INTRODUCTION

Autism Spectrum Disorder (ASD) is a pervasive developmental disorder characterized by impaired social functioning and communication deficits (American Psychiatric Association, 2000, *Diagnostic and statistical manual of mental disorders,* 4th ed. text rev.). Early descriptions of the speech of children with ASD were rife with terms like ‘robotic’, ‘stilted’, ‘monotone’, ‘singsong’, and ‘exaggerated’ (Fay and Schuler, 1980; Baltaxe and Simmons, 1985; Baron-Cohen and Staunton, 1994). These labels refer to *expressive prosody*: roughly speaking, the intonation and rhythm patterns specified by a speaker’s manipulation of F0, amplitude and timing.

Recent studies have corroborated these anecdotal accounts, measuring atypical patterns in the acoustic correlates of both intonation and rhythm produced by children with ASD. Findings on intonation suggest that, compared with typically developing (TD) controls, children with ASD produce pitch contours that are highly variable. In a study that examined narratives taken from two separate ASD populations (17 children and 21 adolescents), Diehl *et al.* (2009) reported high F0 range relative to TD children, especially in the younger group. In a picture-naming task performed by 41 children with ASD (ages 4;0 to 6;6), Bonneh *et al.* (2011) found high pitch variability in addition to expanded range. Sharda *et al.* (2010) likewise corroborated the findings on range, and also reported differences in local pitch excursion.

In addition to intonation differences, duration and amplitude measures point to disordered rhythm. In a lexical production task, Grossman *et al.* (2010) found longer durations in words produced by children with ASD relative to TD controls. The researchers also reported exaggerated within-word pauses, especially when stress fell on the second syllable. Diehl and Paul (2011) also found longer durations and attributed the differences to “difficulties in controlling the precise temporal aspects of word production” (p.22). In a task involving imitation of nonsense syllable strings, Paul *et al.* (2008) reported that adolescent speakers with ASD had difficulty reproducing the correct stress patterns. Finally, in a repetition task designed to elicit affect, Hubbard and Trauner (2007) found some differences in amplitude range and peak placement.

Taken together, the results of these studies point to measureable disruptions in expressive ASD prosody. However, as Diehl and Paul (2011, p.23) point out, what remains unknown is the extent to which these acoustic differences matter to the average listener. Can listeners perceive disorder based on prosody alone, independent of other linguistic and communicative markers that characterize the ASD population? If so, to what extent are their judgments based on intonation versus rhythm patterns? Since the speech and language patterns of children with ASD may lead to stigmatization that impedes their social development (Paul *et al.*, 2005), these questions are non-trivial.

In a preliminary investigation that utilized filtered speech, Redford, Kapatsinski and Cornell-Fabiano (in prep.) showed that naïve listeners distinguished between ASD and TD samples, reliably judging the former as more disordered. However, the filtering method failed to eliminate all traces of intelligibility from the samples, so it is unclear to what extent the listeners relied on language when making dysprosody judgments. Furthermore, since filtering preserves both rhythm and intonation cues to some degree, the relative salience of these two components could not be investigated. The present study addressed these problems by employing speech resynthesis methods that allowed for independent control over F0, duration, intensity and lexical information. These methods have proven useful in investigating how listeners categorize languages based on prosodic cues (Ramus and Mehler, 1999; Ramus, Dupoux, and Mehler, 2003; White, Mattys, and Wiget, 2012). Here, we utilized them in two experiments that addressed listeners’ perception of dysprosody in utterances produced by children with ASD.

EXPERIMENT 1

In the first experiment, we asked whether listeners could reliably judge delexicalized speech samples produced by children with ASD as more disordered than samples produced by TD controls. Furthermore, we investigated the effects on listener performance when either rhythm or intonation is removed from the delexicalized speech.

