Lay listener classification and evaluation of typical and atypical children's speech

Melissa A. Redford[1], **Vsevolod Kapatsinski**[1], and **Jolynn Cornell-Fabiano**[2]

[1]Department of Linguistics, University of Oregon

[2]César E. Chávez Elementary School, Eugene School District 4J

Abstract

Verbal children with autism spectrum disorder (ASD) often also have atypical speech. In the context of the many challenges associated with ASD, do speech sound pattern differences really matter? The current study addressed this question. Structured spontaneous speech was elicited from 34 children: 17 with ASD, whose clinicians reported unusual speech prosody; and 17 typically-developing, age-matched controls. Multiword utterances were excerpted from each child's speech sample and presented to young adult listeners, who had no clinical training or experience. In Experiment 1, listeners classified band pass filtered and unaltered excerpts as "typical" or "disordered". Children with ASD were only distinguished from typical children based on unaltered speech, but the analyses indicated unique contributions from speech sound patterns. In Experiment 2, listeners provided likeability ratings on the filtered and unaltered excerpts. Again, lay listeners only distinguished children with ASD from their typically developing peers based on unaltered speech, with typical children rated as more likeable than children with ASD. In Experiment 3, listeners evaluated the unaltered speech along several perceptual dimensions. High overlap between the dimensions of articulation, clearness, and fluency was captured by an emergent dimension: intelligibility. This dimension predicted listeners' likeability ratings nearly as well as it predicted their judgments of disorder. Overall, the results show that lay listeners can distinguish atypical from typical children outside the social-interactional context based solely on speech, and that they attend to speech intelligibility to do this. Poor intelligibility also contributes to listeners' negative social evaluation of children, and so merits assessment and remediation.

Keywords

Autism; likeability; intelligibility; prosody; acoustics

INTRODUCTION

Temple Grandin, diagnosed with autism as a child, characterized her early speech in the following way: "My voice was flat with little inflection and no rhythm. That alone stamped me as different (Grandin & Scariano, 1986:21)." Especially in childhood, any difference that deviates from peer-defined norms can jeopardize a child's access to peer friendships and social support. For example, in the domain of speech and language, Rice and colleagues

have shown that children with specific language impairment and those who learn English as a second language have less access to positive social interactions in US schools than typically developing, native English-speaking children (Rice, Sell, & Hadley, 1991; Gertner, Rice, & Hadley, 1994). Atypical speech patterns may function like language disorder or incomplete second language acquisition by creating barriers to social interaction for children with ASD. But unlike language disorder or incomplete second language acquisition, the problem may have less to do with communication per se and everything to do with the conveyance of difference; specifically, a difference that results in a negative social evaluation. This possibility was noted even by Asperger (1944, as translated by Frith, 1991:70), who concluded his description of atypical speech in children with ASD by saying that "the language feels unnatural, often like a caricature, which provokes ridicule in the naïve listener." The goal of the present research was to test the hypothesized link between atypical speech patterns and a negative social evaluation of children with ASD. A secondary goal was to identify the patterns that lay listeners attend to when classifying children's speech as typical or not and when evaluating children's likeability as a function of their speech.

Background

A fairly substantial body of scientific work confirms that individuals with autism who speak do indeed produce atypical speech sound patterns, and especially atypical speech prosody. For example, Shriberg and colleagues (Shriberg, Paul, McSweeny, Klin, Cohen, & Volkmar, 2001) used the versatile and comprehensive but unstandardized Prosody-Voice Screening Profile (PVSP; Shriberg, Kwiatkowski, & Rasmussen, 1990) in their study of spontaneous speech in 30 individuals with ASD: trained listeners evaluated speech across a variety of predetermined prosodic and voice categories, including loudness, pitch, voice quality, resonance quality, phrasing, rate, and stress. The most consistent voice deficit was in resonance quality: an inappropriate nasal quality was more often identified in speech produced by individuals with ASD than by the controls. The most consistent prosodic deficit was stress; a category that included excessive/equal stress, prolongation and blocking type disfluencies, as well as incorrect production of lexical stress patterns and accentual patterns at the level of the phrase (labeled "contrastive stress"). The authors report that the perceived differences between individuals with ASD and control speakers was due to a greater number of perceived disfluencies and "misplaced word stress." (p. 1107). Insofar as a number of other perceptual studies similarly report autism-related deficits in the production of lexical stress and phrasal accenting (e.g., Baltaxe & Guthrie, 1987; McCaleb & Prizant, 1985; Paul, Augustyn, Klin, & Volkmar, 2005; Grossman, Bemis, Plesa-Skwerer, & Tager-Flusberg, 2010), it is likely that deficits in "stress" production is a defining feature of atypical prosody in ASD.

In contrast to perceptual studies, acoustic studies on autistic speech have focused on global patterns relevant to characterizing the prosodic differences in individuals with ASD. For example, Diehl et al. (2009) extracted average F0 (the acoustic correlate of pitch) and its standard deviation (= pitch variability) over large temporal windows (250+ milliseconds) in narrative speech produced by a total of 38 children and adolescents with ASD and 38 typical controls. They found that pitch variability was generally higher in speech produced by

children with ASD than in typically developing children. Other studies report similar results (Sharda, Subhardra, Sahay, … & Singh, 2010; Bonneh, Levanon, Dean-Pardon, Lossos, & Adini 2011; Depape, Chen, Hall, & Trainor, 2012; Nadig & Shaw, 2012), which could be consistent with the above-mentioned finding of impaired phrasal accenting in autistic speech if pitch variability captures something about accentual patterning.

In addition to deficits in lexical stress production and phrasal accenting, there is some evidence that individuals with ASD produce speech at slower rates than typical controls (for a review see Shriberg, Paul, Black, & van Santen, 2011). For example, Diehl and Paul (2011) used the same global F0 measures as Diehl et al. (2009) on utterances obtained from 24 children with ASD and 22 typically developing children in different subtests of the formally structured but unstandardized Profiling Elements of Prosody in Speech-Communication test (Peppé & McCann, 2003), but found no difference between the groups. They also took measures of F0 range as well as utterance-level measures of acoustic duration and intensity. Only duration differences were significant: children with ASD produced longer utterances on average than the typical controls.

Studies that use perceptual evaluation or measure global acoustics to identify prosodic deficits/differences assume an expanded definition of prosody relative to that which is assumed in careful linguistic phonetic studies. But whereas the linguist's interest is in the realization of abstract metrical and intonational structures and the alignment between them, the clinician's interest is in identifying those patterns that could adversely affect an individual's ability to communicate with—in order to be accepted by—peers. The search for prosodic deficits is thus expanded to include any suprasegmental pattern in autistic speech that might reliably signal difference or disorder.

Yet, even with its wide purview, existing research on the differences between autistic and neurotypical speech returns few consistent findings. Consider for example the contradictory results on the status of atypical pitch variability in individuals with ASD: Diehl and colleagues (2009) find significant differences between children with and without ASD in one task, but Diehl and Paul (2011) find no such differences in another task. Such inconsistent results may be due to the absence of deficits in a significant proportion of high-functioning individuals with ASD. Although systematic data on the distribution of atypical prosody in autism do not yet exist, it has been suggested that prosodic deficits may only be evident in the speech of half of all individuals with ASD (e.g., Paul, Shriberg et al., 2005).

Despite contradictory findings and the possibility that prosodic disorder occurs in only half of all high-functioning individuals with ASD, the topic of prosodic difference/deficits in autism has received increasing attention in recent years. Whereas McCann and Peppé (2003) found that between 1980 and 2002 only 16 studies on prosody in autism, a scholar.google.fr search in early 2017 on "prosody autism prosodic ASD" returned 1,010 results. The top 100 results, which included of articles, conference proceedings and book chapters, indicated at least 57 new studies (i.e., those published in peer-reviewed journals) that were focused on some aspect of prosody or prosodic processing in the speech of individuals with ASD. This kind of attention is surprising given that autism is, after all, characterized primarily by deficits in social communication and interaction as well as by restricted and repetitive

patterns of behavior that are disabling (American Psychiatric Association, 2013). The diagnostic criteria set forth in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) do not even mention atypical speech sound patterns, let alone prosody. They do however reference behaviors such as "failure to initiate or respond to social interactions" and "abnormalities in eye contact and body language… to a total lack of facial expressions" as well as stereotyped or repetitive motor movements… or speech" (e.g., echolalia or idiosyncratic phrases). Given the disabling nature of the criterial behaviors in autism, which are social or perseverative in nature rather than prosodic, is there a good reason to focus ever more research on the unusual speech patterns of individuals with ASD? This question motivated the current study.

The Current Study

The practiced clinician apparently has no trouble identifying individuals with autism based on atypical speech sound patterns when these are present (Klin, Sparrow, Marans, Carter, & Volkmar, 2000). Clinicians in training can also reliably distinguish speakers with and without autism based on short conversational speech samples that are measurably different with respect to pitch variability (Nadig & Shaw, 2012). These abilities are perhaps not surprising given that, as language speakers, we are attuned to sound patterns in speech that index social information. This sensitivity allows listeners to adeptly categorize speech by ethnolect (Lass, Almerino, Jordan, & Walsh, 1980; Purnell, Idsardi, & Baugh, 1999), dialect (Clopper & Pisoni, 2004; Clopper & Bradlow, 2009), gender identity (Lass et al., 1980; Munson, 2007), and sexual orientation (Munson, McDonald, DeBoe, & White, 2005; Munson, 2007). Moreover, listeners do not themselves need to be expert in the social categories of interest to effectively classify speech based on sound pattern differences (Purnell et al., 1999; Clopper & Bradlow, 2009). For example, Clopper and Bradlow (2009) used a free classification task to show that non-native speakers were able to cluster American English speech samples by dialect, albeit less accurately than native speakers of American English. Other studies strongly suggest that the perception of social information in speech has social consequences (van Bezooijen, 1995; Purnell et al., 1999). For example, Purnell and colleagues found that landlords discriminated against tenants based solely on speech patterns that indexed an African American or Chicano identity.

Our automatic identification of social information in speech coupled with impressionistic descriptions of atypical speech sound patterns in autism as too loud, shrill, sing-songy, mechanical and so on (see, e.g., Asperger, 1944 in Frith, 1991; Klin et al., 2000) suggest that autistic speech may be noticeable to and socially evaluated by listeners who have no known prior experience with autism. But the question remains as to how impactful these speech differences really are. More specifically, is it the speech per se that identifies a person with autism as "other" or is it perhaps the suite of unusual communicative and repetitive behaviors typical of autism that engender the perception of otherness? Also, if otherness can be discerned based solely on speech patterns, would such an identification result in an inherently negative social evaluation? To begin to answer these questions, the current study tested the ease with which lay listeners can distinguish children with ASD from their typically developing peers based solely on short snippets of their prosodically impaired

speech. We also investigated how listeners evaluated this speech absent a social or linguistic context.

