Lexical and Affective Prosody in Children With High-Functioning Autism

Ruth B. Grossman
University of Massachusetts Medical School Shriver Center, Waltham

Rhyannon H. Bemis
University of New Hampshire, Durham

Daniela Plesa Skwerer
Helen Tager-Flusberg
Boston University School of Medicine, Boston, MA

Purpose: To investigate the perception and production of lexical stress and processing of affective prosody in adolescents with high-functioning autism (HFA). We hypothesized preserved processing of lexical and affective prosody but atypical lexical prosody production.

Method: Sixteen children with HFA and 15 typically developing (TD) peers participated in 3 experiments that examined the following: (a) perception of affective prosody (Experiment 1), (b) lexical stress perception (Experiment 2), and (c) lexical stress production (Experiment 3). In Experiment 1, participants labeled sad, happy, and neutral spoken sentences that were low-pass filtered, to eliminate verbal content. In Experiment 2, participants disambiguated word meanings based on lexical stress (*HOTdog* vs. *hot DOG*). In Experiment 3, participants produced these words in a sentence completion task. Productions were analyzed with acoustic measures.

Results: Accuracy levels showed no group differences. Participants with HFA could determine affect from filtered sentences and disambiguate words on the basis of lexical stress. They produced appropriately differentiated lexical stress patterns but demonstrated atypically long productions, indicating reduced ability in natural prosody production.

Conclusions: Children with HFA were as capable as their TD peers in receptive tasks of lexical stress and affective prosody. Prosody productions were atypically long, despite accurate differentiation of lexical stress patterns. Future research should use larger samples and spontaneous versus elicited productions.

KEY WORDS: autism, prosody, lexical stress, affective prosody, perception, production

*P*rosody is a suprasegmental device that can best be described as the "melody" or "rhythm" of speech. It is a complex vocal signal composed primarily of the pitch, intensity, and duration of an utterance. Despite its complexity, typically developing (TD) children and adults are able to perceive and comprehend this prosody automatically (Shriberg & Kent, 2003) and do so from a very early stage in development (Jusczyk, Cutler, & Redanz, 1993; Mehler et al., 1988). There is evidence to suggest that children accurately perceive and use prosodic cues to inform their acquisition of expressive vocabulary from the first-word stage through the preschool years. Some of the typical errors in early word production shown by toddlers, such as weak syllable omission, can be traced directly to their interpretation of prosodic cues (Gerken & McGregor, 1998).

For individuals with autism spectrum disorder (ASD), however, this nonverbal aspect of communication poses a much greater challenge. Abnormal prosody production has been one of the hallmark characteristics of autism since its first description by Kanner (1943/1968), and it continues to be a consistent aspect of the ASD communication profile (Baltaxe & Simmons, 1985, 1992; Rapin & Dunn, 1997; Shriberg et al., 2001). Many

observations have described expressive prosody in this population as flat, monotonous, or abnormally modulated (Baltaxe, 1984; Fay & Schuler, 1980). Atypical prosody production also contributes to the perception of reduced social and communicative competence in individuals with ASD, making it one of the earliest and most salient features of abnormal communication noted by typical listeners (Paul, Shriberg, et al., 2005).

Prosody serves several functions in speech, including, but not limited to, the transmission of affect and the marking of grammatical and lexical constructs (McCann & Peppé, 2003). *Affective prosody* refers to a speaker varying the pitch and rate of an utterance to indicate his or her emotional state. *Grammatical prosody* encompasses the use of pitch to indicate the type of statement being made (e.g., a rising intonation at the end of an utterance defines it as a question) and the application of lexical stress, which is the emphasis of one syllable over another in order to convey or disambiguate meaning (e.g., *HOTdog* is a type of food, whereas *hot DOG* is an overheated canine; Merewether & Alpern, 1990).