Method

Participants

Speech samples provided by 36 children in Redford *et al.* (in prep.) were used in this study. Eighteen of the children were receiving special education services under an autism eligibility, and were recruited for the study through local speech and language pathologists (SLPs). The children ranged in age from 6;4 to 11;8 years (*M* = 8;11 years). Table 1 shows the sex, age, symptom severity, language delay and standardized receptive vocabulary score for each child. Symptom severity was established via the Childhood Autism Rating Scale (Schopler *et al*., 2010), which was completed by the referring SLPs. Language delay was established by parental report. Vocabulary skills were scored using the Peabody Picture Vocabulary Test (PPVT-4; Dunn and Dunn, 2007). The control group consisted of eighteen TD children (5 girls, 13 boys) ranging in age from 7;4 to 10;10 years (*M* = 8;11 years). Typical development was based on speech, hearing and language development as reported by the parents.

TABLE 1. Characteristics of child participants with ASD.

Child ID	Sex	Age	Symptoms	Lg. Delay	PPVT
1	M	6;4	severe	No	106
2	M	7;1	minimal	Yes	119
3	M	7;9	severe	Yes	126
4	M	7;11	severe	Yes	83
5	M	8;1	moderate	Yes	85
6	M	8;1	severe	No	107
7	M	8;1	moderate	Yes	112
8	M	8;1	minimal	Yes	113
9	M	8;10	severe	No	86
10	F	9;3	moderate	No	83
11	F	9;4	severe	No	80
12	M	9;6	severe	Yes	104
13	M	9;6	severe	Yes	78
14	M	10;3	severe	Yes	94
15	M	10;9	minimal	Yes	102
16	M	11;1	severe	No	64
17	F	11;7	severe	Yes	90
18	M	11;8	severe	Yes	94

Prosody judgments were provided by 12 college-aged listeners. These subjects were recruited from a pool of students enrolled in introductory psychology and linguistics classes, and they received course credit for participation.

Stimuli

The stimuli consisted of resynthesized and delexicalized versions of the speech samples obtained in Redford *et al.* (in prep.). In that study, the children were asked to choose one picture book from the "frog stories" series by Mercer Meyer. After taking some time to become familiar with the book, each child provided a narrative based on the sequence of events shown in the pictures. The stories were recorded using hat-mounted lavalier microphones, and four short, non-consecutive excerpts were subsequently extracted from each recording. The excerpts constituted prosodically complete utterances without internal pauses or disfluencies. To facilitate accurate tracking of F0, samples marked by extensive presence of creaky voice were excluded in favor of ones characterized by modal phonation. The ASD utterances ranged from 1.1 to 3.4 seconds in length (*M* = 2.2 sec) and contained between 3 and 14 syllables (*M* = 7.5). The TD samples were between 1.4 and 3.1 seconds (*M* = 2.2 sec) and consisted of 4 to 14 syllables (*M* = 8.3).

Prior to resynthesis, the F0 contour of each sample was stylized by (1) running the MOMEL script (Hirst, 2007) to identify the peaks and valleys, (2) hand-correcting the results, and (3) linearly interpolating between the points (including across voiceless intervals). These steps were done in order to eliminate the adverse effects of microprosody and pitch-tracking error on the subsequent resynthesis process. Each utterance was also phonetically transcribed by trained listeners, and segmented into phone-sized units in Praat (Boersma, 2001), using standard acoustic segmentation criteria (see, e.g., Klatt, 1976).

Resynthesis and delexicalization were accomplished with MBROLA software (Dutoit *et al.*, 1996), which works by concatenating pre-synthesized diphones from a database specified by the user. For this study, we chose a U.S. English database synthesized in a female voice. First, the phonetic labels and durations, as well as the F0 peaks and valleys generated by the stylization, were fed into the program. Next, following Ramus and Mehler (1999), three experimental conditions were created. In the Intonation (I) condition, all of the phonetic segments were replaced with /a/, producing an uninterrupted melody of the stylized pitch contour. In the Rhythm (R) condition, all the vowels and syllabic sonorants were replaced with /a/, all the remaining segments with /s/, and the F0 contour was held constant at 230Hz, resulting in a series of monotone *sasasa*-like sequences. In the Rhythm and Intonation (RI) condition, the phonemes were also replaced with /a/ or /s/, but the F0 values were kept faithful to the stylized contours. All of these manipulations were subsequently output to .wav files. Because MBROLA can only preserve durational correlates of rhythm, intensity cues to rhythm were reintroduced into the R and RI conditions extracting the original contours from the utterances and overlaying these onto the resynthesized utterances.

