Spontaneous-Speech Acoustic-Prosodic Features of Children with Autism and the Interacting Psychologist

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

Atypical prosody, often reported in children with Autism Spectrum Disorders, is described by a range of qualitative terms that reflect the eccentricities and variability among persons in the spectrum. We investigate various word- and phonetic-level features from spontaneous speech that may quantify the cues reflecting prosody. Furthermore, we introduce the importance of jointly modeling the psychologist's vocal behavior in this dyadic interaction. We demonstrate that acoustic-prosodic features of both participants correlate with the children's rated autism severity. For increasing perceived atypicality, we find children's prosodic features that suggest 'monotonic' speech, variable volume, atypical voice quality, and slower rate of speech. Additionally, we find the psychologist's features inform their perception of a child's atypical behavior– e.g., the psychologist's pitch slope and jitter are increasingly variable and their speech rate generally decreases.

Index Terms: atypical prosody, autism spectrum disorder, intonation, psychologist, voice quality, ADOS

1. Introduction

Social interaction is a process in which participants constantly receive, process, plan, and transmit multi-modal, pragmatic and affective cues. A person who is impaired at any stage in the communicative process may have difficulties in effectively interacting with others.

Autism spectrum disorders (ASD) are developmental disorders that result in impaired social communication and reciprocity, as well as restricted, repetitive, and/or stereotyped behavioral patterns [1]. ASD is considered a spectrum disorder due to the heterogeneity of symptomatology. Recent prevalence studies suggest that as many as 1 out of 110 children has ASD [2].

Atypical prosody, a commonly reported symptom in ASD, is intertwined with impaired social reciprocity. In particular, Theory of Mind purports people with autism have a compromised ability to gauge the mental state of another person [3]– such a deficit will lead to impairments in receptive and expressive prosodic skills (which are correlated [4]), in addition to other knowledge of successful social interaction. Other theories implicate impaired speech planning and motor systems [5]. While there is ample documentation of the presence of atypical and impaired prosody in ASD, a precise characterization is still lacking [4, 6]. A more stratified and objective analysis of the speech properties can help toward a better understanding of the nature of the prosodic deficits.

Qualitative descriptions of atypical prosody are general and contrasting, having few well-defined acoustic correlates. Atypicality has been attributed to, "exaggerated or monotonous intonation, slow syllable-timed speech, fast rate of speech, or an adopted accent different from that of peers" [7]. The Autism Diagnostic Observation Schedule (ADOS) rates atypical prosody as any of the following: slow, rapid, jerky and irregular in rhythm, odd intonation or inappropriate pitch and stress, markedly flat and toneless, or consistently abnormal volume [8].

Toward obtaining a more objective characterization of speech prosody, we employ signal processing methods on spontaneous interactions between a child and a psychologist recorded during administration of the ADOS; this is in contrast to analyzing manual annotations of prosody elicited in targeted speaking tasks [4, 9]. Such structured assessment may not capture the extent to which atypical prosody affects social functioning apart from pragmatic expression. Accordingly, atypicality has also been attributed to a range of distinct acoustic speech properties such as pitch slope [10], breathiness [11], and nasality [12]. Few studies have sought acoustically-derived correlates of speech (e.g., [6, 13]), and even fewer have simultaneously assessed spontaneous speech (e.g., [5]).

In this report, we demonstrate the importance of studying the psychologist's acoustic behavior in addition to the child's acoustics to obtain a more detailed description of the interaction. Our approach is further unique in the domain of ASD research because we analyze automatic signal-derived prosodic measures (automatic given lexical transcription and turn-level alignments). Signal processing techniques have the potential to support researchers and clinicians with quantitative description of qualitative behavioral phenomena, to promote understanding of effective psychological methods, and to further facilitate more precise stratification within this spectrum disorder. This is a primary goal of the emerging field of behavioral signal processing (BSP) [14].

