Prosody in Young Verbal Children With Autism Spectrum Disorder

Ashlynn McAlpine
Laura W. Plexico
Allison M. Plumb
Auburn University, Auburn, AL
Julie Cleary
Le Bonheur Children’s Hospital, Memphis, TN

ABSTRACT: Purpose: The purpose of this study was to examine the expressive prosody of young verbal children with autism spectrum disorder (ASD).
Method: Vocalizations were examined in 14 children between 24 and 68 months of age: 7 diagnosed with ASD and 7 with typical development, matched for expressive language abilities. To obtain precise measures of prosody, spontaneous speech samples were obtained and then coded using The Prosody Voice-Screening Profile (PVSP; Shriberg, Kwiatkowski, & Rasmussen, 1990).
Results: The 2 groups did not differ in their production of rate, loudness, or pitch. The group with ASD did, however, produce atypical stress patterns significantly more often, such as misplacement of stress in multisyllabic words or reduced stress.
Conclusion: Results of the current investigation highlight the importance of considering prosody, specifically misplaced stress, when assessing and treating young children with ASD.

KEY WORDS: autism spectrum disorder, prosody, early identification

Kanner’s (1943) original description of autism included several features of disordered social communication, such as echolalia, pragmatic difficulties, and unusual expressive prosody, and these features continue to be considered characteristic of autism spectrum disorder (ASD; American Psychological Association, 2013). Individuals with ASD are also known to exhibit difficulty with emotions and affect, which could be reflected in their difficulties understanding and using prosody in social communication (McCann, Peppe, Gibbon, O’Hare, & Rutherford, 2007).

Prosody is a term that describes the suprasegmental aspects of language that we use to communicate, modify, or highlight the meaning of spoken messages (Couper-Kuhlen, 1986). Phrasal stress, boundary cues, and meter are suprasegmental aspects of language that are often used to modulate the prosodic intent of a message. Gerken and McGregor (1998) provided a tutorial on prosody in which these terms were explained through the lens of normal and disordered child language. When using phrasal stress, for example, normal speakers will increase the saliency of a word by increasing its loudness, duration, and pitch. Boundary cues serve a syntactic purpose in that they are often used to segment units of language (e.g., phrases, clauses, or sentences). These syntactic boundaries are marked by the inclusion of pauses, changes in duration, and pitch adjustments. Meter refers to the arrangement of syllables within an utterance in a rhythmic manner that reflects the stress patterns of a language. There are two types of meter: lexical meter...
and phrasal meter. **Lexical** meter is used to denote the pattern of strong (stressed) and weak (unstressed) syllables within a word. For example, in English, it is common to have an alternation between strong and weak syllables, such as in the word “AL a BAM a.” **Phrasal** meter refers to the pattern of strong and weak syllables within a phrase. Syllables are often arranged in a manner to avoid having more than two unstressed syllables in a row (e.g., “He WANTed a DOG.”)

These suprasegmental aspects of prosody serve to enhance the communicative message for a variety of purposes. These purposes can be divided into three subdomains of prosody: grammatical, pragmatic, and affective prosody (Paul, Augustyn, Klin, & Volkmar, 2005). **Grammatical** prosody is used by a speaker to communicate syntactic information (Warren, 1996). For example, stress can be used within words to signal the part of speech of a word, such as “INSult” (a noun) and inSULT (a verb). Grammatical prosody can also signal whether an utterance is a question or a statement through the use of pitch contours, in which a rising pitch indicates a question and a falling pitch indicates a statement.

**Pragmatic** prosody, sometimes referred to as emphatic or contrastive stress, is used by a speaker to relay social information that would not be considered part of the syntactic structure of the message. In this manner, pragmatic prosody would be used to draw attention to certain elements within a sentence. It could be used to separate the presentation of new information from old information in a given conversation (Paul et al., 2005) or to highlight unfamiliar or unexpected information, thereby making it more salient for the listener.

**Affective** prosody differs from grammatical and pragmatic prosody in that it conveys information beyond grammaticality and emphasis to include more global social purposes, such as the conversational register (i.e., the way a speaker may change his or her speaking style based on the audience), as well as the suprasegmental changes that can occur due to a change in the speaker’s general emotional state (Bolinger, 1989; Hargrove, 1997). The same sentence can sound very different based on suprasegmental features reflecting one’s emotional state. For example, “What are you doing here?” could sound differently depending on whether the speaker was genuinely happy and excited to see his or her visitor or if it was a visitor he or she had hoped not to see.

**Prosodic Characteristics Observed in Children With ASD**

By 30 months of age, typically developing (TD) children display mastery of prosodic cues, such as phrase stress, intonation, and boundary cues (Allen & Hawkins, 1980; Klein, 1981; Snow, 1994). In contrast, individuals with ASD display deficits throughout their life span in many aspects of prosody, including rhythm, stress, fluency, phrasing, intonation, and affect. Although abnormal prosody has continuously been identified as one of the core features of children with ASD, there is a limited amount of research within this area that explores the relationship between the prosodic characteristics of children with ASD and how these characteristics influence their communication. To date, the majority of studies have concluded that individuals with ASD encounter more difficulty with receptive and expressive aspects of grammatical, pragmatic, and affective prosody than their TD peers (McCann et al., 2007; Paul et al., 2005).