Table 2 lists the values for several acoustic correlates of intonation and rhythm in the resynthesized samples. The intonation measures were based on the peak and valley points of the stylized pitch contours. *Mean F0* simply averaged the pitch values of these points. *F0 range* represents the difference between the highest peak and the lowest valley. *Declination slope* was calculated by fitting a least-squares linear regression line to the pitch points. *PMD* (pitch movement density) was calculated by dividing the total number of peaks and valleys by the total number of syllables. It is a new measure intended to capture local F0 variability, normalizing for speech rate.

TABLE 2. Acoustic measures of intonation and rhythm correlates in the resynthesized files. The values represent group means, with standard deviations in parentheses. Normal distributions allowed the use of *t*-tests.

Measure	**ASD**	**TD**	***t*-test**
INTONATION:			
mean F0 in Hz	246 (42)	251 (40)	ns
F0 range in Hz	135 (61)	129 (80)	ns
F0 declination slope in Hz/sec	-14.96 (42.88)	-24.03 (24.86)	ns
PMD in elbows/syl	1.59 (.42)	1.30 (.38)	**p < .0001**
RHYTHM:			
speech rate in syl/sec.	3.48 (.81)	3.82 (.80)	**p < .05**
%V	42.31 (8.25)	41.78 (7.50)	ns
VarcoV	57.52 (18.72)	54.79 (16.63)	ns
VarcoC	67.09 (14.42)	59.20 (12.26)	**p < .001**
intensity slope in dB/sec.	-1.69 (10.84)	-7.81 (11.10)	**p < .005**

Speech rate was measured using the standard formula of syllables per second (recall that the utterances contained no internal pauses). Calculation of *%V* was performed by adding the durations of vocalic segments, dividing the result by the total duration of all segments, and multiplying by 100. *VarcoV* was measured by taking the standard deviation of vocalic interval durations, dividing by the mean of these durations, and multiplying by 100. *VarcoC* was the standard deviation of consonantal durations, divided by the mean and multiplied by 100. The *intensity slope* was calculated by fitting a linear regression line to the intensity tiers.

Unlike in previous work, the groups in this study did not differ in *F0 range*. However, the ASD group did exhibit more local pitch variability as measured by *PMD*. Table 2 also shows differences in three rhythm-related measures: *speech rate*, *VarcoC* and *intensity slope*. Again, these differences characterize the resynthesized files and not the original speech samples. In fact, due to the F0 and intensity manipulations performed in creating the stimuli, several of the raw measures differed from ones taken from the original files. However, *t*-tests performed on original measurements yielded similar results, with only the differences in, *rate*, *VarcoV* and *intensity slope* reaching significance (*PMD*, being dependent on a stylized contour, could not be calculated for the natural speech samples). We therefore concluded that the resynthesis method was successful in preserving group differences in the measures of interest.

Following resynthesis, the files were grouped by speaker and concatenated into four different sequences, with 300ms silence intervals separating the resynthesized excerpts. The concatenating procedure pseudo-randomized the order of the samples in order to ensure that each appeared in sequence-initial and sequence-final position only once across concatenated sequences. Concatenation was performed because we were interested in listeners' holistic evaluations of each speaker, and pilot work indicated that the isolated excerpts were too short to provide the subjects with enough information to form these types of judgments. Each condition thus yielded a set of 144 stimuli of approximately 10 seconds in length (36 speakers X 4 sequences).