Acoustic studies that find group differences between autistic and neurotypical speech strongly suggest that speech-specific differences are there and should be perceptible even to lay listeners. Yet, even this suggestion has yet to be rigorously tested. The perceptual rating study conducted by Nadig and Shaw (2012) provides some evidence in favor of the perceptual saliency of difference, but their listeners were speech-language pathologists in training who were also specifically taught to attend to pitch changes in the speech samples from children with and without ASD. In addition, the samples that Nadig and Shaw's listeners rated were excerpted from conversations with no control over language differences (word choice, appropriateness of response, etc.) that could influence judgments of typicality. The current study provides a more rigorous test of the hypothesis that the atypical speech produced by children with ASD and unusual prosody is perceptually distinct from typical children's speech (Experiment 1). After showing that it is, we go on to investigate whether perceived difference results in a negative social evaluation of the speaker (Experiment 2). We follow up by identifying the perceptual dimensions that listeners use both to distinguish atypical from typical speech and to socially evaluate a speaker (Experiment 3). The methods we use control as best as possible for influences of social-interactional context on listeners' classification and evaluation of speech: we elicited children's spontaneous speech in a narrative task, rather than in a conversational one, and further decontextualized even this speech by extracting only a few, fluently produced utterances from each narrative. The stimuli and judgment tasks were also designed to test the fundamental saliency of prosodic differences, and to control for possible effects of language on listeners' evaluation of speech patterns. Finally, we used lay listeners, who had no known a priori experience with ASD and relatively little experience on average with children, to maximize the possibility that listener judgments were reflective of broader societal norms rather than dependent on personal histories and experiences. In these ways, we sought to answer our motivating question of whether the atypical speech sound patterns often associated with ASD are themselves important to address directly given the many other socially-disabling behaviors more central to autism.

EXPERIMENT 1

The goal of Experiment 1 was to test lay listeners' ability to distinguish children with ASD from their typically developing peers based solely on their atypical speech. The hypothesis that listeners would be able to do this follows from our automatic identification of social information in speech and from acoustic studies that have measured and found differences in speech produced by many children and adults with ASD compared to those without ASD, particularly in the prosodic domain. Since these studies also show that prosodic deficits do not appear in all individuals with ASD, we obtained speech samples only from children with ASD who were already deemed by their speech–language pathologist to have unusual speech prosody.

Methods

Participants—Sixteen college-aged adults provided lay judgments of disorder. These adults were recruited from introductory psychology and linguistics classes, and received course credit for their participation. All reported normal hearing and English as their native language. The listeners evaluated speech obtained from 34 children, who also had normal hearing and English as their native language. Seventeen of the children (14 boys) had special education (SPED) eligibilities for receiving speech and language services in the schools under autism spectrum disorder (ASD). Their mean age was 9;0 years (SD = 18 months). The other 17 children (13 boys) were typically developing. Their mean age was 8;9 years (SD = 15 months).

Children with ASD were recruited for the study through a local network of school speech–language pathologists (SLPs) in 2011. The SLPs were asked to refer children from their caseload who had (1) a SPED eligibility under ASD, (2) normal cognition, and (3) unusual prosody. In the state of Oregon, a determination of autism for SPED eligibility is made by the school district and involves a team of education professionals including a school SLP and clinical psychologist. This team makes their determination based on observation of the child in and out of the classroom, and uses assessment instruments to evaluate target behaviors, cognition, speech and language (see Oregon Administrative Rules for Special Education at http://www.ode.state.or.us/offices/slp/spedoars.pdf). The Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, Pamela, Dilavore, & Risi, 2008) is not used, nor is a medical diagnosis of autism required or accepted as a substitute for a full evaluation by the district team. Regarding unusual prosody, we relied on the school SLP's interpretation of what this might mean in keeping with both the goals of the study and the very broad definition of what constitutes prosody in research on autistic speech.

We collected data from 18 children who were thusly referred, but one child's speech was accidentally not recorded. Symptom severity for the remaining 17 children was measured by asking the referring SLP to complete the Childhood Autism Rating Scale (CARS2; Schopler, Van Bourgodien, Wellman, & Love, 2010). Parental report was used to establish whether or not the child had experienced language delay. Receptive vocabulary scores were obtained using the Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2007). Table 1 reports the sex, age, ASD symptom severity, language delay, and standardized receptive vocabulary scores for each child.

Typically developing children were not specifically recruited for the current study. Instead, these children represented a subset of 100 children who had participated in a longitudinal study on the typical acquisition of prosody in school-aged children, which was on-going at the time of the current study. Speech samples were selected from 17 of the children who best matched the ASD group in age and sex at the time of the present study. Typical development was assessed based on parent report regarding the child's hearing, speech, language, and medical history. The mean age in this group was not significantly different from the mean age in the group of children with ASD. Table 2 reports the sex, age, and standardized receptive vocabulary scores of the typically developing children.

Elicitation Procedure—Structured spontaneous speech samples were obtained using a storytelling task. An experimenter presented children with 4 picture books that depicted different adventures of a frog and/or a boy and a dog (i.e., the frog story books by Mercer Mayer). Children were asked to choose the picture book that s/he would most like to narrate. Typically developing children familiarized themselves with their book of choice under experimenter supervision, then told their story to a parent or caregiver two times. Story repetition was used to control for the effects of language planning on the speech produced (see Redford, 2013). As our interest is in speech and not in cognitive influences on language production, the samples used in the present study were extracted from the children's second storytelling.

Pilot work indicated that children with ASD resisted telling the same story twice to a parent or caregiver. The elicitation protocol was therefore adapted in the following way. Children chose a book to narrate, then developed a story in response to questions and prompts from the experimenter while looking through their book of choice. The child's caregiver was then invited back into the room (or the conversation if they had not left), and the child was instructed to "tell their story" to the caregiver. The scaffolding procedure served the same function as a first story telling in our study on typically developing children's speech acquisition: it allowed children to plan their story and practice the language they would use to tell the story.

All children's stories were digitally recorded onto a Marantz PMD660 (with a sampling rate of 44,100 Hz) using a Shure ULXS4 standard wireless receiver and a lavaliere microphone, which was attached to a baseball hat or headband that the speaker wore. Children with ASD would sometimes refuse to wear the hat or headband, in which case the microphone was placed either on the child's clothing or on the table between the experimenter and child.

Speech Samples—Four multiword utterances were extracted from the middle of each child's story to avoid the stereotyped language and prosody associated with story beginnings and endings. Utterances were selected on the basis of prosodic completeness and length: utterances that were roughly 2 seconds long were selected over those that were shorter or much longer. The goal was to select phrases that would best represent phrase-level rhythm and intonation (hence the relatively long minimum length) while controlling for significant junctures due to pauses and other disfluencies (hence the relatively short maximum length). Samples from children with ASD ranged in length from 1.12 to 3.37 seconds ($M = 2.21$) and from 4 to 14 syllables ($M = 8.08$). Samples from typically developing children ranged in length from 1.45 to 3.06 seconds ($M = 2.17$) and from 5 to 14 syllables ($M = 9.15$).

Once utterances were extracted, filtering was used to create stimuli for a pure prosody condition. The unaltered utterances were used to create stimuli for the corresponding full speech condition. The goal of the manipulation was to test the extent of prosodic difference in the speech of ASD versus typically developing children, keeping in mind that filtering always substantially degrades the speech signal.

Two different band pass filters were used: one preserved lower frequencies (250 to 750 Hz) and the other higher frequencies (500 to 1000 Hz). The band pass filter that preserved lower

frequencies was used to capture rhythmic information associated with the alternation of sonorant and nonsonorant speech intervals. The band pass filter that preserved higher frequencies eliminated the remaining low frequency energy associated with those sonorant consonants produced with significant oral occlusion in onset position. The goal of this filter was therefore to preserve information about the spacing of vowel onsets, which have been referred to as the perceptual centers of rhythmic beats in speech (Port, 2003). Note that pitch (i.e., F0) information was preserved under both manipulations because listeners calculate pitch from the frequency spacing between preserved harmonics (Moore, 1994). In the end, filter type had no significant effect on listener judgments and so were combined in the analyses reported below, which therefore only refer to filtered and unaltered speech samples.

Filtered and unaltered utterances were amplitude normalized to 75 dB using the scale function in Praat (Boersma & Weenik, 2009). Individual utterances from each speaker were then concatenated in random order to create stimuli that were blocked by speaker. Utterances were separated by 300 milliseconds of silence in each block. Blocking was deemed necessary because pilot work indicated that naïve listeners need to hear more than 2 seconds of speech from a single individual to make a reliable judgment. The blocks ranged in duration from 7.27 to 12.51 seconds ($M = 9.80$) for children with ASD and from 8.18 to 10.86 seconds ($M = 9.51$) for typically developing children.

Classification Task—Listeners were asked to make judgments of "typical" and "disordered" based on the speech samples. Listeners also indicated their confidence in each judgment on a 5-point rating scale. All response options were displayed as labeled buttons on a computer monitor. The two alternative judgments, "typical" and "disordered", were placed in a row above the confidence rating scale, which also appeared as a row. Listeners would hear a stimulus block, make a judgment, rate their judgment, then press "OK" to continue in the task. Stimulus blocks were presented in random order. Listeners heard one low-frequency band pass filtered stimulus block and one high-frequency band pass filtered stimulus block per speaker, and so made a total of 68 typicality judgments (34 speakers × 2 repetitions) in the filtered speech condition. Listeners heard the same unaltered stimulus block twice, and so also made a total of 68 typicality judgments in the full speech condition.

Listener instructions included information about the speakers, about how the speech stimuli were created, and the effect of band pass filtering on speech. Specifically, listeners were told that the speech samples had been elicited from children between the ages of 5 and 11 years old, and that half of the children were receiving speech-language services in the school. Listeners were assured the research objective was to help target speech-language deficits in children with disability, and so they should not feel inhibited in assigning a judgment of "disordered" to any speech sample they thought was produced by a child with disability. Listeners were also told that the utterances that made up each sample were taken from a larger story telling task, and that the story context was not preserved. Finally, listeners were told that some of the speech they would hear was filtered to remove as much language information as possible. They were instructed to make judgments based on what remained perceptible in the filtered speech; namely, rhythm and intonation. The experimenter also made sure that listeners understood what speech rhythm and intonation were. These instructions were given to provide listeners with every opportunity to distinguish children

with ASD from their typically developing peers when prosody alone was preserved in the filtered signal.

Finally, in order to gain insight into the basis for judgments of "typical" versus "disordered," we asked listeners to provide written feedback regarding their perceptual criteria. Specifically, listeners were asked to: "Please describe the characteristics in the speech samples that led you to distinguish between "typical" and "disordered" speech (please, especially, note what you think made the speech sound disordered)." Listeners responded to this request twice: first after completing judgments in the filtered speech condition, and then again after completing judgments in the unaltered speech condition.