Studies of prosody in children with ASD can include acoustic and/or auditory–perceptual analyses. Acoustic measures include amplitude, frequency, and duration; auditory–perceptual measures include rater judgments of intonation patterns, the placement of accent and stress, and boundary and phrasing placement in speech. Although a challenging and time-consuming endeavor, a growing number of acoustic studies are providing much needed information on the differences in the speech of children with ASD (Diehl & Paul, 2012, 2013; Diehl, Watson, Bennetto, McDonough, & Gunlogson, 2009; Fosnot & Jun, 1999; Green & Tobin, 2009; Grossman, Bemis, Skwerer, & Tager-Flusberg, 2010; Hubbard & Trauner, 2007; Nadig & Shaw, 2012; Van Santen, Prud’hommeaux, Black, & Mitchell, 2010). The results of these studies have indicated prosodic differences in the duration, intonation, and pitch of children with ASD. More specifically, it has been found that the sentence and word durations produced by children with ASD are longer than those produced by their TD peers (Diehl & Paul, 2012, 2013; Fosnot & Jun, 1999; Grossman et al., 2010; Van Santen et al., 2010). The intonation patterns of children with ASD have been found to be influenced by increased and inappropriate use of pitch accents as well as difficulty producing high-frequency boundary tones (Fosnot & Jun, 1999). In addition, the pitch of individuals with ASD has been documented to have an increased fundamental frequency range and a higher fundamental frequency average than the pitch of TD individuals (Diehl et al., 2009; Green & Tobin, 2009; Hubbard & Trauner, 2007; Nadig & Shaw, 2012).

In one of the few studies of prosody to use a TD control group as well as a control group of children with learning disabilities (LD), Diehl and Paul (2013) used acoustic measures of prosody to compare children/adolescents with ASD to chronological
age–matched controls. All of the participants were between 8 and 17 years of age. Speech was elicited for analysis using the Profiling Elements of Prosodic Systems—Children test (PEPS–C; Peppe & McCann, 2003) and was analyzed for average fundamental frequency across the entire utterance, fundamental frequency range, standard deviation of the fundamental frequency, utterance duration, and utterance intensity. The results indicated that the group with ASD had longer single-word utterance durations on multiple subtests of the PEPS–C than the comparison groups. On one subtest, the ASD and LD groups were also found to have higher pitch ranges and showed greater pitch variance than the TD group. The results of this investigation indicated that many of these differences were present even when the differences were judged to be perceptually appropriate.

In a follow-up study by Diehl and Paul (2012) that used the same participants, imitative utterances from the Form Output subtests of the PEPS–C were analyzed for the acoustic measures that were used in Diehl and Paul (2013). Results provided further support for the previous findings, as the ASD group produced longer durations than the comparison groups when attempting to imitate an intonation pattern.

Several studies have also used perceptual methods to investigate aspects of prosody, including stress, phrasing, and intonation patterns. Results of these perceptual investigations indicated that individuals with high-functioning autism (HFA) misassign or inappropriately use contrastive stress (Baltaxe, 1984; Baltaxe & Guthrie, 1987; McCaleb & Prizant, 1985; McCann et al., 2007; Paul et al., 2005; Peppe, Cleland, Gibbon, O’Hare, & Castilla, 2011; Shriberg et al., 2001). Phrasing differences observed included fewer nongrammatical pauses, increased disfluency, and failure to appropriately use tone boundaries (Fine, Bartolucci, Ginsberg, & Szatmari, 1991; McCann et al., 2007; Paul et al., 2005; Peppe et al., 2011; Shriberg et al., 2001; Thurber & Tager-Flusberg, 1993). Intonation patterns are influenced by difficulty producing stress for grammatical and affective communication (Fine et al., 1991; McCann et al., 2007; Shriberg et al., 2001).

Using the Prosody-Voice Screening Profile (PVSP; Shriberg, Kwiatkowski, & Rasmussen, 1990), Shriberg et al. (2001) documented the speech profiles of 15 males between the ages of 10 and 49 years with HFA, 15 males of similar age with Asperger’s disorder (AD), and 53 similar-age males with normal speech development. Prior to analyzing the PVSP prosody-voice (PV) codes, PVSP exclusion codes (to be discussed further in the Method section) were assigned to those utterances that were to be excluded from the sample. Based on this analysis, the HFA group had significantly more excluded utterances (42.3%) than the control group with normal speech development (21.3%). The most frequent exclusion codes used were Code 5: Interruption/Overtalk and Code 7: Only One Word.

Results of the subsequent PV code analysis revealed that individuals with HFA and those with AD used excessive or misplaced stress in their speech as well as inappropriate phrasing. Out of the 30 individuals with ASD, Shriberg et al. (2001) found that 53% of the individuals with HFA and 26.7% of the individuals with AD used stress improperly, and 66.7% with AD misused phrasing. In comparison, only 5.7% of the TD control group misassigned stress, and 26.4% used incorrect phrasing in their speech. Shriberg et al. noted that the majority of inappropriate stress produced by the HFA and AD speakers involved the pragmatic use of stress, which was consistent with findings by Baltaxe (1984), who found that children with ASD were twice as likely to misassign contrastive stress compared to TD children and children with aphasia. In addition, the children with ASD tended to stress more than one stressable syllable—an error that was not found in either of the other groups.