Many studies of prosody in ASD tend to focus on single aspects of prosody, for example, looking either at production or perception, or focusing only on affective prosody or only on the use of emphatic stress, and the findings from these studies are often contradictory. Some studies of affective prosody have shown that children with ASD are less able than their TD peers to use prosody to match simple emotions from social scenes with pictures of emotional facial expressions (Lindner & Rosén, 2006) or determine complex states of mind from sentences (Golan, Baron-Cohen, Hill, & Rutherford, 2007). On the other hand, Boucher, Lewis, and Collis (2000) found that children with autism were equal to their TD peers in the ability to label six basic emotions based on the affective prosody of single-word utterances (e.g., days of the week), despite being deficient in matching the prosodically expressed emotion to a photograph of an emotional face. The differences in these results may be based on the varied methodologies and stimulus selections, ranging from single words and word lists to complete sentences and using labeling versus face matching.

Investigations of lexical stress production and perception have also yielded an uneven picture of prosodic skills in ASD. TD children appear to have mastered the ability to disambiguate compound nouns (*hotdog*) from noun phrases (*hot dog*) using lexical stress cues by the age of 10 (Cruttenden, 1974; Vogel & Raimy, 2002) or 12 (Atkinson-King, 1973) and can produce appropriate lexical stress even earlier (Cutler & Swinney, 1987). Smith and Robb (2005) asked TD children and children with speech delay aged 5 through 7 years to repeat novel two-syllable words with stress either on the first or second syllable and used the acoustic measure of whole-word duration to capture their productions. Both groups produced

second-syllable stress words that were, overall, longer than those with first-syllable stress, indicating that whole-word duration is a valid measure of lexical stress production in children.

Children with ASD appear to share many of those skills, being able to recall stressed words better than unstressed words (Fine, Bartolucci, Ginsberg, & Szatmari, 1991) and showing no statistical difference from TD peers in disambiguating verbs (e.g., *reCALL*) from nouns (e.g., *REcall*; Paul, Augustyn, Klin, & Volkmar, 2005). Their main deficits appear to lie in the production of prosody, and several studies have noted the inappropriate production of contrastive stress (Baltaxe, 1984; Fine et al., 1991; see McCann & Peppé, 2003, for a review), inappropriate phrasal stress (Baltaxe & Guthrie 1987; Baltaxe & Simmons, 1985; McCaleb & Prizant, 1985), or excessive stress in spontaneous speech (Shriberg et al., 2001). There are also data to suggest that people with ASD have greater variation in the use of prosodic pitch both within an utterance and across individuals (Baltaxe, 1984; Fosnot & Jun, 1999); however, most assessments of inappropriateness have relied on subjective coding instead of acoustic analyses.

A few recent studies have gone a long way toward capturing a more complete picture of prosody perception and production in ASD. Shriberg et al. (2001) used the Prosody-Voice Screening Profile (Shriberg, Kwiatkowski, & Rasmussen, 1990), which measures production for a range of prosodic functions, and found few areas of deficits, which were concentrated mostly in fluency, phrasing, and hypernasality. Paul, Shriberg, et al. (2005) expanded on this work and determined that there was a strong correlation between these prosodic differences and an increased perception of communication and socialization deficits in this population, as measured by the Autism Diagnostic Observation Schedule (ADOS)—Generic (Lord et al., 2000) and the Vineland Adaptive Behavior Scales—Survey Form (Sparrow, Balla, & Cicchetti, 1984). Paul, Augustyn, et al. (2005) used a comprehensive set of 12 tasks to assess the production and perception of grammatical and affective prosody in a group of children with ASD. They determined that the ASD group showed significant deficits in the perception and production of emphatic stress and the production of grammatical stress, but not in the processing of grammatical and affective intonation or in grammatical phrasing. The authors attributed this lack of group differences to possible ceiling effects for some of the tasks included in the battery. They also noted that objective measures of prosody production would be helpful in furthering our understanding of prosody competence in ASD.

