Perceptual Characteristics of Spontaneous Speech in Autism: Are We Discriminating Listeners?

Mary V. Andrianopoulos, Ph.D.*
Elena Zaretsky, Ph.D.†
Carlene McGuigan, B.S.*
Rachel Warshaw, B.S.*

1. Introduction

The Centers for Disease Control (CDC) suggest that one in 68 children will be diagnosed with Autism Spectrum Disorder (ASD) in the United States (CDC 2014a). Signs of autism are evident in early childhood and range from unusual vocal and social development, lack of social interest and social responsiveness, absence of eye contact, fewer vocalizations and imitation of sounds (Volkmar, Chawarska, & Klin, 2005; Sheinkopf, Mundy, Oller, & Steffens, 2000). Parents of children with autism report slow and atypical speech and language development, including nonverbal aspects of communication such as, gestures and facial expression (Tager-Flusberg, Paul, & Lord, 2005; Volkmar et al., 2005). Unusual behaviors and developmental delays become more apparent from infancy to three years of age.

The DSM-5 (APA, 2013) lists persistent deficits in social aspects of communication and interaction, as well as restricted, repetitive patterns of behavior as characteristics associated with the autism spectrum disorder (ASD). Language abilities are one area that is very variable across the spectrum, with skills ranging from non-existent to rather sophisticated. However, some distinct features of ASD speech and language may be present across the entire spectrum such as, abnormal prosody, structural language deficits, and problems with narrative coherence (Boucher, 2012; Eigsti, Bennetto, & Dadlani, 2007; Kjelgaard & Tager-Flusberg, 2001; McCabe, Hillier & Shapiro, 2013; Nadig & Shaw, 2012).

Published research supports the presence of speech characteristics that contribute to the perception of “oddness” in individuals with ASD (Paul, Augustyn, Klin, & Volkmar, 2005). However, it is not clear what the underlying mechanisms are that contribute most to the acoustic form of “speech atypicalities” among speakers with ASD compared to the speech of their

* University of Massachusetts Amherst, Department of Communication Disorders, 358 North Pleasant St. Amherst, MA. 01003. Email: mva@umass.edu
† Clark University, Department of Psychology, Worcester, MA. 01610
neurotypically developing (NTD) peers. It is also not clear whether the “monotonous” and atypical prosody observed in the speech of children with ASD are due to cognitive-linguistic, perceptual, motoric, or prosodic factors, or a combination thereof (Nadig & Shaw, 2012).

Some researchers attribute the speech atypicalities observed among some children with ASD to impaired perspective-taking skills (Theory of Mind, ToM), which enable one to correctly interpret different cues, accurately assess social situations and determine how to best respond (Baron-Cohen, 1991; Joliffe & Baron-Cohen, 1999). Research is lacking whether deficits in ToM or issues related the “speech attunement framework” affect the acoustic-perceptual or prosodic qualities of verbal expression among individuals with autism (Shriberg, Paul, Black, Santen, 2011).

Moreover, approximately 30-50% of individuals diagnosed with ASD never gain the ability to speak (Prizant, 1996; Gernsbacher, Sauer, Geye, Schweigert, & Hill Goldsmith, 2008, Peppe, McCann, Gibbon, O’Hare, & Rutherford, 2007). Atypical motor skills, poor motor imitation, groping and difficulty with both non-verbal oral and verbal motor movements, increased instances of repetitions, revisions, and speech errors have been empirically demonstrated among children with ASD (Adams, 1998; Peppe, McCann, & Gibbon, 2007; Shriberg, Paul, Black, Santen, 2011; Belmonte, Saxena-Chandhok, Cherian, Munee, George, & Karanth, 2013). Abnormal speech of individuals with ASD varies from individual to individual and underlying motoric issues may contribute to these speech differences (Bonneh et al. 2011, Diehl & Paul, 2013).