Procedure

Listeners were presented with the stimuli over headphones, with the 3 conditions counterbalanced for order. Due to the unfamiliar nature of the resynthesized and delexicalized samples, each condition was previewed with a familiarization phase that consisted of 36 stimuli (one sequence per speaker). During this phase, the subjects were told that the samples came from a mix of children with both typical or atypical language development. They were instructed to pay attention to the relevant prosodic components (i.e. rhythm, intonation, or both) and try to get an idea of how the two groups might differ. Following familiarization, the subjects were presented with 72 stimuli in random order (a different pair of sequences per speaker was used in each condition), and asked to decide whether each sample came from a child with typical or disordered language development by clicking on the appropriate button on the screen.

Predictions

Based on (a) the findings reviewed above, (b) the preliminary results in Redford *et al*. (in prep.), and (c) the differences shown in Table 2, we hypothesized that listeners would not require lexical information to reliably judge the speech of children with ASD as atypical. We also anticipated an effect of condition. In terms of reliable categorization of the TD and ASD groups, we predicted best performance in the RI condition, which provided listeners with the most information. In addition, since the acoustic measures in Table 2 showed more differences in rhythm than intonation correlates, we hypothesized better performance in the R condition than in the I condition.

Results

Responses indicating "disordered" judgments were summed across listeners as a function of group and condition. In the RI condition, the mean values were 10.56 for ASD and 8.22 for TD. In the I condition, the averages were 11.44 for ASD and 9.22 for TD. In the R condition, the values were 11.39 for ASD and 8.56 for TD. Figure 1 illustrates these numbers.

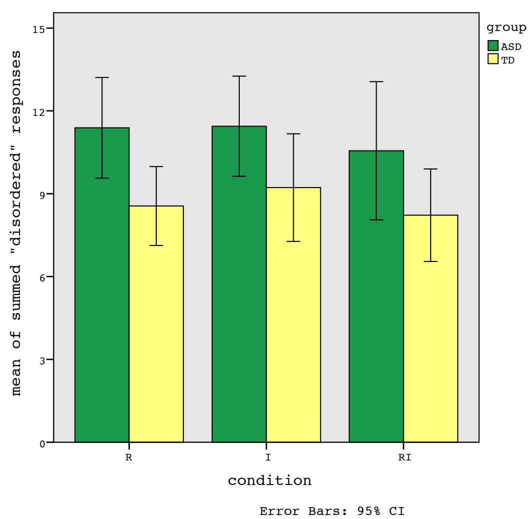

FIGURE 1. Mean summed "disordered" responses as a function of three conditions: Rhythm (R), Intonation (I) and Rhythm and Intonation (RI).

Responses to the ASD stimuli were not normally distributed. To test the hypothesis that listeners could judge the ASD samples as more disordered than the TD samples, Mann-Whitney U tests were performed for each condition. Contrary to our first hypothesis, there was no group effect in the RI condition ($U = 122.5$, $Z = -1.26$). There was also no difference between the groups in the I condition ($U = 117.5$, $Z = 1.42$). However, in the R condition, listeners judged the ASD samples (mean rank = 22.31) as more disordered than TD samples (mean rank = 14.69) , consistent with our prediction ($U = 93.5$, $Z = -2.18$, $r = .36$, $p = .029$).

EXPERIMENT 2

Experiment 1 suggests that intonation alone does not provide listeners with enough information to reliably judge ASD samples as more disordered than TD samples. In Experiment 2, we asked whether intonation would be more informative in the presence of lexical information.

Method and Predictions

Eleven college-aged adults participated in Experiment 2. The subjects were drawn from the same pool as those who took part in Experiment 1. Two conditions were investigated, which we termed RI+LEX and R+LEX. The stimuli for these were created by modifying the MBROLA files generated in Experiment 1. Specifically, the sole manipulation involved undoing the phoneme replacement: in Experiment 2, the phonetic labels that provide input to the synthesizer were not changed to /a/ and /s/, but remained faithful to the transcriptions. Thus, the RI+LEX condition preserved the duration, F0, intensity and lexical information of the original utterances. The R+LEX condition was identical except that F0 was held constant at 230Hz, preserving language and rhythm, but not intonation. The task procedure was identical to that of Experiment 1.