Control Tasks—Two additional tasks were included in the experiment; both were used to control for the effects of language (e.g., word choice, syntactic structure, sample coherence) on "typical" and "disordered" judgments. The control task for the filtered speech condition investigated the effectiveness of the filter at removing language information: listeners indicated roughly how many words they thought they understood in the filtered speech samples on a 5 point scale: *none*, *some*, *half*, *most*, or *all*. These 5 response options were presented in a row of boxes on the monitor. Confidence ratings were again collected on each judgment that the listener made. The control for the unaltered speech condition directly investigated effects of word choice, syntactic structure, and other language features on typicality judgments. This task required listeners to make "typical" or "disordered" judgments based on written versions of the utterances rather than on aurally presented versions. The written versions of a speaker's utterances were displayed for 10 seconds to approximate the duration of the audio samples. Listeners wore headphones when making text-based judgments just as in all other conditions, they then clicked on "typical" or "disordered" as in the other classification tasks, and also indicated their degree of confidence in their judgment.

The experimental and control tasks were always completed in the following order: filtered speech, comprehension, text (i.e., written stimuli), unaltered speech. The fixed order had to do with the fact that filtering only effectively eliminates access to language content if listeners have no a priori expectations about this content.

Analyses—The binary responses were coded "0" for "typical" and "1" for "disordered" and summed across listeners to obtain a single score for each sample in each of the tasks where this response was elicited. Judgment scores in the experimental tasks were then stripped of language influence in the following way. Standardized comprehension rating scores for each sample were used to predict the summed "disordered" scores in the filtered speech condition (i.e., across filter type) using linear regression. Similarly, summed "disordered" scores from the text-only judgment task were used to predict the summed "disordered" scores in the unaltered speech condition. The residual scores from these analyses were saved, and the two scores per speaker averaged to obtain the dependent variable for the principle analysis, which investigated listeners' ability to distinguish speech produced by children with ASD from speech produced by children with typical development. Listeners' confidence in their typicality judgments were also analyzed as a function of speaker group. These were standardized within listener using a z-transform, then

averaged across listeners and within speaker to obtain the mean listener confidence rating per speaker.

The principle analyses used a mixed-design ANOVA to assess the between-subjects effect of speaker group (ASD versus typical) and within-subjects effect of speech condition (filtered versus unaltered) on the dependent variables: residualized "disordered" judgments and standardized confidence ratings. A Shapiro-Wilk test indicated that values associated with these dependent variables were normally distributed. A significant effect of group on "disordered" judgments would indicate that listeners can distinguish children with ASD from their typically developing peers based on speech patterns alone. An interaction between group and speech condition would indicate the influence of non-prosodic factors on "disordered" judgments. Speaker's age-in-months and standardized PPVT scores were included as covariates in the analyses to control for their effects on the dependent variables.

Results

Listener Responses by Judgment Task and Speaker Group—The untransformed and unstandardized responses across the different tasks were not the focus of analysis, but are summarized in Table 3 to give the reader a fuller sense of the data. Note that cumulative "disordered" scores were higher on average for children with ASD compared to their typically developing peers across all tasks. Note also that the mean difference between the two groups of children was especially large in the unaltered speech task.

Effect of Speaker Group on Residualized Judgments of Disorder—Recall that to investigate the influence of speech patterns on judgments absent the influence of language we regressed the scores obtained from the control tasks (comprehension and text-based judgments) against the scores obtained from the experimental tasks (filtered speech and unaltered speech judgments). The analysis on the residuals indicated that listeners were more likely to judge speech produced by children with ASD as "disordered" than speech produced by children with typical development, $F(1, 30) = 4.56$, $p = .041$, $\eta_p^2 = .132$. This effect of group interacted with condition, $F(1, 30) = 4.52$, $p = .042$, $\eta_p^2 = .131$, in that listeners were better able to distinguish between ASD and typically developing children in the unaltered condition than in the filtered condition (see Figure 1). The simple effect of condition was not significant.

Given the significant interaction between condition and group, and the seemingly weak effect of group on "disordered" judgments in the filtered speech condition (Figure 1), follow up analyses were conducted to determine whether listeners were in fact able to distinguish children with ASD from their typically developing peers in the filtered speech condition. The data were split by condition and the effect of group on residualized "disordered" judgments was tested again. The results were that listeners distinguished between groups of children based on speech patterns in the unaltered speech condition, $F(1, 30) = 5.99$, $P = .020$, $\eta_p^2 = .166$, but not in the filtered speech condition, $F(1, 30) = 0.70$, $p > .1$. Note that the effect of group on "disordered" judgments in the unaltered speech condition is significant even when alpha is corrected for multiple comparisons (i.e., the Bonferroni corrected alpha is .025).

Written Feedback from Listeners on their Judgments—Again, the effects of language were controlled in the unaltered speech condition by partialling out variance in judgments due to word choice, syntactic structure, and so. Thus, the results shown in Figure 1 indicate that listeners differentiated children with ASD and prosodic disorder from their typically developing peers based on speech sound patterns alone. Yet, there was no effect of group on the residualized "disordered" judgments in the filtered speech condition. This suggests either that prosodic differences between the two groups of children were subtle at best and listeners only gained sensitivity to those differences as the experiment progressed or that listeners attended to something other than prosody when making judgments of disorder. Listener feedback suggested both explanations could be true. Specifically, listeners reported attending to articulation, speech clarity, fluency (= flow, rate, rhythm), and accenting (= word emphasis, intonation) when making judgments. Fluency and accenting are clearly prosodic in nature; articulation and speech clarity reference segmental characteristics. Table 4 summarizes the feedback we obtained from each of the 16 listeners who provided judgments on the filtered and unaltered speech samples.

Listeners' Confidence in their Judgments—The analysis on listeners' standardized confidence ratings support the suggestion that the speech samples obtained from children with ASD were subtly different from those obtained from typically developing children. They may also have been more variable. Specifically, there was a significant effect of group on listeners' standardized ratings, $F(1, 30) = 10.75$, $p = .003$, $\eta_p^2 = .264$, as shown in Figure 2. Listeners were less confident in their judgments of speech produced by children with ASD than in their judgments of speech produced by children who were typically developing. There was no effect of condition or any interaction between condition and group.

Discussion

The results from Experiment 1 indicate that listeners correctly identify speech produced by children with ASD as "disordered" more often than speech produced by typically developing children. The sound pattern differences are sufficiently salient that listeners can distinguish the groups outside the social-interactional context and based on very short speech samples. Yet, the absence of a group effect on judgments in the filtered speech condition suggests that the sound pattern differences are subtle enough to be obscured in degraded speech. This suggestion is further supported by listeners' weaker confidence in their ratings of atypical children's speech compared to typical children's speech. Listeners' weaker confidence in their ratings of atypical speech may also indicate that the samples provided by children with ASD were more variable than the ones obtained from their typically developing peers.

EXPERIMENT 2

The goal of Experiment 2 was to test whether listeners' social evaluation of speakers based on speech sound patterns also differentiates children with ASD from their typically developing peers. This goal is in service of our question about whether atypical speech sound patterns matter, where the notion of what matters is understood as a question about the effects of difference on listeners' attitude towards the speaker. The relationship between

lay listeners' judgments of disorder (Experiment 1) and their likeability ratings was also directly explored.

Methods

Participants—Twelve new college-aged adults provided likeability ratings on the speech samples from Experiment 1. These adults were recruited from introductory psychology and linguistics classes, and received course credit for their participation. All reported normal hearing and English as their native language.

Rating Task—Listeners rated both the filtered and the unaltered speech samples from Experiment 1 for likeability on a 7-point Likert scale. The low value on the scale was anchored with a positive social statement ("1 = Awesome! Love this kid.") and the high value with a negative social statement ("7 = Nah. Wouldn't like this kid."). Listeners sat in front of a computer monitor with the anchors displayed above the rating scale, which was represented as a numbered sequence of 7 buttons. They listened to a speech sample over headphones, rated it along the likeability scale, clicked "OK" and advanced to a new screen and a new sample. As in Experiment 1, listeners completed the filtered speech condition before the unaltered speech condition. Speech samples were played in different random orders. Listeners were told only that the children were between 5 and 11 years old. Unlike in Experiment 1, they were not told that half of the children were receiving speech and language therapy in the schools.

Analyses—Likeability ratings were standardized within listener across conditions by converting the rating scores into z-scores. Ratings were then averaged across listeners and within speakers. A mixed-design ANOVA was used to assess the between-subjects effect of speaker group (ASD versus typical development) and within-subjects effect of speech condition (filtered versus unaltered) on likeability ratings. A significant effect of group on likeability ratings would indicate that listeners evaluate children with ASD differently from their typically developing peers based on speech patterns alone. An interaction between group and speech condition would indicate the influence of non-prosodic factors on listeners' ratings. Speaker's age-in-months and standardized PPVT scores were included as covariates in the analyses to control for their effects on the dependent variable.

A second analysis used linear regression to investigate the relationship between the residualized "disordered" judgments from Experiment 1 and the likeability ratings obtained from a different group of listeners in Experiment 2. Likeability ratings were the dependent variable, and judgments of disorder the predictor variable. Age-in-months, PPVT scores, and speech condition were entered as control predictor variables.

Results

Effect of Speaker on Likeability—The analysis indicated a significant interaction between group and speech condition, $F(1, 30) = 5.17$, $p = .030$, $\eta_p^2 = .147$, which is shown in Figure 3. Neither the simple effect of group nor that of speech condition was significant, though the effect of group approached significance ($p = .051$).

Given the interaction, the effect of group was explored further in analyses split by condition. These analyses indicated a significant effect of group on likeability ratings in the unaltered speech condition, $F(1, 30) = 6.79$, $p = .014$, $\eta_p^2 = .184$, but not in the filtered speech condition. As shown in Figure 3, children with ASD were perceived as significantly less likeable than children who were typically developing based on the unaltered speech samples[1]. Note that the effect is significant even when alpha is corrected for multiple comparisons (i.e., the Bonferroni corrected alpha is .025).

Relationship between Judgments of Disorder and Likeability—The results on likeability recall those from Experiment 1 in that listeners only discriminated between the groups of speakers in the unaltered speech condition. The similarity between the two experiments was confirmed in a linear regression analysis, which showed that the residualized judgments of disorder could be used to explain/predict likeability ratings, $\beta = .59$, $t(67) = 6.01$, $p < .001$, even with age-in-months, PPVT, and speech condition controlled in the model. By itself, the predictor "disordered" judgments accounted for 42% of the variance in likeability ratings. None of the control variables explained a significant proportion of the variance.

The relationship between perceived disorder and likeability is shown in Figure 4. Speech samples that elicited a higher "disordered" response elicited a more negative social evaluation ("Nah. Wouldn't like this kid."). When split by speech condition, we find that the correlation between the judgments of disorder and likeability ratings was stronger in the unaltered speech condition, $r(34) = .72$, $p < .001$, compared to the filtered speech condition $r(34) = .48$, $p = .004$, though this difference was not significant, $z = 1.51$, $p = .066$ (one-tailed).

Discussion

The results from Experiment 2 indicate that lay listeners evaluate children with ASD more negatively than their typically developing peers. Thus, like in Experiment 1, listeners were able to distinguish between speakers based on the short speech samples provided. The results also parallel those from Experiment 1 in that listeners were not able to distinguish between the different groups of speakers in the filtered speech condition. A regression analysis confirmed that the similarities between Experiment 1 and Experiment 2 were due to a significant correlation between lay listeners judgments of disorder and their social evaluation of children. Next, we investigate whether this correlation means that the same sound patterns trigger both the perception of disorder and a negative social evaluation.