2. Experimental Design

We investigate acoustic-prosodic cues of child and psychologist speech relating to pitch, intensity, duration, and voice quality. Child-psychologist interactions are from the USC CARE Corpus [15].

2.1. The USC CARE Corpus

The USC CARE Corpus is comprised of spontaneous child-psychologist interactions of youth at risk for autism [15] collected in the context of the Autism Diagnostic Observation Schedule (ADOS). The psychologist determines which of four modules to administer depending on the subject's expressive language level and age.

The current analysis focuses on interactions involving 28 children that were administered the ADOS Module 3 (designed for verbally fluent children). Demographics are given in Table 1. Sessions are conducted in English– in which all subjects are fluent. Two subjects (from an original 30) were excluded– one due to lack of vocal activity and another due to a primarily Spanish discourse. The sessions were roughly equally divided amongst three trained psychologists (led by co-author M. E. Williams).

We analyze a subset of the sessions consisting of two subtasks: "Emotions" and "Social Difficulties and Annoyance", chosen since they offer continuous samples of conversational speech. From the subtasks start, we analyze up to five minutes per child (μ=264s, min=101s).

2.2. ADOS Codes of Interest

The ADOS Module 3 consists of 28 codes, not all of which are used in the final ADOS decision. We focus on one 'Atypical Prosody' code and the three totals that are used in ADOS determination. They are: "Speech abnormalities associated with autism" (Atyp. Pros.), and the "Communication" (Comm.), "Social Interaction" (Soc. Int.), and combined "Communication and Social Interaction" (CS&I) Totals.

The 'Atypical Prosody' code quantifies atypical prosody on an integer scale from '0' to '2', with '0' designating appropriate prosody, '1' signifying slight deviations from typicality, and '2' used to report 'clearly abnormal prosody'. Our data comprises 4, 12, and 12 instances of atypical prosody scores '0', '1', and '2', respectively. The ADOS Totals are highly correlated with 'Atypical Prosody' in our data, Spearman's ρ=0.74 (p<10e-6). We correlate with ADOS Totals because atypical prosody is difficult to describe, relying on subjective interpretation of multiple factors, and because the session totals are higher resolution and may be more indicative of global phenomena relating to prosody.

Table 1: *Demographic statistics of the 28 recorded children in this study that were administered Module 3 of the ADOS.*

Category	Count/Statistic
Age (years)	mean: 9.8, std. dev.: 2.5, range: 5.8-14.7
Gender	male: 22, female: 6
Native language	Spanish: 8, English: 9, Sp.&Eng.: 4, unk: 7
Ethnicity	Hispanic/Latino: 20, White/White+Other: 8
ADOS module	#3: 28
ADOS diagnosis	autism: 17, ASD: 5, below ADOS cutoffs: 6

2.3. Spontaneous-Speech Acoustic-Prosodic Features

We extracted 25 prosodic features that address, at the session level, the four areas of prosody described in the ASD literature: intonation, volume, rate, and voice quality. All word-level features are extracted on turn-end words since pitch contours are most perceptually salient at phrase boundaries. We utilize turn-level alignments and lexical transcriptions to provide more reliability in our automatically extracted features. Forced-alignment is performed with context using HTK with adult-models trained on the Wall-Street Journal Corpus and with child-models trained on the CU Kid's Corpus.

Intonation and volume contours (pitch and intensity) are extracted per word using Praat [16]. To remove variability across sessions and speakers, log-pitch and intensity are normalized by subtracting means per speaker, per session [14]. Each contour is bounded in the range [-1,1], then parameterized as as second-order polynomial (curvature, slope, and zero-crossing). Mean (μ) and standard-deviation (σ) functionals are computed, totaling 12 features. Speaking rate and rhythm total 9 features, including: mean and 90% quantile of turn-end and non-turn-end syllabic-speaking rate, means and standard deviations of vowel and consonant durations, and the proportion of vowel speech to total speech. Voice quality is captured by four features: median and inter-quartile ratio (IQR) of jitter and shimmer values. Jitter and shimmer were calculated on extended vowels (at least $3T_0$) using the 'local' method in which pitch period and pitch magnitude are quantized once per pitch period interval (such as in [16]).