Children with ASD have been found to perform poorly on some tasks that require the use of prosody to achieve linguistic/communicative goals like expressing affect or focusing a lexical item (Peppé et al., 2007). Since these tasks often involve voluntary manipulation of F0 by the speaker, and since only the RI+LEX stimuli contained variable F0 information, we reasoned that listeners would pick up on any irregularities in tune-to-text alignment and judge the RI+LEX samples more accurately than the R+LEX samples.

Results

Summing "disordered" responses across listeners yielded the results seen in Figure 2. In the R+LEX condition, the mean values were 12.17 for ASD and 7.78 for TD. In the RI+LEX, the values were 12.33 and 8.05 for ASD and TD, respectively.

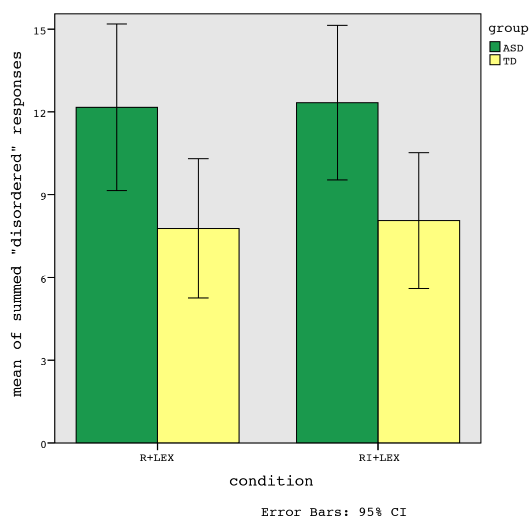

FIGURE 2. Mean summed "disordered" responses as a function of two conditions: Rhythm and Lexical Information (R+LEX), and Rhythm, Intonation and Lexical Information (RI+LEX).

Once again, the responses were not normally distributed. As in Experiment 1, Mann-Whitney *U* tests were performed to evaluate the hypothesis that ASD speech would be perceived as more disordered than TD speech. In the RI+LEX condition, listeners performed in the expected direction, judging the ASD samples (mean rank = 22.50) more disordered than TD samples (mean rank = 14.50) at a significant level ($U = 90.0$, $Z = -2.28$, $r = .38$, $p = .022$). In the R+LEX condition, the ASD group (mean rank = 22.17) was also judged more disordered than the control group (mean rank = 14.83), with the difference reaching significance ($U = 96.0$, $Z = -2.09$, $r = .35$, $p = .037$).

DISCUSSION

Taken as a whole, the results of Experiments 1 and 2 indicate that differences in rhythm-related measures (*speech rate*, *VarcoC* and *intensity slope*) are robust enough to translate into reliable perceptual judgments that distinguish ASD speech from TD speech. This was the case regardless of the presence or absence of lexical information. In fact, the effect sizes in the R and R+LEX conditions were similar (.36 and .35, respectively), suggesting that language context played little role (if any) in guiding the listeners.

Intonation paints a different picture. Presented on its own, it did not provide the listeners with enough information to differentiate ASD samples from TD samples, despite the measureable group difference in local pitch variability (*PMD*). This result is perhaps not surprising, since it is consistent with other findings where listeners have trouble categorizing 'disembodied' intonation contours (Ramus and Mehler, 1999). It is also consistent with related work on ASD prosody. In a recent study, Nadig and Shaw (2012) measured higher pitch variability in the speech of children with ASD relative to controls (unlike here, the difference was in *F0 range*). However, when asked to focus on pitch variability in natural language samples, trained listeners (speech-language pathology students) rated both groups equally (the groups were distinguished only when the task called for an 'overall impression' of the children's speech, suggesting that F0 differences played no part in the judgments). Together with our findings, these results suggest that high pitch variability that characterizes ASD speech (whether measured in *F0 range* or *PMD*) is within the 'acceptable' range of listeners.