Paul et al. (2005) expanded on the Shriberg et al. (2001) study in an investigation that examined the perception and production of prosody in 27 individuals with ASD with an average age of 16;8 (years;months) and 13 typical peers with an average age of 16;7. Individuals participated in 12 experimental tasks that were designed to examine the perception and production of three aspects of prosody: intonation, stress, and phrasing. Paul et al. found significant differences between the two groups in the grammatical production of stress as well as the pragmatic perception and production of stress. Paul et al. concluded that individuals with ASD have difficulty understanding how to properly use stress in their speech, which hinders their ability to produce affective and appropriate stress patterns. However, no significant differences were found between the two groups regarding intonation or phrasing, which may indicate that speakers with ASD encounter the greatest difficulty when using stress that serves pragmatic or affective functions in communication.

The speech of children with ASD has consistently been described in the literature as containing atypical vocalizations and unusual prosody (Baltaxe, 1984; Fosnot & Jun, 1999; Grossman et al., 2010; Hubbard & Trauner, 2007; McCann & Peppe, 2003; McCann et al., 2007; Shinkopf, Mundy, Oller, & Steffens, 2000; Shriberg et al., 2001). Although a growing body of evidence exists investigating the prosody of children with ASD, studies have not established
consistent patterns that have enabled clinicians to develop clear theoretical or clinical implications. The inconsistent findings can be attributed to methodological differences found across studies (e.g., sample size, assessment instruments) as well as variability in participant characteristics (e.g., age, age range, diagnosis; McCann & Peppe, 2003). A majority of studies have focused on school-age individuals with HFA and AD, resulting in a need for further information on the prosodic patterns of young verbal children with ASD.

Although a majority of studies of prosody in children with ASD have included participants up to 14 years of age (Baltaxe 1984; Baltaxe & Guthrie, 1987; McCaleb & Prizant 1985; McCann et al., 2007), there is a growing body of evidence that investigates younger children specifically. Bonneh, Levanon, Dean-Pardo, Lossos, and Adini (2011) investigated the pitch variability of 83 children between 4 and 6.5 years of age through the acoustic analysis of recordings during the naming of pictures. The results indicated that the ASD group demonstrated significantly more variable pitch range and variability across time than the TD children. This finding is in contrast to the common belief that the speech of individuals with ASD is monotonic.

Sharda et al. (2010) compared the intonation of 15 children with ASD between 4 and 10 years of age with 10 TD children matched on chronological age. The findings showed that the ASD group demonstrated exaggerated pitch, pitch range, and pitch contours. Using acoustic and perceptual measures, Van Santen et al. (2010) analyzed the speech of children with ASD between 4 and 8 years of age during imitation and picture description tasks. The children with ASD were found to perform significantly better than the TD children matched on nonverbal IQ during the picture description task. There were only weak differences on the imitation tasks. Interestingly, Van Santen et al. found that distinguishing features identified by auditory perceptual judgment were not the same distinguishing features found with instrumental measures. There was a bias toward frequency differences when using perceptual measures, whereas instrumental measures indicated that durational differences more clearly distinguished the two groups. With that said, however, the findings indicated that there was not a salient feature that clearly differentiated the ASD children from the TD children. Rather, it was a balance or interplay among the prosodic features that distinguished the children.

Purpose

The acoustic and perceptual studies discussed here provide pertinent information regarding the atypical prosody of individuals with ASD, yet they do not provide conclusive evidence due to methodological differences. Moreover, only a small number of studies have specifically investigated the prosodic characteristics of verbal preschool-age children with ASD. Due to the lack of conclusive evidence concerning the prosody of children with ASD, more specifically young children, the present study was undertaken to provide a detailed perceptual analysis of the prosody components of young verbal children between the ages of 24 and 68 months who have been diagnosed with ASD. Specifically, this study explored prosodic characteristics such as loudness, pitch, and quality of TD individuals and individuals with ASD. We used the PVSP to evaluate the degree to which atypical prosodic characteristics appeared in the speech of TD young children and young children with ASD.

We asked the following three questions:

- Do children with ASD differ from TD children with regard to the number of utterances related to content and/or context that must be excluded from the speech sample?
- Do the number and type of prosodic characteristics present in the speech of children with ASD differ from those present in the speech of TD children?
- When using the PVSP cutoff criterion, are young children with ASD more likely than TD children to have greater than 10% of their utterances coded as inappropriate for each of the PV variables?

METHOD

Participants

The participants in this study included two groups of children between the ages of 24 and 68 months: verbal children with ASD and TD children. These children were selected from a research study that was conducted at The University of Memphis and Auburn University investigating early markers of ASD in young children. To be included in the study, children had to meet the following criteria: (a) no known hearing or visual impairments or comorbid diagnoses; (b) monolingual, English speakers; (c) no prematurity; (d) no low birth weight; and (e) self-identified as either African American or Caucasian race when provided with the race options defined by the U.S. Census Bureau (2010; i.e., White, Black or African American, American Indian and Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Other Race). To be selected for the current study, the children also
were required to be verbal communicators (i.e., producing at least two-word combinations).

Children in the TD group tested above the 10th percentile (less than \(1\frac{1}{4}\) SD below the mean) on the subtests of fine motor, visual reception, receptive language, and expressive language on the Mullen Scales of Early Learning (MSEL; Mullen, 1995) and had no concerns about ASD reported by either the clinician or parent as measured by the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003). For the ASD group, a diagnosis of ASD within the previous 6 months was required. The diagnosis was confirmed by the investigators through administration of the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999), which is a standardized diagnostic instrument for direct observation of ASD symptomatology. To confirm or rule out a diagnosis of ASD, the investigators also used the Autism Diagnostic Interview—Revised (ADI–R; Rutter, LeCouteur, & Lord, 2003), which is a standardized interview measure that is designed to assess the range of behaviors relevant to the differential diagnosis of ASD. In addition, the diagnostic criterion established by the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (American Psychological Association, 2000) was adhered to at the time of assessment when confirming or disconfirming a best diagnosis of ASD.