3. Analysis of Acoustic-Prosodic Features

In Sections 3.1 and 3.2, the significant pair-wise correlations between the 25 child and psychologist prosodic features and the four considered codes (themselves strongly correlated) are examined and interpreted.

3.1. Child's Prosody

We first consider pitch contour parameterization functionals. Our data shows a medium-strong significant negative correlation (ρ=-0.56) between turn-end pitch-slope and rated atypicality– meaning lower average pitch slope indicates more atypicality. Negative pitch slope (or curvature) at turn-end is often associated with statements, and we noticed that average pitch slope was positive only for children with the least atypical ratings. Although this result should be interpreted with care, it may be that the

Table 2: *Participants' acoustic-prosodic features with significant correlations to ADOS code labels. '+' and '-' denote positive (i.e., increasing feature values and increasing atypicality) and negative correlation. If* **bold** *α=0.01, else α=0.05.*

	Code Label			
	Atyp. Pros.	Comm. Total	Soc.Int. Total	C&SI Total
Child's Acoustic-Prosodic Features				
f0_slope μ	-0.45	$\mathbf{-0.57}$	$\mathbf{-0.50}$	$\mathbf{-0.56}$
f0_curve μ		-0.46	-0.41	-0.45
Int_intercept σ		$+0.39$		
Jitter $median$	$+0.42$	$+0.39$	$+0.41$	$+0.41$
Jitter iqr	$\mathbf{+0.55}$	$+0.47$	$\mathbf{+0.48}$	$\mathbf{+0.50}$
syl_SR-nonBoundary $q_{0.9}$		-0.41		
Psychologist's Acoustic-Prosodic Features				
f0_slope μ		$+0.38$		
f0_intercept σ	$+0.44$	$\mathbf{+0.62}$	$+0.40$	$+0.47$
f0_slope σ		$+0.47$		
f0_curve σ	$+0.42$	$\mathbf{+0.58}$		$+0.39$
Jitter $median$	$\mathbf{+0.53}$	$\mathbf{+0.77}$	$\mathbf{+0.58}$	$\mathbf{+0.69}$
Jitter iqr	$+0.46$	$\mathbf{+0.57}$	$+0.39$	$+0.47$
syl_SR-Boundary $q_{0.9}$		$+0.46$		
syl_SR-nonBoundary $q_{0.9}$			$\mathbf{-0.48}$	-0.43
vowel_dur σ		$\mathbf{+0.59}$		

atypical-rated children possess the reported 'monotone' voice often cited in the literature, or that more typical behavior in these sessions may involve asking questions.

Voice quality descriptions such as 'breathy', 'hoarse', and 'nasal' are frequent for children with ASD. McAllister et al. (1998) found jitter to correlate with breathiness, hoarseness, and roughness, while shimmer correlated with breathiness [17]. In our data, a child's standard deviation of the normalized-intensity intercept at boundary words increases with increasing rated atypicality. This type of global energy variability relates to perceptions of 'consistently abnormal volume', but does not capture voice quality. Voice quality as measured by jitter shows significant positive correlations with all four codes– higher local variability in pitch occurs with increasing perceived atypicality.

The sixth and final correlated children's prosody feature is the 90% quantile syllabic speaking rate of non-turn-end words. This feature can be considered a robust measure of maximum speaking rate. A maxima was desired because it may indicate maximal ability, and other considered features capture rate variability. The results show that the slower a child talks, the more likely they are to be atypical in Communication Total evaluation. The features we have analyzed have meaningful interpretations and show promise for spontaneous-speech acoustic-prosodic measurements of autistic children's speech.