Despite reliable listener performance in the R condition, adding intonation to the rhythm (RI) produced an unexpected result: subjects were no longer able to distinguish the groups even though they had more prosodic cues at their disposal. It seems that, when intonation falls within the range of acceptability, its presence attenuates the relative salience of atypical rhythm, resulting in generous judgments of typicality. This would suggest that, in a sense, intonation is unhelpful to the listener. However, results in the RI+LEX condition show that this is not exactly the case. When language context was added to the equation, listeners once again performed at a significant level. It seems that intonation information is useful, but unlike rhythm, it needs to map onto lexical content. It may be the character of these mappings, and not the acoustic properties of the intonation cues *per se*, that forms the basis of typicality ratings.

To sum up, our results indicate a complicated relationship between rhythm, intonation and lexical information in listeners' judgments of dysprosody in ASD speech. We suggest that acoustic correlates of rhythm are perceptually robust independent of language context. By contrast, intonation is judged on the way it is mapped onto language to convey intended meaning. Thus, dysprosody in the speech of children with ASD appears to be both global (in rhythm) and functional (in intonation).

This preliminary study leaves open several questions for future investigation. First, although our analysis implied that listeners somehow *decided* that the pitch variation measured in the ASD samples was within the typical range, it is possible that they simply did not *perceive* a difference between the groups. Thus, it is unclear to what extent the intonation judgments were based on perceptual discrimination versus preconceived notions of what constitutes disordered prosody. It is possible that, for all conditions, judgments were considerably more conservative than discrimination thresholds. We plan to address this question in a follow-up ABX task. Second, with respect to the acoustic markers of dysprosody, the experimental conditions created here do not exhaust the possibilities afforded by the resynthesis method. For example, it may be of interest to leave out the intensity information to judge its relative impact on the rhythm judgments. Also, by replicating the judgment task on natural speech samples and comparing the results to those obtained in RI+LEX, one could investigate the contribution of voice quality, which has been found to differ acoustically between the groups (Bonneh *et al.*, 2011). Finally, while we found group differences on the *PMD* measure, we are cautious about interpreting them at this point. Recall that the measure was based on counting peaks and valleys generated by hand-correcting the output of a pitch stylizing algorithm. The corrections led to perceptual equivalence (in the sense of 't Hart, Collier and Cohen, 1990) between the original and stylized versions as judged by one of the authors. In order to validate the *PMD* results, these judgments need to be corroborated by other listeners.

ACKNOWLEDGMENTS

This research was supported by Award Number R01HD061458 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD). The content is solely the responsibility of the authors and does not necessarily reflect the views of NICHD. The authors are also grateful to Jolynn Cornell-Fabiano, who introduced us to these wonderful children and their families, which made this work possible.

REFERENCES

American Psychiatric Association. **(2000)**. *Diagnostic and statistical manual of mental disorders* (4th ed.,text rev.). Washington, DC: Author.

Baltaxe, C.A.M., and Simmons, J. **(1985)**. "Prosodic development in normal and autistic children," In E. Schopler and G. B. Mesibov (eds.), *Communication Problems in Autism* (pp. 95–125). New York: Plenum Press.

Baron-Cohen, S., and Staunton, R. **(1994)**. "Do children with autism acquire the phonology of their peers? An examination of group identification through the window of bilingualism," First Lang. **14**, 241–248.

Boersma, Paul **(2001)**. "Praat, a system for doing phonetics by computer," Glot Int. **5**(9/10), 341-345.

Bonneh, Y. S., Levanon, Y., Dean-Pardo, O., Lossos, L., and Adini, Y. **(2011)**. "Abnormal speech spectrum and increased pitch variability in young autistic children," Frontiers Hum. Neurosci. **4,** 237

Diehl, J. J., and Paul, R. **(2011)**. "Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children by children with autism spectrum disorders," Appl. Psycholing. **34**(1)**,** 135-161.