Fourteen participants met the study criteria and were selected for inclusion in the study. The participants for the ASD group consisted of six males and one female between the ages of 38 and 62 months (\(M = 47.57; SD = 8.73\)). Participants for the TD group consisted of five males and two females between the ages of 30 and 64 months (\(M = 44.14; SD = 13.26\)). Results from the Wilcoxon-Mann-Whitney rank-order difference test showed no significant difference between the two groups with regard to chronological age (\(z = –.643, p = .520\)). All of the participants were Caucasian and were recruited from communities in the southern region of the United States (see Table 1 for demographic information).

### Procedure

The participants were evaluated individually at The University of Memphis or Auburn University. All evaluations included the MSEL, the Communication and Symbolic Behavior Scales Developmental Profile (CSBS DP; Wetherby & Prizant, 2002), and an informal play sample. The MSEL is used to evaluate the cognitive skills of children through motor, visual reception, and expressive and receptive language subscales. The MSEL can be administered to infants and children up to 68 months of age. The ASD group tested 1.5 SDs below the mean on the fine motor and receptive language scales. However, the ASD group was within 1 SD of the mean on the visual reception and expressive language subscales. The TD group was within 1 SD of the mean on all four scales (see Table 2).

The CSBS DP was used to determine a child’s communicative competence through the use of eye gaze, gestures, sounds, words, understanding, and play. The test was designed for children with a functional communication age between 6 and 24 months. However, it may be used for children up to 6 years of age whose functional communication skills are below 24 months (Wetherby & Prizant, 2002). For the current study, administration of the CSBS DP allowed for a structured context for sampling communication behaviors, regardless of the child’s language level.

The informal play sample was used to evaluate a child’s communicative behaviors during interactions

<table>
<thead>
<tr>
<th>Group</th>
<th>ASD</th>
<th>TD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Child’s age at initial evaluation in months ((M, SD))</td>
<td>47, 8.73</td>
<td>44.14, 13.26</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female ((n, %))</td>
<td>1, 14</td>
<td>2, 28</td>
</tr>
<tr>
<td>Male ((n, %))</td>
<td>6, 86</td>
<td>5, 72</td>
</tr>
<tr>
<td>Parents’ education in years completed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother ((M, SD))</td>
<td>16, 2.23</td>
<td>17, 2.69</td>
</tr>
<tr>
<td>Father ((M, SD))</td>
<td>16, 2.15</td>
<td>17, 2.27</td>
</tr>
<tr>
<td>Parents’ age at child’s evaluation in years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother ((M, SD))</td>
<td>33.00, 3.68</td>
<td>34.44, 3.84</td>
</tr>
<tr>
<td>Father ((M, SD))</td>
<td>41.41, 7.73</td>
<td>35.79, 4.52</td>
</tr>
</tbody>
</table>

**Note.** ASD = autism spectrum disorder, TD = typically developing.
with his or her parent or caregiver. During the assessment, novel toys were arranged within and outside of the child’s reach in a clinical evaluation room. Caregivers and their children were asked to play as they typically would with toys and during social games. As discussed earlier, the children who demonstrated language delays on the MSEL received two additional assessments to confirm or rule out a best estimate diagnosis of ASD: ADOS and ADI–R.

Participants were matched based on mean length of utterance (MLU) and MSEL Expressive Language Developmental Quotient (see Table 2), which was derived from the Expressive Language scale of the MSEL. The MLU was obtained by entering the transcribed speech samples into the Systematic Analysis of Language Transcripts program (SALT; Miller & Iglesias, 2008). Results from the Wilcoxon-Mann-Whitney rank-order difference test showed no significant difference between the two groups with regard to the MSEL Expressive Language Developmental Quotient ($z = –1.154$, $p = .249$) or MLU ($z = –.703$, $p = .482$). Charman (2004) suggested that combining measures of language from more than one source provides a more global assessment of language competence than a single measure does.

Recording of speech samples. All of the evaluations were audio- and video-recorded in a room that was filled with a variety of toys. The children wore vests that follow a design developed by Buder and Stoel-Gammon (2002). The vests were equipped with a Countryman Isomax microphone and a Samson wireless transmitter that sent a signal to a Samson receiver. TF-32 software (Milenkovic, 2002) operating a Data Translation acquisition card was used to digitize the child’s signals at 48 kHz after low-pass filtering at 20 kHz using a Data Translation antialiasing board.

Data analysis using the PVSP. The speech samples obtained from the audio and video recordings were transcribed verbatim using WAVpedal (n.d.) software. Following transcription, each sample was segmented into utterances according to the segmentation rules established by Shriberg et al. (1990). Segmentation rules include how an utterance is defined, how to segment utterances that include conjunctions, and how to separate a run-on string of words or ideas. Each speech sample, consisting of speech from the CSBS DP and informal play sample, was analyzed.