With respect to the acoustic perceptual features of speech produced by children with ASD compared to their NTD counterparts, research also supports the presence of increased loudness, high pitch, exaggerated stress, “machine-like” or sing song speech, fast or slow rate, atypical and shallow vocal quality, speech distortions, and differences in duration of syllables (Bonneh, Levanon, Dean-Pardo, Lossos, & Adini, 2010; Shriberg, Paul, Black, & van Santen, 2011; Grossman & Tager-Flusberg, 2012; Nadig & Shaw, 2012; Belmonte, Saxena-Chandhok, Cherian, Munee, & Karanth, 2013; Grossman, Edelson, & Tager-Flusberg, 2013; Filipe, Frota, Castro, & Vicente, 2014). Prosodic function and form atypicalities are also evident among children with ASD that include differences in identifying and producing statements and questions (turn-end functions) and contrastive stress, interpreting and producing affect and prosodic phrase breaks, and poor auditory discrimination (Diehl, Watson, Bennett, McDonough, & Gunlogson, 2009; Bonneh, Levanon, Dean-Pardo, Lossos, & Adini, 2010; Grossman, Bermis, Plesa Skwerer, & Tager-Flusberg, 2010; Peppé, Cleland, Gibbon, O’Hare, & Martinez Castilla, 2011; Nadig & Shaw, 2012; Diehl & Paul, 2013).

Published empirical findings are inconsistent regarding the underlying mechanisms that contribute to the speech atypicalities or the vocal awkwardness observed among children with ASD. It is not clear if the speech atypicalities of
children with ASD are due to cognitive-linguistic, perceptual, motor speech, articulation, vocal, or receptive and expressive prosodic differences.

Regardless of the possible underlying mechanisms for the unusual speech characteristics that contribute to the perception of speech “oddness” in individuals with ASD (Paul, Augustyn, Klin, & Volkmar, 2005), it is difficult to identify precisely what sounds different to the human ear. Most research investigations to date have focused on dissecting the speech and language of children with ASD with respect to the length and quality of their sentences and verbal expression. These methodological approaches do not take into consideration what specific aspects of verbal expression it is that human listeners distinguish as being different from an acoustic-perceptual perspective among individuals with autism.

The purpose of this study is to address the following research questions: 1) Are there distinct speech and language differences as perceived by human listeners that distinguish children with autism as sounding “different” compared to their neurotypically developing (NTD) age- and gender-matched peers?; 2) Do experienced and non-experienced judges acoustically rate the narratives of children with autism differently than they rate the narratives of their NTD peers?

2. Methods

This study was conducted in two phases. First, oral narratives were collected on two homogeneous groups of children with and without autism. Second, two cohorts of human listeners were recruited to evaluate the acoustic perceptual characteristics of the oral narratives produced by the group of children with autism as compared to the narratives produced by the group of age- and gender-matched NTD peers.

2.1. Procedures

2.1.1. Oral Narrative Collection

Oral narratives were collected on two groups of children between the ages of eight to 11 years comprised of 24 participants with a formal diagnosis of autism and 23 age- and gender-matched NTD “control” participants. In addition to age and gender, both the ASD and NTD participants were matched based on their performance on the Peabody Picture Vocabulary Test-Fourth Edition (PPVT-IV; Dunn & Dunn, 2007) to control for receptive vocabulary.

In addition to age and gender, the prospective participants with ASD were controlled for homogeneity in that they also met the following inclusionary criteria: 1) a formal diagnosis of autism, Asperger’s syndrome, or pervasive developmental disorder-not otherwise specified (PDD-NOS), given by a psychologist, psychiatrist, neurologist, or pediatrician; and 2) a documented IQ score of above 70 for all children diagnosed with ASD. Both the ASD and NTD groups of participants were recruited from different regions across Connecticut, Massachusetts, and New York.
It is important to note that the participants with ASD were formally diagnosed by qualified specialists using the DSM-IV-TR (2000) and prior to the publication of the DSM-5 (2013). While high-functioning autism spectrum disorder (HFA) is not a formal diagnosis, all participants had an Intelligence Quotient (IQ) above 70, which was typically used to refer to HFA (Klin, 2006). Among the 24 participants in the experimental group, eight (8) participants were diagnosed with Asperger’s syndrome (33.33%), two (2) participants were diagnosed with autism (8.33%) and fourteen (14) participants were diagnosed with pervasive developmental disorder—not otherwise specified (58.33%).