EXPERIMENT 3

In Experiment 3, further perceptual evaluation of children's speech and acoustic measures of segmental and suprasegmental features were used to identify the speech sound patterning that mediated listeners' judgments of "typical" versus "disordered" in Experiment 1 and their likeability ratings in Experiment 2. The goal was to answer the question of whether

---

[1]Recall that the 7-point likeability scale was anchored such that higher scores indicate a more negative social evaluation than lower scores ("1 = Awesome! Love this kid." versus "7 = Nah. Wouldn't like this kid.").

both types of judgments reference the same patterns in order to understand why a judgment of disorder may lead to a negative social evaluation. Thus, the focus of Experiment 3 is on listeners' behavior. We do not investigate this behavior as a function of the speaker groups, but do confirm that children with ASD differ from typically developing children along the perceptual and acoustic dimension(s) that predict judgments of disorder and their likeability ratings.

Methods

Participants—Sixteen new college-aged adults rated the unaltered speech samples along several dimensions. These adults were recruited from introductory psychology and linguistics classes, and received course credit for their participation. All reported normal hearing and English as their native language.

Rating Task—Listeners rated the unaltered speech samples from Experiment 1 and 2 on a 7-point Likert scale along the 5 dimensions derived from the subjective feedback acquired in Experiment 1: articulation, speech clarity, fluency, and accenting. A fifth dimension, monotony, was included to differentiate ratings that may have been due to differences appropriately pitch accenting linguistic content (i.e., tune-to-text alignment) from those due simply to pitch variability. Listeners were only told that they would be rating children's speech along different perceptual dimensions. Listeners were also told the age range of child speakers, but no mention was made of the fact that some of the children were receiving speech and language therapy in the schools. The speech samples were explained as before with reference to the elicitation method. Only unaltered speech samples were used in Experiment 3 because these elicited the highest correlations between judgments of disorder and likeability (see Experiment 2), and because we needed to limit the number of stimuli presented to listeners in order for them to rate all dimensions within the 45 to 55 minutes allotted to "work with breaks" during the experiment.

Listeners sat in front of a computer monitor with the anchors displayed above the rating scale, which was represented as a numbered sequence of 7 buttons. The low value on the scale was anchored with a description of typical speech and the high value with a description of atypical speech. The low and high anchors for <u>articulation</u> were "1 = ALL sounds correctly produced" and "7 = NO sounds correctly produced;" for <u>clearness</u>, "1 = Extremely Clear" and "7 = Totally Unclear;" for <u>fluency</u>, "1 = Super Fluent" and "7 = Really Disfluent". The anchors for <u>accenting</u>[2] were, respectively, "1 = Emphasizes Most Important Words" and "7 = Emphasizes Only Irrelevant Words;" and for <u>monotony</u>, "1 = Very Lively" and "7 = Totally Monotonous." Listeners heard a speech sample over headphones, rated the sample on the scale along the specified dimension, clicked "OK" and advanced to a new screen and a new sample. There were two repetitions of each sample as before. All samples were presented in random order. The order in which dimensions were rated was varied across listeners. Fourteen listeners completed ratings along the articulation dimension[3]. All listeners completed ratings along all other dimensions.

---

[2]The anchors for rating along the accenting dimension were meant to draw listeners' attention to the intonational aspects of the phrase, and especially to the appropriate/inappropriate use of prosodic focus, following the literature on prosodic deficits in speech produced by individuals with ASD.

Acoustic Measurements—In order to better characterize speech sound patterning relevant to the dimensions of articulation, speech clarity, fluency, accenting, and monotony, a number of acoustic segmental and suprasegmental measures were taken on the 136 utterances that were the speech samples provided to listeners. Measurements were based on the hand-segmentation of the utterances into consonant and vowel intervals in Praat using standard segmentation criteria (see, e.g., Redford, 2014). Vowel durations, F0, F1, and F2 were automatically extracted. The measurement interval for F0 was set at .01 seconds and the range set from 75 to 600 Hz. Values that deviated one SD from a speaker's mean, were re-measured by hand. Vowel formant measures were based on the Praat LPC algorithm in time steps equal to 10 percent of the vowel interval, and with the maximum number of formants set to 5 and maximum formant frequency to 6000 Hz. Every formant track was visually inspected, then hand-corrected and re-measured if the tracks were off.

Segmental measures: Vowel duration and formant measures were used to derive 3 measures of articulation: schwa duration, mean normalized variability in schwa duration, and F1 × F2 vowel space. Vowel duration and stability in repeated vowel production are indicators of speech motor control, with shorter durations and more stable productions a feature of greater control over targeted articulations (see Redford & Oh, 2017). Schwa vowels are minimally influenced by prosodic factors since they do not receive lexical stress or phrasal accents, so mean schwa duration and variability in duration is less sensitive to contexts. There were also multiple schwa productions in every speech sample, so schwa duration and variability was something we could calculate for all speakers. The F1 × F2 vowel space was defined by full vowels that were also monopthongs. The space was calculated as the mean Euclidian distance of each full vowel from the mean full vowel in each utterance. This measure thus provides information about vowel distinctiveness within an utterance, which is a correlate of speech clarity (see, e.g., Lindblom, 1990).

Suprasegmental measures: The duration and F0 measures were used to derive measures of rate, rhythm and intonation. Articulation rate was calculated as number of syllables per second of speech, excluding any silences due to pausing, which was extremely rare in any case (see speech samples under Experiment 1). The coefficient of variation in vowel durations (i.e., standard deviation/mean) was used to capture speech rhythm. This measure has been proposed in the literature as an effective, simple-to-compute, rate-normalized rhythm metric (White & Mattys, 2007). Final lengthening, a feature of prosodic boundary marking, was calculated as the ratio of phrase-final vowel duration to the average non-final vowel duration. Another feature of prosodic boundary marking, final pitch change, was calculated as F0 change from the penultimate to ultimate syllable of the phrase. Pitch declination, a characteristic of the whole intonation contour, was calculated as the slope of a best fit regression line plotted through sequentially arranged F0 values for each utterance.

In addition to the segmental and suprasegmental measures, several acoustic measures corresponding to voice were taken to determine whether overall voice quality might have

[3]Listeners were run 2 at a time in the laboratory. On one occasion, the experimental session started late and so the listeners who participated in that session were not able to complete the task in time. Since task order was varied across sessions, both listeners in the shortened session were unable to complete the same/last task in the session.

contributed to listeners judgments of disorder (see, e.g., Shriberg et al., 2001). The measures were mean F0 and two measures of the proportion of time during a glottal period where there is no contact between the vocal folds, which is related to vocal quality ranging from vocal fry to breathy. These "openness" measures were the mean amplitude difference between the first harmonic (H1) and the second harmonic (H2) across all vowel intervals, and or the mean amplitude difference between H1 and the first formant (A1) (see Garellek & Keating, 2011). As with the F0 and formant values, phonation measures were calculated automatically using a Praat script. F0 and formant calculations were set as before.

Data Reduction—Ratings along each dimension were standardized within listener by converting the Likert scores into z-scores. Ratings were then averaged across listeners and within speakers. Inter-correlations between the rated dimensions were assessed. Ratings along the dimensions of articulation, clearness, and fluency were exceptionally highly correlated, with correlation coefficients ranging from .92 (articulation and fluency) to .96 (articulation and clearness). Accordingly, these 3 dimensions were reduced to one using principal components analysis (PCA). Scores from the first factor of the PCA were saved and used in place of the articulation, clearness, and fluency ratings. For ease of reference, we will refer to these scores as capturing the dimension of *intelligibility*. This descriptive label was chosen because the constituent components of this shared dimension (e.g., articulation, clearness, and fluency) are all attributes of intelligible speech. Intelligibility, accenting, and monotony were then entered into a multiple regression model to predict judgments of disorder (see below). This model did not suffer from unacceptable collinearity (VIF < 3.5).

Inter-correlations between the acoustic measures were also assessed. Many did not reach significance. All and only the ones that did are shown in Table 5. Note the near perfect correlation between the two measures related to voice quality in the table. Given the high degree of overlap, these measures were reduced to one using PCA. Scores from the first factor of the PCA were saved and used in place of the individual H1-H2 and H1-A1 measures. For ease of reference, we will refer to these scores as measures of *voice quality*.

Statistical Analyses—Stepwise linear regression modeling was used to investigate the relative contribution of listener-identified perceptual dimensions of speech sound patterning to residualized "disordered" judgments and z-scored likeability ratings on unaltered speech samples. Age-in-months and receptive vocabulary scores were included as control variables in the model. Backward elimination was used to reduce the number of predictor variables to only those that contributed significantly to explaining listeners' responses. The same technique was then used to characterize in acoustic terms the perceptual dimension(s) that best explained listeners' behavior in Experiments 1 and 2.

Results

Perceptual Basis for Judgments of Disorder—The full model accounted for 69% of the variance in judgments of disorder and had an adjusted R^2 of .63. The significant model with the fewest predictors accounted for 67% of the variance and had an adjusted R^2 of .65. The only significant predictors in this model were age-in-months, $\beta = .22$, $t(33) = 2.10$, $p = .045$, and intelligibility, $\beta = .84$, $t(33) = 7.98$, $p < .001$; that is, the perceptual dimension

derived from the overlap in ratings of articulation, clearness, and fluency. Whereas intelligibility by itself accounted for 63% of the variance in "disordered" judgments, the model with age-in-months alone was no better than the null model at accounting for variance. Figure 5 shows the relationship between poor intelligibility[4] and "disordered" judgments.

An independent samples t-test confirmed that variance along the intelligibility dimension also varied systematically with the group of speakers who produced the samples, Mean Difference = 1.02, $t(33) = 3.40$, $p = .002$ (two-tailed). Although children with ASD were rated as less intelligible on average than their TD peers, it is also evident from the data presented in Figure 5 that not all children were rated as equally unintelligible nor were they all rated as equally "disordered".

Perceptual Basis for Likeability Ratings—The full model accounted for 57% of the variance in likeability ratings and had an adjusted R^2 of .49. The significant model with the fewest predictors accounted for 49% of the variance and had an adjusted R^2 of .46. The only significant predictors in this model were intelligibility, $\beta = .56$, $t(33) = 4.12$, $p < .001$, and monotony, $\beta = .29$, $t(33) = 2.15$, $p = .040$. Intelligibility accounted for 42% of the variance in likeability ratings by itself, and monotony for 21%. Thus, it would seem that intelligibility and monotony each accounted for some of the same variance in the regression model. Figure 6 shows the relationship between intelligibility and likeability, where higher values along each axis indicate more negative evaluations.