The next step in the PVSP process was to exclude utterances from PV coding. Thirty-two exclusion codes have been developed to improve the validity of PVSP scores, with each code reflecting a type of utterance that may bias or contraindicate PV coding (Shriberg et al., 1990). The exclusion codes are divided into four categories: (a) content/context (utterances excluded due to linguistic content or sociolinguistic context), (b) environment (utterances excluded due to problems with the recording environment), (c) register (utterances excluded because they contain specific prosodic characteristics, such as a whisper or sound effect), and (d) states (utterances excluded due to biological states, such as laughing, coughing, and sneezing). A more detailed explanation of these exclusion codes is provided in the PVSP manual (Shriberg et al., 1990).

Following the exclusion of specific utterances, the first 100 utterances that met the criteria for PVSP coding were coded according to the six suprasegmental variables of prosody and voice as defined by the PVSP: (a) phrasing, (b) rate, (c) stress, (d) loudness, (e) pitch, and (f) quality. Each of the PV codes is accompanied by a definition as well as scoring procedures (Paul et al., 2005; Shriberg et al., 1990).

In order to obtain PVSP scores, it is suggested that a minimum of 12 codable utterances be obtained, with half of the utterances consisting of at least four words. However, in an effort to obtain a representative sample and due to the young age and developmental delays of the participants, we altered this rule.

### Table 2. Descriptive statistics of the standardized language measures.

<table>
<thead>
<tr>
<th>Measure</th>
<th>ASD</th>
<th>SD</th>
<th>ASD</th>
<th>SD</th>
<th>ASD</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean length of utterance</td>
<td>3.06</td>
<td>3.45</td>
<td>2.11–4.34</td>
<td>2.13–4.81</td>
<td>0.87</td>
<td>0.98</td>
</tr>
<tr>
<td>Mullen Scales of Early Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual reception(^a)</td>
<td>40.71</td>
<td>52.29</td>
<td>20–58</td>
<td>40–63</td>
<td>12.93</td>
<td>6.70</td>
</tr>
<tr>
<td>Fine motor(^a)</td>
<td>32.14</td>
<td>47.14</td>
<td>20–45</td>
<td>35–58</td>
<td>9.75</td>
<td>8.47</td>
</tr>
<tr>
<td>Receptive language(^a)</td>
<td>32.71</td>
<td>57.00</td>
<td>24–49</td>
<td>48–80</td>
<td>9.66</td>
<td>11.17</td>
</tr>
<tr>
<td>Expressive language(^a)</td>
<td>42.00</td>
<td>48.00</td>
<td>33–48</td>
<td>44–54</td>
<td>6.43</td>
<td>3.70</td>
</tr>
<tr>
<td>Expressive language quotient(^b)</td>
<td>87.64</td>
<td>92.96</td>
<td>73.6–97.7</td>
<td>79.7–102.0</td>
<td>10.10</td>
<td>7.43</td>
</tr>
</tbody>
</table>

\(^a\)The Mullen Scales of Early Learning (Mullen, 1998) standard scores are based on $M = 50$ and $SD = 10$. \(^b\)The Mullen Scales of Early Learning expressive language developmental quotient is based on the expressive language age equivalent divided by the chronological age multiplied by 100.
to include utterances containing three or more words. On average, children between the ages of 24 and 36 months typically produce utterances consisting of two to three words (Garvey & BenDebba, 1974).

Once the speech sample was coded, the supra-segmental variables were tallied in order to obtain a frequency of occurrence (total number of occurrences/100 × 100), which was represented as a percentage. A score of less than 100% on any one variable indicates that utterances were judged as containing inappropriate prosody or voice characteristics. Shriberg et al. (1990) set criterion for passing at 90% in order to differentiate individuals who would not require further evaluation from those who would be recommended to receive clinical services.

A total of two to three investigators listened to and transcribed each speech sample in order to judge the intelligibility of given utterances before they were coded as unintelligible. Interjudge reliability was evaluated for 100% of the samples for PV coding. For several PV codes, including phrasing, loudness, laryngeal features, and resonance, two investigators analyzed 100% of the speech sample in order to obtain reliability. Three investigators listened to 40% of each speech sample in order to determine rate, stress, and pitch. Each of these investigators overlapped across 10 utterances per child, which were used to obtain reliability. The mean interjudge reliability calculated was 94%, with values ranging from 80% to 100%. These results are similar to those obtained in reliability studies for PV coding described by Shriberg et al. (1990), in which estimates from developmental studies for interjudge agreement ranged from 78% to 96% and estimates from clinical studies ranged from 74% to 99%.

RESULTS

We used nonparametric analyses to examine the prosody measures obtained for the current investigation. Exclusion codes were initially analyzed to determine if the ASD group had significantly more excluded utterances than the TD group. Following exclusion code analysis, PSVP measures were analyzed for group differences.

Exclusion Code Analysis

Percentage of excluded utterances. The first question, concerning the number of utterances related to content and/or context that must be excluded from the speech sample was answered by comparing the percentages of utterances that had to be excluded from the analyses because they met the criteria for one of the exclusion codes. A pairwise, nonparametric group comparison was made using the Wilcoxon-Mann-Whitney rank-order difference test. The ASD group had a mean rank of 85.40, and the TD group had a mean rank of 83.65. There was no significant difference between the two groups in the percentage of utterances that had to be excluded from the analyses, z = −0.251, p = .802.