2.1.2. Oral Narrative Stimuli

The Edmonton Narrative Norms Instrument (ENNI; Schneider, Dubé, & Hayward, 2005) was administered to obtain spontaneous oral narrative samples on both the ASD and NTD participants. The ENNI is an instrument that measures storytelling skills among children four to nine years old. The ENNI includes norms for a number of different analyses typically used in expressive language assessments. Narratives from the ENNI allow the investigator to calculate scores on a number of different aspects of storytelling such as, story information, referring expressions, number of different words, Mean Length of Communication Unit, and Subordination Index. There is some evidence of concurrent validity as correlations between the grammar subtest scores for the ENNI and performance on The Clinical Evaluation of Language Fundamentals (CELF-preschool; Wiig, Secord, & Semel, 1992; CELF-III, Semel, Wiig, & Secord, 1995) range from .39 to .70 with the highest correlations found for Expressive Language. There are currently no reliability data on the ENNI.

The ENNI was administered to each participant according to the manual’s instructions in a room devoid of distractions and controlled for ambient noise levels below 50 decibels. Each ASD and NTD participants were seated at a table in a comfortable position during the collection of the oral narratives. Participants were asked to hold the binder with the story stimuli and create a story to accompany the series of pictures using the ENNI stimuli. Each child first performed a training story, and then completed the test story. Only the spontaneous narratives elicited and recorded for each ASD and NTD participant using ENNI’s Story B3- Balloon stimuli were included in this human listener perceptual study.

2.2.3 Human Judges

A total of 40 full-time graduate students comprised of two (2) cohorts of judges were recruited from the University of Massachusetts Amherst campus. One group of 20 Speech Language Pathology (SLP) graduate students (trained listeners) and 20 randomly selected full-time graduate students from different disciplines (untrained listeners) were recruited across the university campus to participate as “human listeners”. All participants selected as judges met the
following inclusion criteria: 1) normal hearing; and 2) English as a native language. All participants also completed a short screening questionnaire to help control for the homogeneity of each group of judges with respect to participant’s level of experience and contact time with adults and children with ASD.

The 20 SLP graduate student judges were engaged in coursework that was required for earning a masters degree and had fairly extensive clinical experience and contact time working with a variety of adults and children with communicative disabilities, including children with autism. The 20 untrained judges earning graduate degrees in a different discipline had little to no experience with adults or children with ASD.

2.2.3. Acoustic Perceptual Stimuli

The spontaneous narratives recorded using the ENNI B3-Balloon story stimuli obtained from the 24 participants with ASD and the 23 age- and gender-matched NTD controls were edited in length to include approximately the same duration of speaking time per participant from each group. Approximately a two-minute sample of oral narratives, which was also controlled for content, were burned onto four Compact Disks (CDs). A total of four unique CDs were created each containing a random sample of oral narratives produced by 10 children with ASD and 10 children from the NTD group. Each of the four CDs included a different randomized order of the 20 oral narratives.

2.2.4. Human Listener Experiment

Each of the 20 SLP and 20 untrained (Non-SLP) graduate student judges were scheduled to participate in a listening experiment on an individual basis. The listening experiment took place in a secure laboratory devoid of distractions. Each judge underwent an orientation and was briefed on how apply a set of criteria using a Likert scale to rate and “judge” each oral narrative based on the perceptual and linguistic characteristics of each child’s oral narratives. Judges were then seated directly in front of a Mac Book Air computer and they were instructed to listen to and score one practice sample of a narrative using the headphones provided, which they set to a comfortable volume for their ears. At this time, the judges were allowed to ask the investigator questions to clear up any confusion. Following the practice session, each judge listened to the 20 narratives on one of the four CDs to which they were assigned. No time constraints were placed on each judge to complete the listening experiment. On average, each judge needed approximately one hour to complete the study following the orientation and practice session. The judges received a small stipend for their participation.