Using a measure of prosody production and perception across affect and grammatical usage, the Profiling Elements of Prosodic Systems in Children (PEPS–C, Peppé & McCann, 2003), Peppé, McCann, Gibbon, O'Hare,

and Rutherford (2006) showed that the prosodic deficits of a boy with high-functioning autism (HFA) could be captured and differentiated from the standardization sample of TD children. In this case study they documented a tendency for misplaced stress and less accurate performance on affective prosody, as well as relatively poorer performance on short items (words) over longer items (phrases or word lists). Applying the PEPS–C to a large sample of school-age children with HFA, McCann, Peppé, Gibbon, O'Hare, and Rutherford (2007) and Peppé, McCann, Gibbon, O'Hare, and Rutherford (2007) described systematic deficits in this population, including reduced accuracy for the perception and production of affective prosody in single words (food items expressed with like or distaste matched with happy and unhappy faces) and the production of contrastive stress. The prosody skills of both the HFA and TD groups were strongly correlated with expressive and receptive language scores, indicating that mastery of prosody is an integral part of language function.

Although the PEPS–C assesses most aspects of prosody, it does not contain a measure of *lexical stress,* which is a prosodic function more directly related to grammatical processing than either contrastive or emphatic stress and therefore a potential area of strength for high-functioning individuals with ASD whose language skills are in the normal range. The PEPS–C contains a measure for *chunking,* which refers to the use of pauses to indicate whether a list has three items (e.g., chocolate, cookies, and milk) or two items (e.g., chocolate cookies and milk). This is, however, a different process than the use of selective intensity (i.e., stress) to determine the meaning of an ambiguous word, such as *hot dog* versus *hotdog,* which is one of the prosodic structures we investigated in the research reported in this article. Also, the affective prosody tasks of the PEPS–C are based on productions of single food words, indicating like or dislike, thereby limiting the affective prosody information to a very brief utterance instead of the wealth of prosodic information typically contained in natural, emotionally laden, sentence-length speech. The aim of this article is to add to the current knowledge base of prosodic ability in individuals with ASD by presenting tasks of affective prosody using sentence-length stimuli as well as tasks investigating receptive and expressive aspects of lexical stress in two-syllable ambiguous word pairs. We used objective measures of stress production (intensity, pitch, and whole-word duration) to reduce reliance on subjective coding to determine group differences.

Hypotheses for the Three Experiments

On the basis of this population's documented deficits in affective prosody matching and labeling for single words (McCann et al., 2007; Peppé et al., 2007), but potentially preserved skills for sentences (Paul, Augustyn, et al., 2005), we expected children and adolescents with HFA to be able to use the richness of whole-sentence prosodic information to determine affective content, even when we eliminated all verbal content from the stimuli.

Experiment 1. Our hypothesis for the affective prosody task in Experiment 1 is that children with HFA would perform at a level equal to that of their TD peers on a task using sentences that have been filtered to eliminate verbal content but maintain prosodic contours. We expected that both groups would show higher accuracy scores for affective sentences with verbal content than for those without.

Experiment 2. In Experiment 2, we expected, on the basis of the HFA group's high verbal language scores and the previously demonstrated connection between language skills and prosodic ability (McCann et al., 2007), combined with the highly grammatical nature of the lexical stress stimuli, that this group would be able to use lexical stress information to disambiguate compound nouns (e.g., *wetsuit*) from noun phrases (e.g., *wet suit*). We therefore hypothesized that the HFA group would perform at a level equal to that of their TD peers when asked to use lexical stress to determine the meaning of an ambiguous word pair.

Experiment 3. In Experiment 3, on the basis of the notion that a deficit in prosody production is one of the few consistently reported results in the autism prosody literature, with both quantifiable and subjective deficits noted, we expected that participants with HFA would show quantifiable deficits in lexical stress production for ambiguous word pairs. We also hypothesized that this expressive prosody difference would be captured in the objective acoustic measure of whole-word duration.