Acoustic Characterization of Intelligibility and Monotony—Segmental, suprasegmental, and voice measures were entered as predictor variables in a stepwise regression models to provide acoustic characterizations of the perceptual dimensions, intelligibility and monotony. The full models accounted for 46% of the variance in intelligibility and for 68% of the variance in monotony; they had adjusted R^2s of .20 and .55, respectively. The significant models with the fewest predictors accounted for 41% of the variance in intelligibility and for 56% of the variance in monotony; they had an adjusted R^2s of .17 and .50, respectively. Articulation rate was the sole significant predictor of intelligibility, $\beta = -.41$, $t(33) = -2.55$, $p = .016$: slower rates were associated with poorer intelligibility. By contrast, the reduced model of monotony included two significant segmental measures—mean schwa duration, $\beta = .31$, $t(33) = 2.54$, $p = .016$, and vowel distinctiveness, $\beta = -.25$, $t(33) = -2.04$, $p = .050$—and two significant voice measures—mean F0, $\beta = -.37$, $t(33) = -2.45$, $p = .020$, and voice quality, $\beta = .44$, $t(33) = 2.91$, $p = .007$. The direction of the effect of schwa duration is consistent with the effect of speech rate: longer schwa were associated with greater perceived speech monotony than shorter schwas. The effect of vowel distinctiveness indicates that more monotonous speech was also likely less clear. Lower pitch and less modal (more breathy) voice qualities were also associated with more monotonous speech. Given this list of significant predictor variables, it would seem that although the perceptual dimensions of intelligibility and monotony are

---

[4]Recall that the 7-point rating scale was anchored such that higher scores are more compatible with descriptions of disordered speech. (e.g., "1 = ALL sounds correctly produced" versus "7 = NO sounds correctly produced").

acoustically separable, both are well-characterized by features that are more associated with speech motor and voice factors than with linguistic prosody.

Discussion

The results from Experiment 3 strongly suggest that lay listeners attend to patterns related to motor speech rather than to prosodic ones when they are asked to classify children's speech sound patterns as "disordered" or "typical" and to evaluate how much they like a child based on short speech samples. This conclusion is based on the finding that "accenting," the most overtly prosodic perceptual dimension evaluated, was not a significant predictor of judgments of disorder or of likeability. Moreover, the strongest predictor of both perceived disorder and likeability was intelligibility, a dimension derived from the overlap in ratings of articulation, clarity, and fluency. This dimension was best characterized by articulation rate. Rate may be a suprasegmental feature of speech, but it is one more typically associated with articulatory timing control than with prosody (see, e.g., Redford, 2014). Similarly, the dimension "monotony" was best characterized by measures that captured the accuracy and speed of segmental articulations (schwa duration and vowel distinctiveness) and by measures related to voice characteristics (pitch and voice quality) that are also outside the domain of prosody, strictly defined.

GENERAL DISCUSSION

Although children and adults with ASD are often reported to have atypical speech patterns, many individuals with autism show no prosodic deficits (see Shriberg et al., 2001; 2011; Nadig & Shaw, 2012). Moreover, the behaviors that are central to establishing a diagnosis of autism are more obviously socially disabling (e.g., not responding to one's own name) and deviant (e.g., echolalic speech) than we might imagine unusual speech patterns to be. The current study was motivated by the question of whether atypical speech patterns in children with ASD is important enough to warrant research. In a clinical or educational context, the notion of what behavior is or is not important to understand amounts to a question about the effects that such behavior has on peer relations or on learning. The results from the present study provide evidence in support of the view that atypical speech sound patterns mark children with ASD as different from their TD peers, and are likely to create barriers to positive social interactions. We conclude from this that atypical speech should be addressed directly in children with ASD. The results do not, however, support the view that prosody should necessarily be the focus of these efforts. When input is limited, listeners' judgments are influenced by speech patterns that likely reflect delayed or deviant speech motor control and unusual voice quality rather than prosodic deficits per se. More specifically, the current study identifies intelligibility as the primary perceptual dimension that mediates both listeners' judgments of disorder and their social evaluation of the child based on short snippets of speech. This dimension is closely associated with the perceived goodness and fluency (= speed) of segmental articulation.

Intelligibility is a concept fundamental to studies of motor speech disorders (e.g., dysarthria), where it is usually measured as the number of distinct speech sounds or words recognized by a panel of listeners (see e.g., Kent et al., 1989). We would therefore like to

clarify that none of the children who participated in the present study sounded dysarthric. Their speech was not distorted, just different sounding. This difference was characterized by the referring SLPs as unusual prosody, but listeners' behavior suggests that perhaps speech motor factors, rather than linguistic prosodic ones, were more important to the perception of difference. It is for this reason that we would argue that the notion of intelligibility better captures what listeners were most sensitive to when distinguishing children with ASD from their typically developing peers based on the short speech samples provided. Additional evidence for this view comes from the feedback that listeners provided in Experiment 1. For example, Listeners 1, 2, and 15 all note that their judgments were affected by what they perceived as "slurred" or "slurring." Listeners 11 and 13 refer to the influence of a "lisp" on their judgments. Listeners 3, 4, 5, 7, 8, 11, and 15 comment on aspects of rate. Relatedly, articulation rate was a significant predictor of intelligibility. Elsewhere, we have found elsewhere that default articulation rates index the acquisition of articulatory timing control more broadly, at least in typically developing children (Redford, 2014).

The conclusion that listeners may have focused on speech motor factors to distinguish between children with ASD and TD is at odds with the characterization of autistic speech as prosodically disordered if one understands speech prosody in linguistic terms; that is, as a system of metrical grouping and accent placements governed by a prosodic grammar that is linked, at some level, to syntactic structure and semantic-pragmatic meaning. Deficits in lexical stress placement and contrastive focus meet this definition of prosody and so are consistent with the primary deficits in semantic-pragmatic communication that is associated with ASD, but these were not found here. Moreover, the differences we did find correspond to the many other documented differences between autistic speech and typical speech that are not necessarily prosodic. In particular, acoustic-phonetic studies that characterize differences between ASD and typical speech in terms of global differences in pitch range and variability (Diehl et al., 2009; Nadig & Shaw, 2012), longer word and utterance durations (Diehl & Paul, 2012), and higher rates of disfluencies (Shriberg et al., 2001) are all compatible with differences in planning and production and not with prosody per se.

The findings that listeners are sensitive to speech intelligibility when judging short snippets of speech and that this sensitivity affects their social evaluation of a speaker also has implications for conceptualizing the semantic-pragmatic impairments in children with ASD. In particular, a focus on intelligibility emphasizes the importance of the social interaction for defining "impaired", "atypical" or "disordered", and so supports a view of pragmatic impairment as emergent from the interaction between speakers and listeners (see Perkins, 2007). When the contributions of both are in balance, communication is successful and pragmatics are intact. When one member of the dyad must adapt in an unfamiliar, asymmetrical manner to accommodate the other's behavior, then communication suffers and pragmatics is impaired. This view of impairment has the advantage of allowing us to understand why listeners' perception of disorder might affect their social evaluation of the speaker. Any decrement in speech intelligibility is likely to result in communicative asymmetry; namely, an asymmetry that creates extra work for the listener. If listeners (implicitly) resent doing extra work, then they may blame the speaker. Note that this explanation is similar to that which Rice and colleagues proposed to explain their findings that children with specific language impairment or who are second language speakers of

English have less access to positive social interactions than their typically developing and native English-speaking peers in US schools (Rice et al., 1991; Gertner et al. 1994).

Although we have suggested that listeners are more sensitive to speech sound patterns that suggest deficits in speech motor control rather than in prosody, it is important to note that the strength of this conclusion is limited by the experimental design we used. Specifically, the short, decontextualized speech samples presented to listeners may have forced them to latch on to something other than prosody to make their decisions. If unusual prosody in autism often reflects inappropriate or incorrect prosodic focus (Shriberg et al., 2001; Paul, Augustyn, et al, 2005; Peppé et al., 2007), then access to language context would be critical for listeners to identify the atypical patterns. In fact, one way to interpret the finding that listeners could only distinguish children with ASD from their typically developing peers in the unaltered speech condition is that only this condition provided listeners with the relevant information about tune-to-text alignment. Moreover, some listeners did comment that *how* a speaker used accenting influenced their judgments of disorder. For example, Listener 4 in Experiment 1 noted that their judgment was affected by whether the child "stressed the right syllables" and Listener 16 was sensitive to whether or not emphasis was used correctly. Future work should evaluate the relative influences of intelligibility and ungrammatical prosodic focus assignment and phrasing on listeners' behavior by using longer stretches of speech that provide a discourse context.

In conclusion, the results from the present study show that lay listeners are sensitive to the atypical speech patterns produced by children with ASD. This sensitivity also influences their social evaluation of the speaker. Speakers who produced patterns that were perceived as atypical or disordered and rated as less likeable. These findings validate efforts to understand and treat atypical speech in ASD, even in the context of the many other disabling behaviors associated with the disorder. The present study also adds to a body of literature on listener sensitivity to social-indexical aspects of speech, but suggests that some negative biases towards speakers may emerge from the modeling effort that is needed for listeners to engage in successful communication with an unfamiliar speaker.

Acknowledgments

References

American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. Washington, D.C.: American Psychiatric Association; 2013.

Baltaxe CA, Guthrie D. The use of primary sentence stress by normal, aphasic, and autistic children. Journal of Autism and Developmental Disorders. 1987; 17:255–271. [PubMed: 2440848]

Boersma, P., Weenink, D. Praat, version 5.1.16. 2009. Available at http://www.praat.org

Bonneh YS, Levanon Y, Dean-Pardon O, Lossos L, Adini Y. Abnormal speech spectrum and increased pitch variability in young autistic children. Frontiers in Human Neuroscience. 2011; 4:2347.

Clopper CG, Bradlow AR. Free classification of American English dialects by native and non-native listeners. Journal of Phonetics. 2009; 37:436–451. [PubMed: 20161400]

Clopper CG, Pisoni DB. Some acoustic cues for the perceptual categorization of American English regional dialects. Journal of Phonetics. 2004; 32:111–140. [PubMed: 21451736]

Depape AM, Chen A, Hall GB, Trainor LJ. Use of prosody and information structure in high functioning adults with autism in relation to language abilitiy. Frontiers in Psychology. 2012; 3:1664–1678.

Diehl JJ, Watson D, Bennetto L, McDonough J, Gunlogson C. An acoustic analysis of prosody in high-functioning autism. Applied Psycholinguistics. 2009; 30:385–404.

Diehl JJ, Paul R. Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders. Research in Autism Spectrum Disorders. 2012; 6:123–134. [PubMed: 22125576]

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

Frith, U. Translation and annotation of "Autistic psychopathy" in childhood, by H Asperger. In: Frith, U., editor. Autism and Asperger syndrome. Cambridge, UK: Cambridge University Press; 1991.

Garellek M, Keating P. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. Journal of the International Phonetic Association. 2011; 41:185–205.

Gertner BL, Rice ML, Hadley PA. Influence of communicative competence on peer preferences in a preschool classroom. Journal of Speech, Language and Hearing Research. 1994; 37:913–923.