2.2.5. Perceptual Rating Instrument
The perceptual rating instrument that was applied by each judge to rate the oral narratives was developed by the investigators based on published empirical findings regarding speech and language characteristics or differences most frequently observed among children with autism compared to their NTD peers. The perceptual rating instrument consisted of 13 categories comprised of seven (7) linguistic variables (Story sequencing, Topic organization, Story details, Pronominal references, Causal language, Use of elaborations, Emotional language or ToM) and six (6) speech or acoustic variables (Articulation, Fluency, Intonation, Rate, Pitch, and Loudness). Each category included a three-point Likert rating scale ranging from a score low score of “0” to a high score of “3”. To assist the judges in selecting the most appropriate rating that best represented what they heard or perceived for each of the 13 variables, a description of the speech and linguistic behaviors was included next to each Likert rating score.

2.2.6. Statistical Analysis

Repeated measure ANOVAs were calculated for the main effect diagnosis between the groups (ASD vs. NTD) and the main effect judge status within groups (SLP vs. untrained group). Alpha level was set at 0.05 for statistical significance. Inter-rater reliability on 20% of the data (four SLP judges and 4 untrained judges) revealed substantial agreement at 0.73 (0.61-0.80, Landis & Koch, 1977).

3.0 Results

The results of this study revealed that there are distinct speech and language differences as perceived by human listeners that distinguish children with ASD as sounding “different” compared to their NTD age- and gender-matched peers. For the main effect diagnosis (ASD vs. NTD), statistically significant between group differences were found for four (4) out of 13 variables: Story sequencing (F(1,38)=7.95, p=0.008); Articulation (F(1,38)=12.47, p=0.001); Fluency (F(1,38)=3.39, p=0.005); and Rate (F(1,38)=6.36, p=0.016). A trend, but nonsignificant differences at the 0.05 alpha level were found for the variables: Topic organization (F(1,38)=3.39, p=0.073) and Pitch (F(1,38)=3.82, p=0.058). For all six variables, narratives produced by children with autism yielded lower mean scores and greater variability. Nonsignificant differences for the main effect diagnosis were found for the variables: Story details (F(1,38)=0.92, p=0.343); Pronominal referencing (F(1,38)=0.17, p=0.682); Causal language (F(1,38)=1.16, p=0.289); Use of elaborations (F(1,38)=2.79, p=0.104); Emotional language (ToM) (F(1,38)=0.92, p=0.343); Intonation (F(1,38)=2.33, p=0.136); and Loudness (F(1,38)=0.08, p=0.782).

For the main effect judge status (SLP vs. untrained listeners), statistically significant within group interactions were found for three (3) out of the 13 variables: Emotional language (ToM) (F(1,38)=5.93, p=0.020); Articulation
(F(1,38)=11.23, p=0.002); and Rate (F(1,38)=4.69, p=0.037). A trend, but nonsignificant interaction at the 0.05 alpha level was found for the variable: Causal language (F(1,38)=2.96, p=0.094). Nonsignificant differences for the main effect judge status were found for the variables: Story sequencing (F(1,38)=0.13, p=0.718); Topic Organization (F(1,38)=0.41, p=0.525; Story details (F(1,38)=0.17, p=0.684); Pronominal referencing (F(1,38)=0.89, p=0.355); Use of elaborations (F(1,38)=0.23, p=0.633); Fluency (F(1,38)=0.25, p=0.618); Intonation (F(1,38)=0.04, p=0.837); Pitch (F(1,38)=0.32, p=0.318); and Loudness (F(1,38)=0.09, p=0.762).