Bases for the exclusion codes. Table 3 illustrates the means and standard deviations associated with each of the 12 exclusion codes for content/context. As shown, the most frequent codes used in both groups were interruption (Code 5), one word (Code 7), and unintelligible utterances (Code 12). Group differences were not analyzed at the post hoc level because no significant difference was found between groups at the omnibus level; however, the ASD group did have a higher percentage of utterances excluded due to instances of imitation (Code 4). The ASD group had an average of 3.81 utterances excluded due to imitation, whereas the TD group only had 0.96 excluded. Because a number of research studies have described the speech of children with ASD as echo-lalic (Prizant & Duchan, 1981; Wetherby, 1986), this trend is consistent with the literature.

Analysis of PVSP Group Differences

The second question regarding the number and type of prosodic characteristics present in the speech of

<table>
<thead>
<tr>
<th>Table 3. Exclusion codes by group.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclusion code</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>1 Automatic sequential</td>
</tr>
<tr>
<td>2 Back Channel</td>
</tr>
<tr>
<td>3 I don’t know</td>
</tr>
<tr>
<td>4 Imitation</td>
</tr>
<tr>
<td>5 Interruption</td>
</tr>
<tr>
<td>6 Not 3+ words</td>
</tr>
<tr>
<td>7 Only one word</td>
</tr>
<tr>
<td>8 Only person’s name</td>
</tr>
<tr>
<td>9 Reading</td>
</tr>
<tr>
<td>10 Singing</td>
</tr>
<tr>
<td>11 Second repetition</td>
</tr>
<tr>
<td>12 Unintelligible</td>
</tr>
</tbody>
</table>

Number of utterances analyzed | 193.140 | 41.260 | 179.430 | 37.020 |

Percentage of utterances excluded | 48.220 | 44.270 |

Note. The mean reflects the mean percentage for total utterances.
children with ASD differing from those present in the speech of TD children was answered by determining whether or not specific PV codes appeared more often in the speech of one group versus the other. To determine if the groups differed significantly from each other, we analyzed the data using a series of Wilcoxon-Mann-Whitney rank-order difference tests. Table 4 highlights the means, ranges, and standard deviations for the seven prosody codes assessed. Inappropriate stress was observed more often in the speech of children in the ASD group than the TD group ($z = -3.108$, $p = .002$). The ASD group produced an average of 18.43 instances of inappropriate stress per 100 utterances compared to the TD group, which produced an average of 4.57 instances of inappropriate stress. No significant differences were observed between the groups on the remaining PV codes.

**Group Differences Observed for PVSP Cutoff Criterion**

The third question regarding young children with ASD being more likely than TD children to have greater than 10% of their utterances coded as inappropriate for each of the PV codes was answered by looking descriptively at whether or not the percentages obtained for the ASD group exceeded the 10% criterion. No children in either the ASD or TD group exceeded the 10% cutoff criterion for the PV codes of rate, pitch, loudness, or resonance. We conducted a series of contingency table analyses to evaluate the pairwise differences among the means for the remaining PV codes. Results indicated that significantly more children with ASD ($n = 7$) failed to surpass the 90% cutoff criterion for the stress PV code (Fischer’s exact test = .002) than TD children ($n = 1$). There were no significant differences between the two groups with regard to the number who failed to surpass the 90% cutoff criterion for phrasing (ASD, $n = 2$; TD, $n = 1$) or laryngeal features (ASD, $n = 1$; TD, $n = 0$). The results indicated that children with ASD are more likely than TD children to fail to surpass the 90% cutoff criterion of only the stress portion of the PVSP.

**DISCUSSION**

A number of research studies have indicated that the speech of children with ASD contains atypical vocalizations and unusual prosody (Baltaxe, 1984; Fosnot & Jun, 1999; Grossman et al., 2010; McCann & Peppe, 2003; McCann et al., 2007; Sheinkopf et al., 2000; Shriberg et al., 2001). These studies have shown that children with ASD display atypical rate, rhythm, intonation, and stress patterns when compared to TD children (Baltaxe, 1984; Fosnot & Jun, 1999; Shriberg et al., 2001). However, these studies have differed with regard to the number of participants, age ranges included, and methodology.

The primary purpose of the current study was to evaluate the prosodic characteristics of young verbal children with ASD between the ages of 24 and 68 months. Specifically, we explored characteristics such as the loudness, pitch, and vocal quality of individuals with ASD. We used the PVSP to evaluate the degree to which atypical prosodic characteristics appear in the speech of young children with ASD. This particular assessment was selected because it is a well-validated prosody assessment that can be used with a variety of age ranges and a variety of speech and language disorders. In addition, it contains a detailed coding scheme, which allows for precise coding of the expressive prosody variables of interest. Following the assessment, the prosodic characteristics of the ASD group were compared to the prosodic characteristics of the TD group to determine whether or not differences exist.

**PVSP Exclusion Codes**

In this study, there was no difference between the two groups of children with regard to the percentage of utterances that were excluded from the analyses.