Grossman RB, Bemis RH, Plesa Skwerer D, Tager-Flusberg H. Lexical and affective prosody in children with high-functioning autism. Journal of Speech, Language and Hearing Research. 2010; 53:778.

Kent RD, Weismer G, Kent JF, Rosenbek JC. Toward phonetic intelligibility testing in dysarthria. Journal of Speech and Hearing Disorders. 1989; 54:482–499. [PubMed: 2811329]

Klin, A., Sparrow, S., Marans, WD., Carter, A., Volkmar, FR. Assessment issues in children and adolescents with Asperger Syndrome. In: Klin, A.Volmar, FR., Sparrow, SS., editors. Asperger syndrome. New York: Guilford Press; 2000. p. 309-339.

Lass NJ, Almerino CA, Jordan LF, Walsh JM. The effect of filtered speech on speaker race and sex identifications. Journal of Phonetics. 1980; 8:101–112.

Lindblom, B. Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, WJ., Marchal, A., editors. Speech Production and Speech Modelling. The Netherlands: Kluwer Academic; 1990. p. 403-439.

McCaleb P, Prizant BM. Encoding of new versus old information by autistic children. Journal of speech and Hearing Disorders. 1985; 50:230. [PubMed: 4021450]

Moore, BCJ. An introduction to the psychology of hearing. 3rd. London: Academic Press; 1994.

Munson B. The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. Language and Speech. 2007; 50:125–142. [PubMed: 17518106]

Munson B, McDonald EC, DeBoe NL, White AR. The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. Journal of Phonetics. 2006; 34:202–240.

Nadig A, Shaw H. Acoustic and perceptual measurement of expressive prosody in high-functioning autism: increased pitch range and what it means to listeners. Journal of Autism and Developmental Disorders. 2012; 42:499–511. [PubMed: 21528425]

Paul R, Augustyn A, Klin A, Volkmar FR. Perception and production of prosody by speakers with autism spectrum disorders. Journal of Autism and Developmental Disorders. 2005; 35:205–20. [PubMed: 15909407]

Paul R, Bianchi N, Augustyn A, Klin A, Volkmar FR. Production of syllable stress in speakers with autism spectrum disorders. Research in Autism Spectrum Disorders. 2008; 2:110–124. [PubMed: 19337577]

Peppé S, McCann J. Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. Clinical Linguistics and Phonetics. 2003; 17:345–354. [PubMed: 12945610]

Peppé S, McCann J, Gibbon F, O'Hare A, Rutherford M. Receptive and expressive prosodic ability in children with high-functionning autism. Journal of Speech, Language, and Hearing Research. 2007; 50:1015–1028.

Perkins, M. Pragmatic impairment. Cambridge University Press; 2007.

Port RF. Meter and speech. Journal of Phonetics. 2003; 31:599–611.

Purnell T, Idsardi W, Baugh J. Perceptual and phonetic experiments on American English dialect identification. Journal of Language and Social Psychology. 1999; 18:10–30.

Redford MA. A comparative analysis of pausing in child and adult storytelling. Applied Psycholinguistics. 2013; 34:569–589. [PubMed: 23772097]

Redford MA. The perceived clarity of children's speech varies as a function of their default articulation rate. Journal of the Acoustical Society of America. 2014; 135:2952–2963. [PubMed: 24815275]

Redford MA, Oh G. The representation and execution of articulatory timing in first and second language acquisition. Journal of Phonetics. 2017; doi: 10.1016/j.wocn.2017.01.004

Rice ML, Sell MA, Hadley PA. Social interactions of speech, and language-impaired children. Journal of Speech, Language and Hearing Research. 1991; 34:1299.

Schopler, E., Van Bourgodien, ME., Wellman, GJ., Love, SR. Childhood Autism Rating Scale, Second Edition (CARS2). Torrance, CA: WPS; 2010.

Sharda M, Subhadra TP, Sahay S, Nagaraja C, Singh L, Mishra R, Singhalc N, Erickson D, Singh NC. Sounds of melody–Pitch patterns of speech in autism. Neuroscience Letters. 2010; 478:42–45. [PubMed: 20447444]

Shriberg, LD., Kwiatkowski, J., Rasmussen, C., Lof, G., Miller, J. The Prosody-Voice Screening Profile (PVSP): Psychometric Data and Reference Information for Children. University of Wisconsin-Madison, Waisman Center on Mental Retardation and Human Development; 1992. Phonology Project Technical Report 1

Shriberg LD, Paul R, Black LM, van Santen JP. The hypothesis of apraxia of speech in children with autism spectrum disorder. Journal of Autism and Developmental Disorders. 2011; 41:405–426. [PubMed: 20972615]

Shriberg LD, Paul R, McSweeny JL, Klin A, Cohen DJ, Volkmar FR. Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. Journal of Speech, Language and Hearing Research. 2001; 44:1097.

Tilsen S, Johnson K. Low-frequency Fourier analysis of speech rhythm. Journal of the Acoustical Society of America. 2008; 124:EL34–EL39. [PubMed: 18681499]

Van Bezooijen R. Sociocultural aspects of pitch differences between Japanese and Dutch women. Language and Speech. 1995; 38:253–265. [PubMed: 8816084]