Diehl, J. J., Watson, D., Bennetto, L., McDonough, J., and Gunlogson, C. **(2009)**. "An acoustic analysis of prosody in high-functioning autism," Appl. Psycholing. **30**(3), 385-404.

Dunn, L., & Dunn, D. **(2007)**. *Peabody Picture Vocabulary Test, Fourth Edition* (PPVT-4). Minneapolis, MN: NCS Pearson.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., and van der Vrecken, O. **(1996)**. "The MBROLA project: towards a set of high quality speech synthesizers free of use for non commercial purposes," Proc. ICSLP '96. **3**, 1393-1396.

Fay, W. H., and Schuler, A. L. **(1980)**. *Emerging language in autistic children*. Baltimore: University Park Press.

Grossman, R. B., Bemis, R. H., Plesa Skwerer, D., and Tager-Flusberg, H. **(2010)**. "Lexical and affective prosody in children with High-Functioning Autism," J. Speech Lang. Hear. Res. **53**(3), 778–793.

Hart, J. 't, Collier, R., and Cohen, A. **(1990)**, *A perceptual study of Intonation*. Cambridge: Cambridge University Press.

Hirst, D. **(2007)**. "A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation," Proc. 16th ICPhS. 1233-1236.

Hubbard, K., and Trauner, D. A. **(2007)**. "Intonation and emotion in autistic spectrum disorders," J. Psycholing. Res. **36**, 159–173.

Klatt, D.H. (**1976).** "Linguistic uses of segmental duration in English: acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**, 1208-1221.

Nadig, A., & Shaw, H. **(2012)**. "Acoustic and perceptual measurement of expressive prosody in high-functioning autism: increased pitch range and what it means to listeners," J. Autism and Developm. Disord. **42**, 499-511.

Paul, R., Bianchi, N., Augustyn, A., Klin, A., and Volkmar, F. R. **(2008)**. "Production of syllable stress in speakers with autism spectrum disorders," Res. Autism Spectrum Disord. **2**(1), 110–124.

Paul, R., Shriberg, L.D., McSweeney, J., Cicchetti, D., Klin, A., & Volkmar, F. **(2005)**. "Brief report: relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorder," J. Autism Develop. Disord. **35**, 861–869.

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. **(2007)**. "Receptive and expressive prosodic ability in children with high-functioning autism," J. Speech Lang. Hear. Res. **50**, 1015–1028.

Ramus, F., and Mehler, J. **(1999)**. "Language identification with suprasegmental cues: A study based on speech resynthesis," J. Acoust. Soc. Am. **105**(1), 512–521.

Ramus, F., Dupoux, E., and Mehler, J. **(2003)**. "The psychological reality of rhythm classes: Perceptual studies," Proc. 15th ICPhS. 337-342.

Redford, M. A., Kapatsinski, V., and Cornell-Fabiano, J. "Naïve judgments of prosody produced by children with ASD," manuscript in preparation.

Sharda, M., Subhadra, T., Sahay, S., Nagaraja, C., Singh, L., Mishra, R. , Sen, A., Singhal, N., Erickson, D., and Singh, N. **(2010)**. "Sounds of melody — pitch patterns of speech in autism," Neurosci. Let. **478**(1), 42–45.

Schopler, E., Van Bourgodien, M.E., Wellman, G.J., & Love, S.R. **(2010)**. *Childhood Autism Rating Scale, Second Edition* (CARS2). Torrance, CA: WPS.

Shriberg, L. D., Paul, R., Black, L. M., and van Santen, J. P. **(2010)**. "The hypothesis of apraxia of speech in children with Autism Spectrum Disorder," J. Autism Develop. Disord. **41(**4), 405–426.

White, L., Mattys, S. L., and Wiget, L. **(2012)**. "Language categorization by adults is based on sensitivity to durational cues, not rhythm class," J. Mem. Lang. **66**(4), 665–679.