3.2. Psychologist's Prosody

We examined the speech features of the psychologists to determine potential correlates with the ADOS ratings of the children (Table 2). Mean slope shows positive correlations– higher mean pitch slope happens with higher rated atypicality of the child– a trend which is opposite for the same feature of children. Potentially, the psychologist is responding more with statements for children that are less atypical or exaggerating their pitch to engage more atypical children. We also see positive correlations between standard deviations of pitch parameters and rated atypicality– increasing variability accompanies increasing atypicality. We may expect that a psychologist will vary interaction strategies when struggling to engage the child, which may explain this finding.

The strongest correlation measured from the psychologists' acoustics is between median word-level jitter and the Communication Total, ρ=0.77 (p<10e-5). Such a correlation may indicate that the psychologist is adapting her voice quality, deliberately or spontaneously, to the child's voice quality parameters.

A final interesting finding comes from syllabic speaking rate patterns of the psychologists. The 90% quantile of speaking rate for utterance turn boundary words increases with atypicality, whereas the same feature for non-turn-end words decreases. In other words, the psychologist speaks slower in the middle of a sentence and faster at turn-end with increasing rated-atypicality. We might expect such a result, if we imagine the dynamics of an interaction between an adult and a child who is not engaged. The psychologist may speak slower during the middle of a sentence to make sure the child will comprehend, but also speak quicker (and with rising intonation) at the end to add excitement. While the results are based on a very limited data set, they nevertheless underscore the importance of considering the psychologist's acoustics when modeling the child's spontaneous speech.

4. Predictive Tasks & Discussion

We performed multiple linear regression using the entire child's and psychologist's prosodic feature sets to assess how predictive the feature sets might be (Table 3). A leave-one-session-out modeling is utilized in two layers, one for prediction and another for forward-feature selection parameter tuning. We chose Spearman's rank-correlation coefficient for final analysis and for tuning.

Intriguingly, the considered psychologist's acoustics were more predictive of the child's rated atypicality than were the child's own acoustics. The child's features seem to be predictive of the Communication Total and thus the combined ADOS Total, but the correlations are medium-low, ρ=0.36 and ρ=0.37. The psychologist's features are shown to be more predictive, with medium-strong correlations of ρ=0.61 for both the Communication Total and the Social Interaction Total. No advantage was observed when combining the psychologist's and the child's features. One possibility may be the challenges inherent in processing child speech (aside from the challenge of ASD heterogeneity), and hence the relatively more accurate feature characterization of the adult psychologist's speech leading to stronger correlations.

However, this raises interesting possibilities– 'Can the strategies of a psychologist be modeled? In addition, can we interpret the specific cues of the child's speech

Table 3: *Correlations of prosodic feature sets' predictions with ADOS code labels. [*,**,***]≡ α=[0.10,0.05,0.01]*

	Code Label			
Child's Acoustic-	Atyp.	Comm.	Soc.Int.	C&SI
Prosodic Feature	Pros.	Total	Total	Total
Child		$\mathbf{0.36}^{*}$		$\mathbf{0.37}^{*}$
Psychologist		$\mathbf{0.61}^{***}$	$\mathbf{0.61}^{***}$	$\mathbf{0.45}^{**}$
Both		$\mathbf{0.63}^{***}$		$\mathbf{0.50}^{***}$

they are attuning to and those cues they are changing within their own speech?'. Such data analysis can offer further insights into a more detailed characterization of the prosody in speech communication.

To ensure that we were not modeling any bias in ratings across psychologists, we performed an experiment to predict the psychologist from their acoustics (already having some speaker-normalization), where the regressand representing a psychologist is the mean Communication Total for sessions which she evaluated. We conclude the features hold information beyond speaker identity since neither Pearson's nor Spearman's correlation coefficients showed significance at the α=0.10 level.

It is important to analyze these interesting results further. Foremost, the psychologist is an active participant who influences the interaction during assessment. While we do not have enough data to sort out all of the sources of variability between the behaviors of the psychologists when interacting with children that have different engagement and abilities in the conversation, we have observed that the psychologists' acoustic features are informative of their evaluations. Additional data will allow us to address variability across children and psychologists. We may potentially more robustly predict the child's behavior given sufficient data from a particular psychologist.