4.0 Discussion

The purpose of this investigation was determine whether there were distinct speech and language differences as perceived by trained (SLP) and untrained (non-SLP) human listeners that distinguish children with autism as sounding “different” compared to their NTD age- and gender-matched peers. The results of this study provide empirical evidence that children between the ages eight and 11 years diagnosed with HFA (under the DSM-IV-TR, 2000 criteria) produced oral narratives that were perceived by human listeners as “different” with respect to the Story sequencing and speech/acoustic variables, Articulation, Fluency and Rate. Although nonstatistically significant at the 0.05 level, a trend was observed for the variables, Topic organization and Pitch. The oral narratives produced by the children with autism for all six of these variables demonstrated lower group mean scores and greater variability compared to their NTD counterparts.

Experienced (SLP) and untrained (Non-SLP) judges rated the oral narratives of children with ASD differently compared to the manner in which they rated the oral narratives for the age- and gender-matched NTD peers with respect to Emotional language (ToM), Articulation and Rate. Similarly, a trend, but nonsignificant interaction was found for the linguistic variable, Causal language. Experienced (SLP) judges rated the ASD narratives with higher (or better) scores than the untrained (Non-SLP) judges for all four of these categories. These differences may be due to a number of reasons. The experienced SLP judges have more practice transcribing speech and language samples, including assessing and rating the severity of a communication problem. The SLP judges may be applying significantly stricter criteria for judgment across all linguistic and speech/acoustic measures in that the spontaneous speech samples they rated in this study were not remarkably severe considering the caseloads they encounter ranging from mild to very severe communicative disabilities. The untrained judges may have rated the speech and language productions of participants with lower (or poorer) scores, since they do not have the experience in deciphering and attending to the subtle linguistic and phonotactic features of communication, including varying degrees of severity. Nonetheless, specific parameters are discernable to even untrained ears when
judging connected speech via oral narrative production of children with ASD compared to NTD children.

5.0 Conclusion

This research also supports that experienced (SLP) and non-experienced (untrained) judges acoustically rated the narratives of children with ASD differently than they rated the narratives of NTD children. It is important to note that the acoustic/speech aspects of the connected speech, rather than use of linguistic structures, were found to be the most discriminating features in this study.

It is not clear what underlying mechanisms contribute most to the “oddness” or speech atypicalities observed among the children with autism compared to their age- and gender-matched peers. However, based on results of this study and published empirical findings, the following hypothesis are proposed: 1) pragmatic deficits affecting attunement and emulation of social pragmatic cues (Shriberg et al., 2011; Diehl & Paul, 2013); motor deficits affecting motor execution, motor programming and motor planning (Shriberg et al., 2001; Peppe et al., 2007; Velleman, Andrianopoulos, Boucher, Perkins, Averback, Currier, Marsello, Lippe, & Van Emmerik, 2010; Diehl & Paul, 2012); deficits in mechanisms that control pitch, issues affecting auditory processing, phonological representation and motor functioning (Bonneh et al., 2010; Diehl & Paul, 2013); poor language ability and atypical or delayed prosodic development (Shriberg et al., 2001; Diehl et al., 2009; Bonneh et al., 2010; Peppe et al., 2011; Diehl & Paul, 2013).

The limitations of this study include the small sample size of participants (ASD vs. NTD) and judges (SLP vs. Non-SLP). In addition, a more robust instrument and standardized assessment protocol is needed to better identify the salient and critical markers of atypical speech, such as articulation, prosody and other linguistic features, is needed to better assess the communication abilities of speakers with autism. Researchers should be applying the same criteria when studying the speech and language differences among speakers with ASD so that they are comparing apples-to-apples when determining the underlying mechanisms regarding the speech atypicalities of children with ASD.

6.0 References


McCabe, Allyssa, Hillier, Ashleigh, Shapiro, Claudia. (2013). Brief report Structure of personal