White L, Mattys SL. Calibrating rhythm: first language and second language studies. Journal of Phonetics. 2007; 35:501–522.
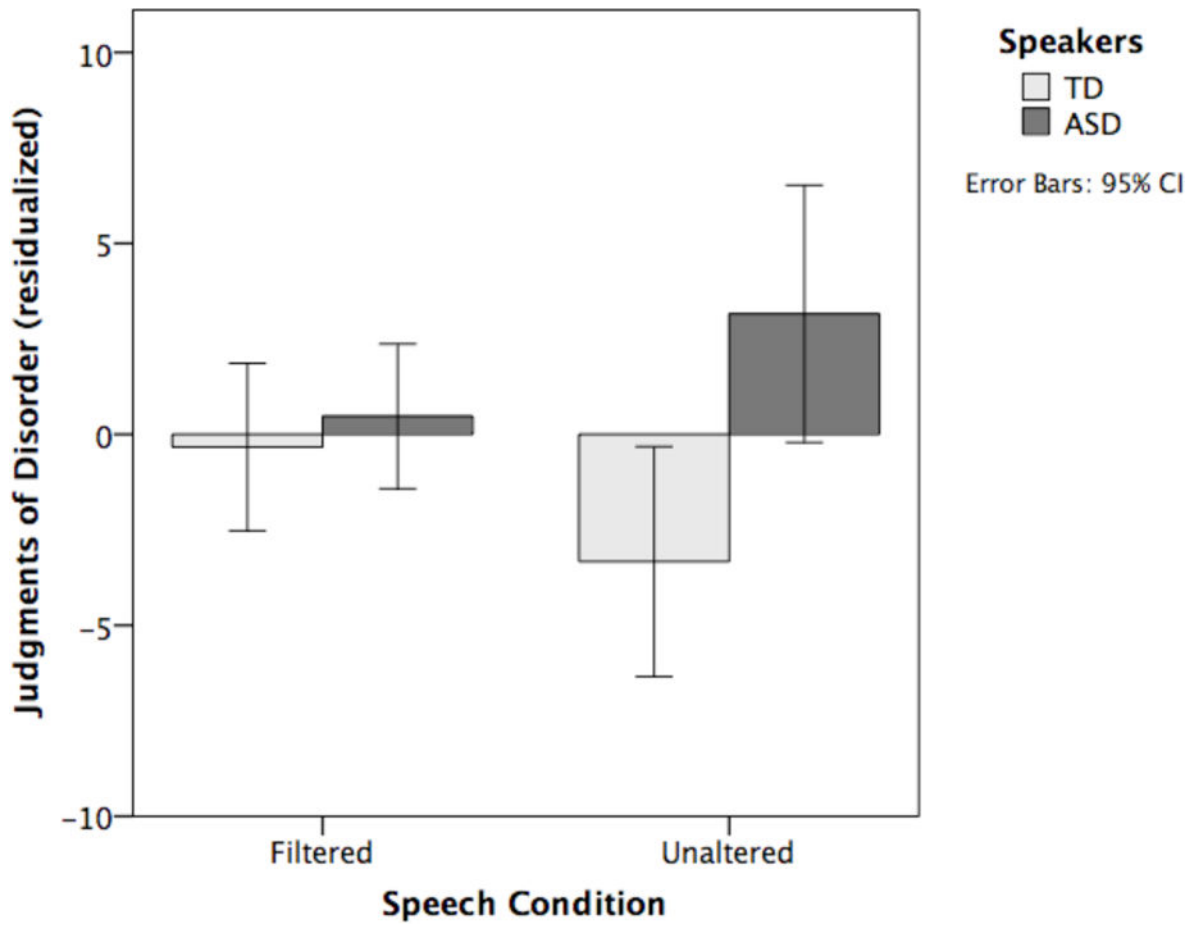
Figure 1.
Listeners' judgments of disorder, absent effects of language, are shown as a function of speech condition and speaker group (TD = typically developing children; ASD = children with ASD and prosodic disorder).
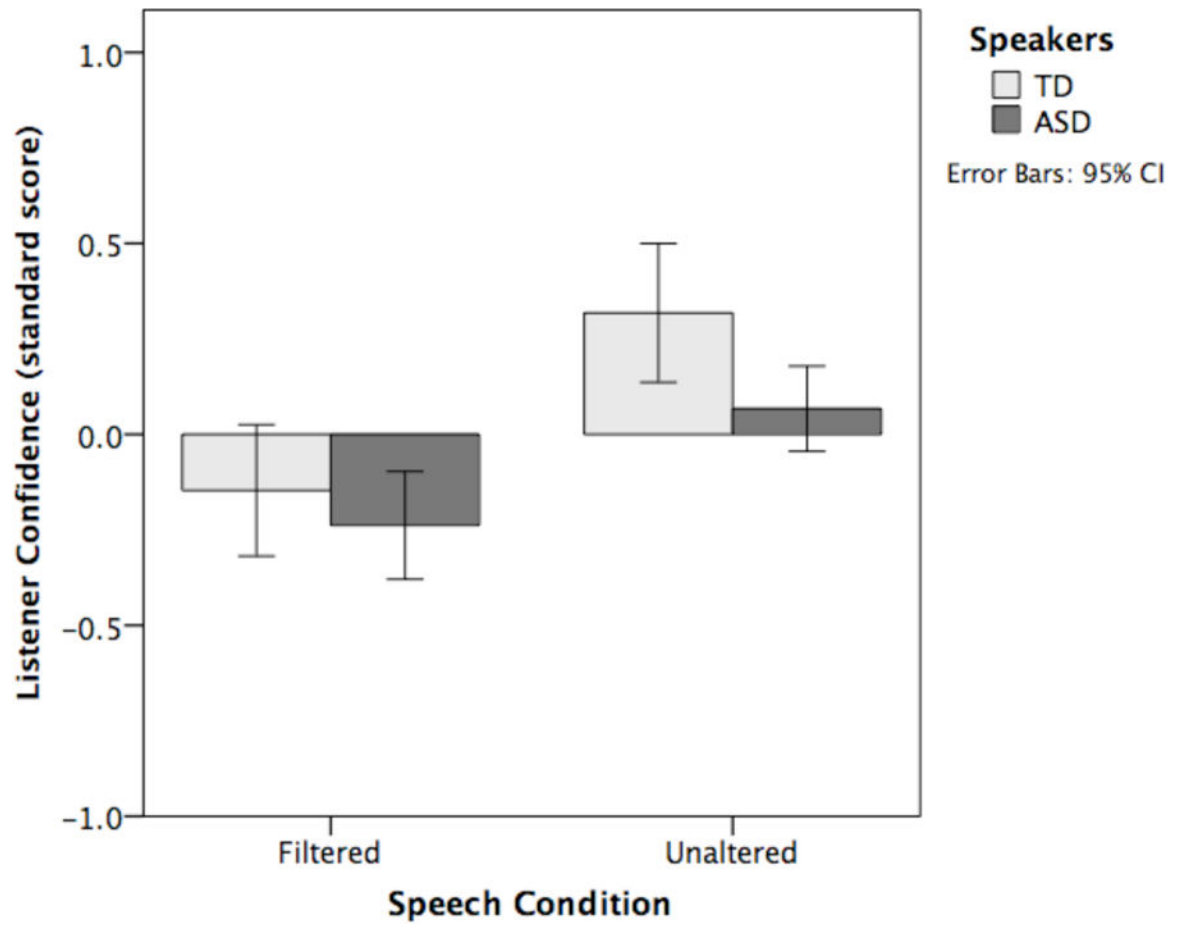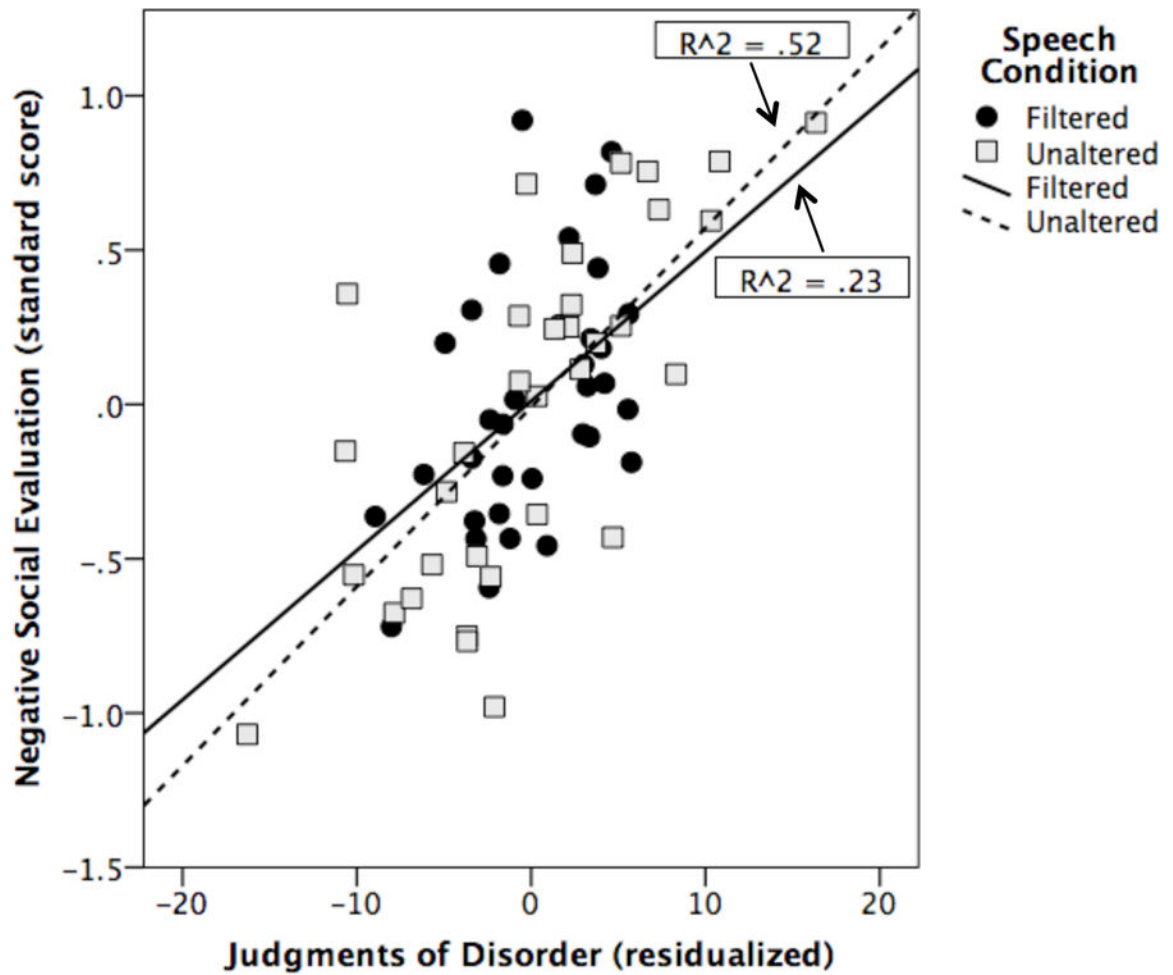
Figure 2.
Listeners' confidence in their judgments of disorder are shown as a function of speech
condition and speaker group (TD = typically developing children; ASD = children with ASD
and prosodic disorder).
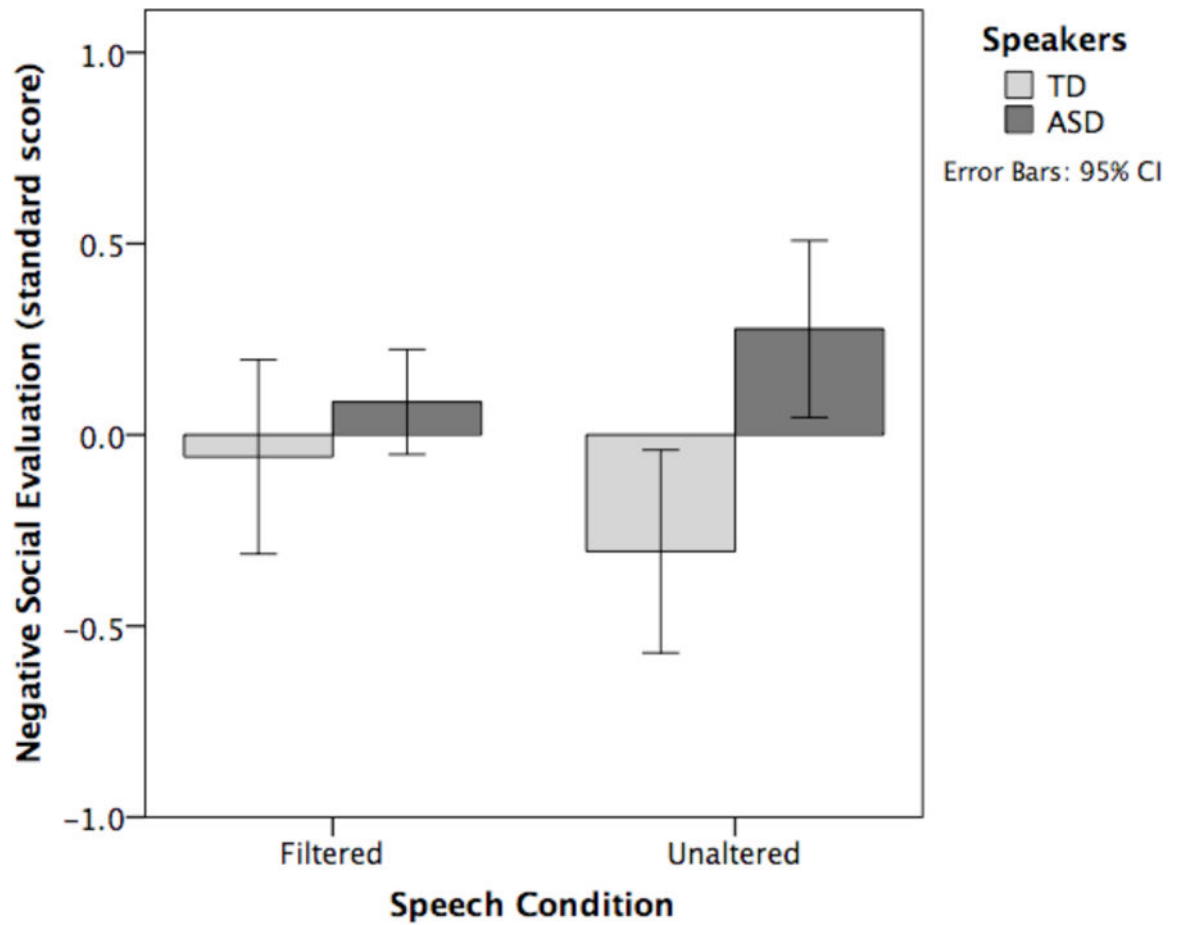
Figure 3.
Listeners' ratings of a child's likeability (low anchor = "Awesome! Love this kid."; high anchor = "Nah. Wouldn't like this kid.") shown as a function of speech condition and speaker group (TD = typically developing children; ASD = children with ASD and prosodic disorder).

