varies her speech and language strategies in attempts to engage the child in social interactions, while children with more severe ASD speak less and use fewer affect words and personal pronouns. Additionally, we found greater predictive power for ASD severity in subtasks with high social demand, while the psychologist's language cues were predictive even during minimal-speech, low social-demand tasks.

Further experiments will be conducted on larger datasets which include data from typically developing children to provide a normative context. Normative data can provide a baseline for expected typical behavior, allowing for greater precision and detail in modeling interaction. Additional topics such as at which point the psychologist makes a decision, the global versus local nature of deficits, and the modeling potential of unsupervised behavioral signal [26, 27] will be examined.

6. Acknowledgements

This research was supported by funds from NSF and NIH. Special thanks to Sylvia Acosta and Irina Zamora for administering the ADOS and Marcia Higareda for recruiting efforts.

7. References

[1] *Diagnostic and Statistical Manual of Mental Disorder, Ed. 4 text revision*, American Psychiatric Assoc., Washington D.C., 2000.

[2] J. Baio, "Prevalence of autism spectrum disorders: Autism and developmental disabilities monitoring network, 14 sites, united states, 2008. morbidity and mortality weekly report. surveillance summaries. volume 61, number 3." *Centers for Disease Control and Prevention*, 2012.

[3] A. Raux and M. Eskenazi, "A finite-state turn-taking model for spoken dialog systems," in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2009, pp. 629–637.

[4] S. Narayanan and P. G. Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," *Proceedings of the IEEE*, vol. PP, no. 99, pp. 1–31, 2013.

[5] O. Golan, S. Baron-Cohen, J. J. Hill, and M. Rutherford, "The reading the mind in the voicetest-revised: a study of complex emotion recognition in adults with and without autism spectrum conditions," *Journal of autism and developmental disorders*, vol. 37, no. 6, pp. 1096–1106, 2007.

[6] S. Peppe, J. McCann, F. Gibbon, A. O'Hare, and M. Rutherford, "Receptive and Expressive Prosodic Ability in Children with High-Functioning Autism," *J. of Speech & Hearing Research*, vol. 50, pp. 1015–1028, 2007.

[7] J. J. Diehl, D. Watson, L. Bennetto, J. McDonough, and C. Gunlogson, "An Acoustic Analysis of Prosody in High-Functioning Autism," *Applied Psycholinguistics*, vol. 30, pp. 385–404, 2009.

[8] R. Paul, L. D. Shriberg, J. McSweeny, D. Cicchetti, A. Klin, and F. Volkmar, "Brief Report: Relations between Prosodic Performance and Communication and Socialization Ratings in High Functioning Speakers with Autism Spectrum Disorders," *J. of Autism and Dev. Dis.*, vol. 35, pp. 861–869, 2005.

[9] D. Bone, M. P. Black, C.-C. Lee, M. E. Williams, P. Levitt, S. Lee, and S. Narayanan, "Spontaneous-speech acoustic-prosodic features of children with autism and the interacting psychologist," in *Proc. Interspeech*, 2012.

[10] C. Lord, S. Risi, L. Lambrecht, E. Cook, B. Leventhal, P. DiLavore, A. Pickles, and M. Rutter, "The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism," *J. of Autism and Dev. Dis.*, vol. 30, pp. 205–223, 2000.

[11] P. A. Heeman, R. Lunsford, E. Selfridge, L. Black, and J. Van Santen, "Autism and interactional aspects of dialogue," in *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, 2010, pp. 249–252.

[12] E. T. Prudhommeaux, B. Roark, L. M. Black, and J. van Santen, "Classification of atypical language in autism," *ACL HLT 2011*, p. 88, 2011.

[13] B. T. Brown, G. Morris, R. E. Nida, and L. Baker-Ward, "Brief report: Making experience personal: Internal states language in the memory narratives of children with and without aspergers disorder," *Journal of autism and developmental disorders*, vol. 42, no. 3, pp. 441–446, 2012.

[14] G. B. Mesibov, E. Schopler, and K. A. Hearsey, "Structured teaching," *Behavioral issues in autism*, pp. 195–207, 1994.

[15] M. P. Black, D. Bone, M. E. Williams, P. Gorrindo, P. Levitt, and S. S. Narayanan, "The USC CARE Corpus: Child-Psychologist Interactions of Children with Autism Spectrum Disorders," in *Proceedings of Interspeech*, 2011.

[16] J. Miller and A. Smith, "Salt transcription manual: Guidelines for transcribing free speech samples," *Unpublished paper, Language Analysis Lab, University of Wisconsin-Madison*, 1983.

[17] K. Gotham, S. Risi, A. Pickles, and C. Lord, "The autism diagnostic observation schedule: revised algorithms for improved diagnostic validity," *Journal of Autism and Developmental Disorders*, vol. 37, no. 4, pp. 613–627, 2007.

[18] K. Gotham, A. Pickles, and C. Lord, "Standardizing ados scores for a measure of severity in autism spectrum disorders," *Journal of autism and developmental disorders*, vol. 39, no. 5, pp. 693–705, 2009.

[19] H. Soltau, G. Saon, and B. Kingsbury, "The ibm attila speech recognition toolkit," in *Spoken Language Technology Workshop (SLT), 2010 IEEE*. IEEE, 2010, pp. 97–102.

[20] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[21] A. McAllister, J. Sundberg, and S. R. Hibi, "Acoustic Measurements and Perceptual Evaluation of Hoarseness in Children's Voices," *Logopedics Phoniatrics Vocology*, vol. 23, 1998.

[22] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, 2001.

[23] T. Chaspari, D. Bone, J. Gibson, C. Lee, and S. S. Narayanan, " Using physiology and language cues for modeling verbal response latencies of children with ASD," in *Proceedings of ICASSP*, 2013.

[24] R. R. Jordan, "An experimental comparison of the understanding and use of speaker-addressee personal pronouns in autistic children," *International Journal of Language & Communication Disorders*, vol. 24, no. 2, pp. 169–179, 1989.

[25] C. A. Baltaxe, "Pragmatic deficits in the language of autistic adolescents," *Journal of Pediatric Psychology*, vol. 2, no. 4, pp. 176–180, 1977.

[26] D. Bone, C.-C. Lee, and S. Narayanan, "A robust unsupervised arousal rating framework using prosody with cross-corpora evaluation," in *Proc. Interspeech*, 2012.

[27] C.-C. Lee, A. Katsamanis, M. P. Black, B. R. Baucom, A. Christensen, P. G. Georgiou, and S. S. Narayanan, "Computing vocal entrainment: A signal-derived pca-based quantification scheme with application to affect analysis in married couple interactions," *Computer Speech & Language*, 2012.