---

**Table 4.** Prosody Voice-Screening Profile (Shriberg, Kwiatkowski, & Rasmussen, 1990) codes by group.

<table>
<thead>
<tr>
<th>PV code</th>
<th>ASD</th>
<th>TD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Min</td>
</tr>
<tr>
<td>Phrasing</td>
<td>9.71</td>
<td>5</td>
</tr>
<tr>
<td>Rate</td>
<td>34.00</td>
<td>20</td>
</tr>
<tr>
<td>Stress</td>
<td>18.43</td>
<td>12</td>
</tr>
<tr>
<td>Pitch</td>
<td>1.57</td>
<td>0</td>
</tr>
<tr>
<td>Loudness</td>
<td>1.43</td>
<td>0</td>
</tr>
<tr>
<td>Laryngeal features</td>
<td>3.86</td>
<td>0</td>
</tr>
<tr>
<td>Resonance</td>
<td>0.00</td>
<td>0</td>
</tr>
</tbody>
</table>
PV Findings

Results from the current study both support and contradict findings from previous studies. Specifically, data from this study indicated that children with ASD produced atypical stress patterns, such as multisyllabic word stress or reduced stress, in relation to the TD group. The majority of inappropriate stress for the ASD group involved the use of lexical stress within multisyllabic words. This suggests that children with ASD encounter difficulty with the grammatical use of prosody, which relates to the appropriate placement of stress within words and sentences. Previous studies have also found that individuals with ASD demonstrate difficulty with lexical stress (McCann et al., 2007; Paul et al., 2005). Specifically, McCann et al. (2007) found that children with ASD misassigned stress placement, often stressing the first syllable within a word. In our study, all but one child with ASD produced inappropriate stress cues within utterances, indicating deficits in the pragmatic use of stress. This pattern is consistent with results described by Baltaxe (1984), Fine et al. (1991), and Shriberg et al. (2001), who also reported significant increases in the inappropriate use of contrastive stress in children with ASD.

In contrast to several previous studies, the ASD group in the current study did not differ from the TD group with regard to the production of rate, loudness, and pitch. Results from the current study support previous findings by Grossman et al. (2010), who found no significant differences in the pitch or loudness of children with HFA when compared to TD children. However, several studies have described atypical pitch accents and timing patterns in the speech of children with ASD (Bonneh et al., 2011; Diehl et al., 2009; Fosnot & Jun, 1999; Sharda et al., 2010). Fosnot and Jun (1999), for example, found that the speech of children with ASD was most deviant from that of TD children in terms of duration, intonation, and pitch during reading and imitative tasks. Specifically, the authors found that declarative and question sentences read by children with ASD were longer in duration than those of children who stutter and TD children. The same results were found for the imitative tasks. Fosnot and Jun also reported that the children with ASD did not use proper intonation patterns, making it difficult to distinguish declarative sentences from interrogative sentences during the reading task. Although these results were not replicated in the current study, the two studies differed in the method of data collection. That is, Fosnot and Jun used reading passages and imitative speech tasks to obtain a spontaneous speech sample, whereas the current study used a variety of structured play activities.

Results from a study by Diehl et al. (2009) also differed from the current study. Diehl et al. used acoustic and perceptual measures to analyze the fundamental frequency of children and adolescents with HFA. The authors found that both age groups exhibited increased fundamental frequency variation and a higher fundamental frequency average when compared to TD children. Perceptual analysis of pitch for the current study did not find any significant differences between groups with regard to pitch variation.

Several reasons may account for the differences between results from the studies discussed here and the current study. First, the children included in the current study were between the ages of 24 and 68 months, which is a younger age group than those included in previous studies of individuals with ASD. At this young age, TD children may not have
mastered the prosodic cues necessary for differences between these two populations to be determined. Although several research studies have suggested that children begin to integrate prosodic cues into their speech before the age of 2 years, differences exist in the literature concerning the age at which children begin to master this aspect of communication (Allen & Hawkins, 1980; Gerken & McGregor, 1998; Klein, 1981; Snow, 1994). Second, a significant number of previous studies used structured reading passages, picture naming, and imitative tasks to elicit speech samples (Baltaxe, 1984; Bonneh et al., 2011; Fine et al., 1991; Fosnot & Jun, 1999; Sharda et al., 2010; Van Santen et al., 2010). In contrast, the current study used spontaneous play speech samples to determine whether or not atypical prosodic characteristics were present in the speech of children with ASD. This method was used because spontaneous speech samples are more likely to be representative of a child’s speech and language abilities than imitative tasks or structured reading passages.

Of significant interest is the fact that both the ASD and TD groups exhibited high instances of inappropriate rate codes in comparison to all of the other PV codes. According to the PVSP, appropriate rate is defined as utterances that contain two to four syllables per second. Of the four PVSP rate codes, both groups exhibited more instances of fast rate than any other rate code. These results stand in contrast to those described by Grossman et al. (2010), who found that children with HFA produced longer word productions than TD children. Additionally, Grossman et al. noted that participants with HFA produced more pauses between syllables than the TD group, making word production sound awkward and disfluent. Results from the current study did not indicate that children with ASD produced utterances that are longer in duration. Instead, both the ASD group and the TD group produced a large number of utterances with a fast rate. This may have been caused by moments of excitement during testing, in which a variety of toys of high interest were used in order to prompt and facilitate communication. This fast rate may also be more characteristic of younger children.