Figure 4.
Listeners' judgments of disorder based on filtered and unaltered speech samples predicts another group's likeability ratings of the speakers who produced those samples. Low values along the y-axis indicate greater likeability than high values.
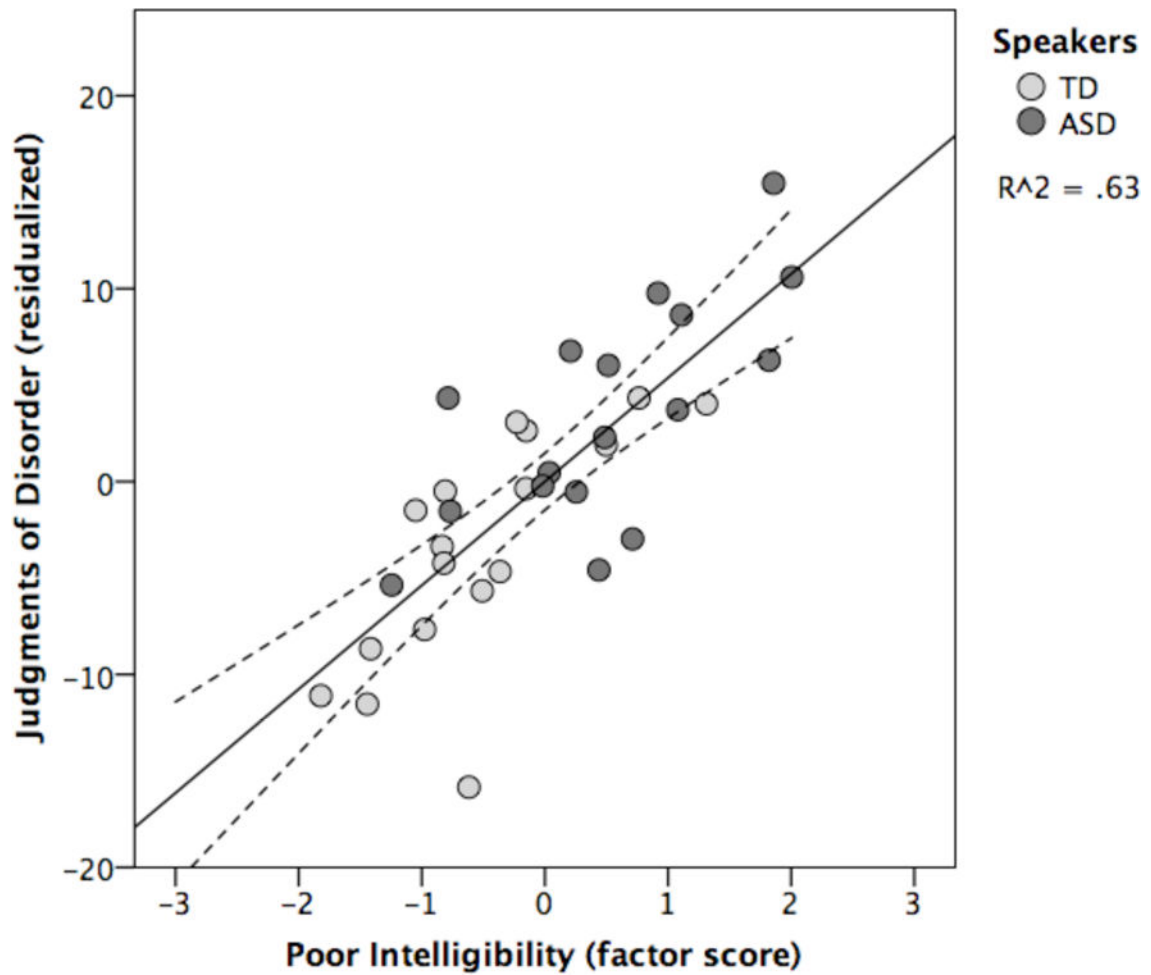
Figure 5.
The intersection of listeners' ratings of the articulation, clearness, and fluency of unaltered speech samples (= intelligibility) predicts another group's judgments of disorder based on those same speech samples. Text-based judgments of disorder are partialled out from the values shown along the y-axis.
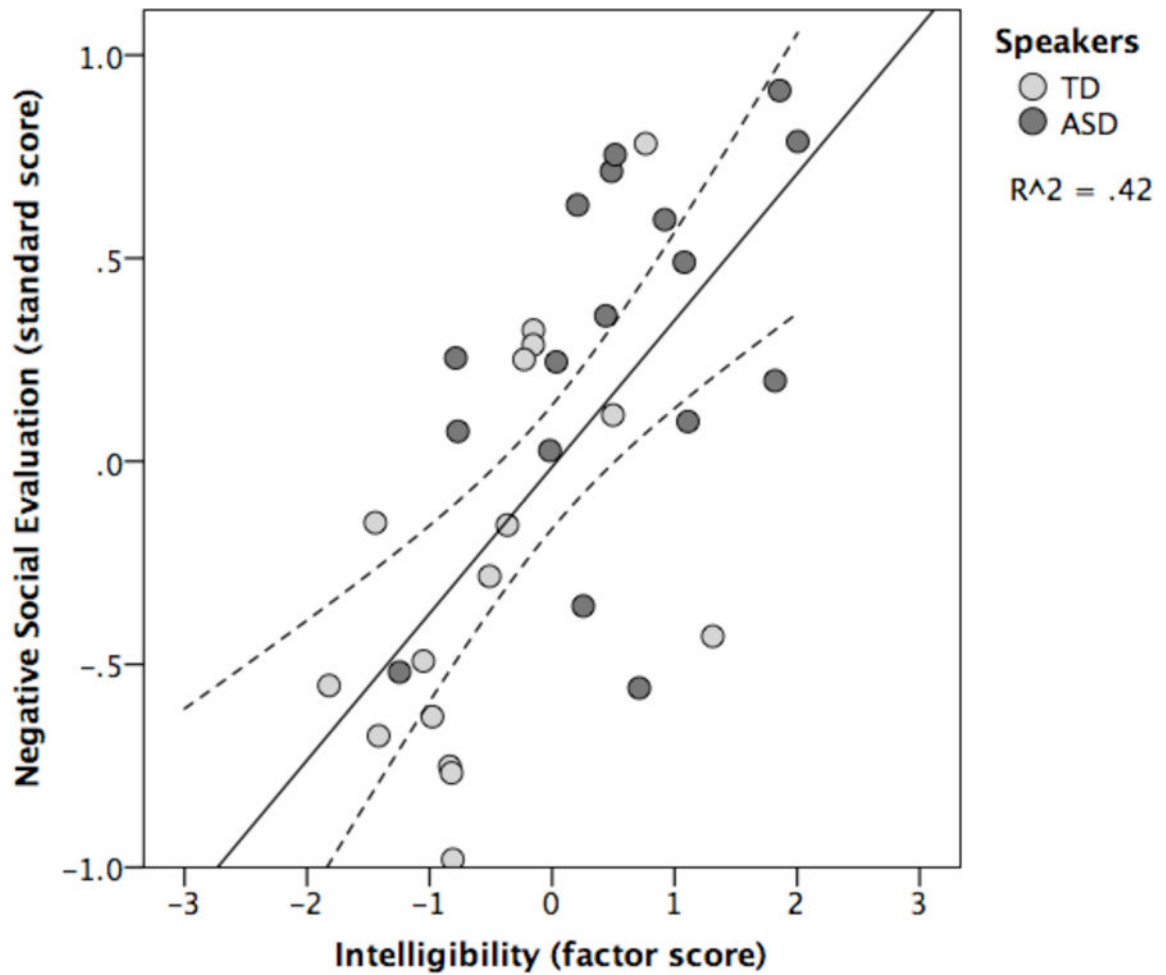
Figure 6.
The intersection of listeners' ratings of the articulation, clearness, and fluency of unaltered speech samples (= intelligibility) predicts another group's likeability ratings of the speakers who produced those samples. Low values along the y-axis indicate greater likeability than high values.

TABLE 1

Characteristics of participants with SPED eligibilities for receiving speech and language services in the schools under ASD.

Speaker	Sex	Age	Symptoms	Lg. Delay	PPVT
p109asd	M	6;4	severe	No	106
p110asd	M	7;1	minimal	Yes	119
p102asd	M	7;9	severe	Yes	126
p100asd	M	7;11	severe	Yes	83
p103asd	M	8;1	moderate	Yes	85
p107asd	M	8;1	severe	No	107
p108asd	M	8;1	moderate	Yes	112
p114asd	M	8;1	minimal	Yes	113
p115asd	M	8;10	severe	No	86
p101asd	F	9;3	moderate	No	83
p111asd	F	9;4	severe	No	80
p118asd	M	9;6	severe	Yes	104
p119asd	M	9;6	severe	Yes	78
p106asd	M	10;3	severe	Yes	94
p112asd	M	10;9	minimal	Yes	102
p113asd	F	11;7	severe	Yes	90
p116asd	M	11;8	severe	Yes	94

TABLE 2

Characteristics of participants with typical development.

Speaker	Sex	Age	PPVT
p1032td	M	6;4	112
p1037td	M	7;2	121
p1061td	M	7;7	90
p1073td	M	7;11	106
p1011td	M	8;0	117
p1083td	M	8;0	147
p1045td	F	8;1	129
p1072td	M	8;2	114
p1016td	M	9;0	134
p1058td	F	9;2	103
p1075td	F	9;4	88
p1026td	M	9;5	110
p1027td	M	9;7	129
p1006td	F	10;0	111
p1010td	M	10;1	111
p1001td	F	10;10	135
p1022td	M	10;10	130

TABLE 3

Mean (and standard deviation) of untransformed and unstandardized responses are shown as a function of speaker group (TD = typically developing). Responses of disorder obtained in the filtered speech, unaltered speech, and text-based judgment tasks were aggregated across samples and listeners to render a by-speaker perceived disorder score (maximum score = 32 for speech-based judgments; 16 for text-based judgments). Comprehension of filtered speech was rated on a 5-point scale (no words understood = 1; all words understood = 5). Listeners also rated how confident they were in their responses (not at all = 1; completely sure = 5).

Judgment Task	Response type	Children w/TD	Children w/ASD
Filtered speech (Experimental)	Cumulative "disordered"	12.53 (7.08)	14.71 (6.45)
	confidence	3.27 (0.31)	3.20 (0.25)
Comprehension (Control)	ratings from 1 to 5	2.32 (0.64)	2.10 (0.50)
	confidence	3.74 (0.89)	3.61 (0.91)
Unaltered speech (Experimental)	Cumulative "disordered"	7.18 (5.27)	18.35 (7.25)
	confidence	3.67 (0.32)	3.47 (0.19)
Text (Control)	Cumulative "disordered"	4.29 (2.80)	7.24 (3.54)
	confidence	3.41 (0.90)	3.38 (0.93)

TABLE 4

Summary of individual listener's responses to the request for feedback on what speech characteristics influenced their judgments of disorder in the filtered and unaltered speech conditions. The responses are categorized by major theme.

Listnr	Articulation	Speech Clarity	Fluency	Accenting
1	*mispronounciation*	*slurring of words*	*fluidity and … fluctuation*	*appropriate flux … to accent*
2		*slurred words, concise pronunciation*	*choppy, flow, pausing at awkward times*	*emphasis of a word, enthusiastic*
3		*clarity, unclear*	*slower speech, streaming words, rhythm*	
4			*steady rhythm, quickly, slowly, smoothly*	*stressed the right syllables, accents*
5		*enunciated*	*slower, sort of flow*	*up and down, placed emphasis*
6	*distorted "r" and "l" sounds, oddly pronounced blends*			
7	*mispronounced words, sounds*		*speech rate*	*stress patterns*
8		*slurred, clearer, easier to understand*	*little slower, faster*	*enthusiasm*
9			*pauses between… phrases*	*mid-phrase inflections*
10			*sounded disjoint*	*emphasize certain points, excitement*
11	*lisp, "r" sounds*	*speak effectively*	*dragged their syllables out*	
12		*muffled, extremely unclear*		
13	*lisp, pronounciating*	*understand…clearly*		
14			*natural flow, rhythm, pace of words*	*tone, clashy*
15		*clear and precise, slurred*	*quickly*	
16				*disorganized tones, emphasized for no purpose*

TABLE 5

All and only significant pair-wise correlations between acoustic measures.

	Schwa Dur.	Artic. Rate	F0 Slope	Final F0	Mean F0	H1 – H2	H1 – A1
Schwa Dur.	1	–.43 *					
Artic. Rate		1					
F0 Slope			1	–.52 **	–.39 *	–.44 **	–.53 **
Final F0 Change				1			
Mean F0				–.41 *	1	–.61 **	–.56 **
H1–H2						1	.95 **
H1–A1							1

*
p < .05;

**
p < .001