5. Conclusions

In this work, we demonstrated that the child's prosodic features correlated with their session level ratings, and we further introduced the concept of modeling a psychologist's prosodic behavior, given that the psychologist is both interlocutor and evaluator, in order to indicate the child's perceived behavior ratings. More interesting was the finding that the psychologist's features were more predictive of those ratings. This suggests the psychologist is attuning to the child's behavioral cues, deliberately or spontaneously. Modeling can leverage this intuitive finding to inform a more precise characterization of prosodic patterns in communication, offering insights into the nature of interaction strategies with children diagnosed with ASD. For example, it is of interest to the psychological community to know at what point a psychologist makes her decision in this assessment scenario.

In the future, our analyses will focus on the study of temporal patterning of speech cues. The child's vocal behavior will be modeled throughout the session to examine whether atypical speech behavior is globally uniform or locally dependent on context. Furthermore, additional data will provide the opportunity to address the types of interaction strategies utilized between the psychologist and child, and offer insights into the salient aspects of prosodic factors that may lead to specific perceived ratings. Modeling will also explore computing vocal entrainment, and its directionality, between the dyads [18]. Given normative data we can find non-linear variability in features– one of the major difficulties in quantifying atypical prosody with greater precision and detail. Accordingly, we plan to collect an analogous corpus with typically-developing children.

6. Acknowledgements

This work was supported by funds from NSF and NIH.

7. References

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

[2] National Center on Birth Defects and Developmental Disabilities, (2010, March) Autism Spectrum Disorders (ASDs). [online]. http://www.cdc.gov/ncbddd/autism/.

[3] S. Baron-Cohen, "Social and Pragmatic Deficits in Autism: Cognitive or Affective?" *Journal of Autism and Developmental Disorders*, vol. 18, no. 3, pp. 379–401, 1988.

[4] 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.

[5] L. D. Shriberg, R. Paul, L. M. Black, and J. P. van Santen, "The Hypothesis of Apraxia of Speech in Children with Autism Spectrum Disorder," *J. of Autism and Dev. Disord.*, vol. 41, pp. 405–426, 2011.

[6] 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.

[7] J. McCann and S. Peppe, "Prosody in Autism Spectrum Disorders: A Critical Review," *Int. J. Lang. Comm. Dis.*, vol. 38, pp. 325–350, 2003.

[8] 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.

[9] 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.

[10] J. Paccia and F. Curcio, "Language Processing and Forms of Immediate Echolalia in Autistic Children," *Journal of Speech and Hearing Research*, vol. 25, pp. 42–47, 1982.

[11] M. C. Wallace, J. E. Cleary, E. H. Buder, W. Pettit, and D. K. Oller, "An Acoustic Inspection of Vocalizations in Young Children with Autism Spectrum Disorders," in *IMFAR*, 2008.

[12] 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*, vol. 44, pp. 1097–1115, 2001.

[13] J. P. H. van Santen, E. T. Prud'hommeaux, L. M. Black, and M. Mitchell, "Computational Prosodic Markers for Autism," *Autism*, vol. 14, pp. 215–236, 2010.

[14] M. P. Black, A. Katsamanis, B. R. Baucom, C.-C. Lee, A. C. Lammert, A. Christensen, P. G. Georgiou, and S. S. Narayanan, "Toward Automating a Human Behavioral Coding System for Married Couples' Interactions Using Speech Acoustic Features," *Speech Communication*, 2011, in Press.

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

[17] 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.

[18] C. Lee, A. Katsamanis, M. Black, B. Baucom, P. Georgiou, and S. Narayanan, "An Analysis of PCA-based Vocal Entrainment Measures in Married Couples' Affective Spoken Interactions," in *Proceedings of Interspeech*, 2011.