Finally, our results indicated that significantly more children with ASD than TD failed to surpass the 90% cutoff criterion for the stress PV code. However, there were no significant differences between the two groups with regard to the number who failed to surpass the 90% cutoff criterion for phrasing, rate, loudness, pitch, laryngeal features, or resonance PV codes. These results are similar in some aspects to those obtained by Shriberg et al. (2001), who found that significantly more speakers with HFA and AD failed the stress PV code (HFA = 53.3%, AD = 26.7%) compared to TD speakers (5.7%). However, Shriberg et al. also found that significantly more speakers with HFA and AD failed the resonance PV code (HFA = 40.0%, AS = 26.7%) than TD speakers (1.9%), and significantly more speakers with AD (66.7%) failed the phrasing PV code than TD speakers (26.4%). Those results were not replicated in the current study. Neither group in our study had any utterances coded with inappropriate resonance. Additionally, both groups had a similar number of utterances coded as inappropriate phrasing (ASD = 9.71%, TD = 6.29%). Again, this difference may be explained by the difference in the age groups that were included in the current study and the group included in the study by Shriberg et al. Results from the current study stand in agreement with research by Yairi (2005), who found that the speech of young preschool children is more disfluent than the speech of school-age children. This is likely due to the fact that at a young age, the motor planning of speech is developing. Therefore, the speech of young children may be more awkward and disfluent than the speech of older children.

The results of the current investigation provide preliminary evidence that differences in stress patterns may be the only prosodic red flag of ASD for young children. Because no differences were found in rate, loudness, or pitch between the two groups, it appears that deficits in these areas of prosody may not be as readily apparent in young children. The possibility exists that even TD children may not have mastered the prosodic cues to the level necessary to differentiate the groups for children less than 68 months. The findings of the current investigation highlight the importance of considering prosody, specifically stress, when assessing and treating young children with ASD. Prosody measures should be included in the assessment battery when determining a diagnosis of ASD because misplaced stress may serve as a supralinguistic marker of possible ASD in young children. As prosodic differences could hinder affective and pragmatic communication, they should also be considered in intervention planning; however, to date, there has been little research on treating prosodic deficits in children with autism.

Limitations and Future Directions

Because the present study included a small number of participants, it will be important for future studies to include a larger sample of children with ASD in order to better understand the prosodic patterns that are present in this population. It would also be interesting to determine whether echoed utterances, which were excluded from the current study, differ...
from non-echoed utterances based on PVSP variables. Currently, it is not clear whether there are specific atypical prosodic patterns associated with echoed utterances that contribute to the perception of atypical prosody in young verbal children with ASD. Research is needed to differentiate the prosodic patterns of spontaneous versus echoed utterances of children with ASD.

Further investigation should also be conducted to compare the type and frequency of atypical prosodic patterns of children with ASD to those of other populations, such as children with developmental disabilities. This analysis would help determine whether or not the prosodic patterns observed are related to a general developmental delay or if the patterns are specific to children with ASD. Additionally, this information would provide clinicians with important information concerning the prosodic characteristics of a given population and whether or not these prosodic patterns warrant treatment. Also of interest would be to compare the prosody characteristics of each child with ASD over several years in order to better understand the development and progression of prosody within this population. In doing so, researchers and clinicians could obtain information regarding the development of prosody in children with ASD, whether it improves or becomes more atypical with age, and whether or not there are any unique developmental patterns in this population of children.

Finally, as there is a lack of literature on interventions for expressive prosody in children with ASD, further research is warranted to investigate the challenges associated with treating prosody in this population. A primary challenge is modeling prosodic patterns in a population known to frequently demonstrate echolalia. Therefore, the clinician runs the risk that a child may communicate in a way that sounds “canned” or artificial. Studies examining these challenges would be helpful to practicing clinicians who wish to address prosody in their intervention plans for children with ASD. For example, as opposed to directly targeting expressive prosody through modeling, would improving a child’s receptive prosody have an impact on his or her expressive prosody?

**Conclusion**

Currently, clinicians and caregivers look for evidence of a communication delay in verbal children who are suspected to have ASD (Wetherby & Prizant, 1996; Wetherby, Goldstein, Cleary, Allen, & Kublin, 2003). Because there may be no apparent delay in the spoken language of verbal children with ASD, it is difficult for clinicians to identify these children early in development. Young verbal children with ASD may obtain scores within normal limits on standardized speech and language tests administered to preschool-age children and therefore may not be identified as having a communication delay. This apparent issue necessitates more precise indicators of ASD in order to improve early and accurate diagnosis. Results from this study suggest that atypical stress patterns consistently appear in the speech of children with ASD, which coincides with previous data collected (Baltaxe, 1984; McCane et al., 2007; Shriberg et al., 2001). It may therefore be beneficial to evaluate the frequency and type of stress patterns present in the speech of children suspected of ASD. Such an evaluation might provide clinicians with an additional diagnostic tool.

In summary, the results of this study indicate that the prosodic patterns of young children with ASD do not differ significantly from those of TD children, with the exception of grammatical and pragmatic stress. Although these results differ from those of previous studies, the current study involved a much younger age group. Therefore, it is difficult to determine whether or not the participants had reached a level of mastery with regard to the use of prosodic cues. Additionally, two of the prosody and voice characteristics measured in the current study involved the use of instrumental assessment, which may provide more accurate results than the perceptual measures used in a significant number of previous studies (Baltaxe, 1984; McCaleb & Prizant, 1985; Paul et al., 2005; Shriberg et al., 2001). Although the majority of the prosodic cues assessed did not differ between groups, the presence of atypical stress patterns within the ASD group may serve as an early indicator of ASD in verbal children and is a subject that necessitates further investigation in future research.

**REFERENCES**


Contact author: Ashlynn McAlpine, 68 Harvard Street, Brookline, MA 02445. E-mail: ashlynn.campagna@gmail.com