Experiment 1: Perception of Affective Prosody in Filtered and Unfiltered Sentences

To test perception of affective prosody, the lexical content of the stimulus needs to be controlled (Peppé et al., 2006). In Experiment 1, we used sentences that were filtered to remove all recognizable semantic content but contained fully preserved prosody. Our aim was to assess whether children and adolescents with HFA are able to make affect judgments based solely on a speaker's emotional prosody without any assistance of linguistic content. We also wanted to compare accuracy for these filtered sentences with accuracy in a baseline comparison task using unfiltered sentences with semantically biased lexical content.

Method
Participants

Two groups participated in this study: (a) children and adolescents with HFA (n = 16) and (b) TD control children (n = 15). The participants were recruited through advocacy groups for parents of children with autism; local schools; and advertisements placed in local magazines, newspapers, and the Internet.

Diagnosis of autism. The participants in the autism group met *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; American Psychiatric Association, 1994) criteria for autistic disorder, based on expert clinical impression and confirmed by the Autism Diagnostic Interview—Revised (Lord, Rutter, & LeCouteur, 1994) and the ADOS (Lord, Rutter, DiLavore, & Risi, 1999), which were administered by trained examiners. Participants with known genetic disorders were excluded. To create a group of participants with a narrowly constrained diagnosis, we excluded all participants who scored in the range of ASD, pervasive developmental disorder not otherwise specified, or Asperger syndrome and kept only that subgroup of participants who were diagnosed as falling in the "full autism" range of the spectrum.

Standardized testing. We used the Kaufman Brief Intelligence Test, Second Edition (Kaufman & Kaufman, 2004) to assess IQ in all participants and the Peabody Picture Vocabulary Test—Revised (Dunn & Dunn, 1981) to measure receptive vocabulary. Participants were selected on the basis of standardized scores within the normal range to further create a more homogeneous and high-functioning group of participants with autism. The descriptive characteristics of both groups can be found in Table 1. Using a multivariate analysis of variance (ANOVA), with group as the independent variable, we verified that the HFA and TD groups did not differ significantly in age, $F(1, 30) = 0.046$, $p = .83$; verbal IQ, $F(1, 30) = 1.763$, $p = .195$; nonverbal IQ, $F(1, 30) = 0.761$,

$p = .39$; and receptive vocabulary ability, $F(1, 30) = 0.598$, $p = .445$.

Language competence in the HFA group. All participants with HFA had receptive vocabulary and verbal IQ within 2 *SD*s of the mean as assessed through standardized measures, which placed them in the high-functioning range of the spectrum. Their social, pragmatic, communication, and other deficits, however, were significant enough to reach the diagnostic threshold for full autism on the ADOS. It is important to note that most standardized measures of language do not include detailed measures of prosodic or pragmatic competence, meaning that an individual may have a normal verbal IQ but still exhibit great difficulty perceiving and producing nonverbal language, such as prosody or affective facial expressions.

Materials

The stimuli used in the affective prosody task were complete, natural sentences containing commonly used words. We used sentences spoken by a female native English speaker that were previously used by Plesa Skwerer, Schofield, Verbalis, Faja, and Tager-Flusberg (2007). Six different sentences were produced in each of the three emotions (happy, sad, and neutral). In accordance with established features of affective prosody (Bachorowski & Owren, 1995; Banse & Sherer, 1996; Cosmides, 1983; Murray & Arnott, 1993), Plesa Skwerer et al. selected the 18 sentences that best portrayed their target emotions, verified by acoustic characteristics as measured using Praat software (Boersma & Weenink, 2005). Happy prosody samples were verified to contain higher pitch, faster rate, and a complex tone ending (a rise–fall pattern). Sad sentences had lower pitch, slower rate, and a low tone ending, and neutral sentences were portrayed by midrange pitch, accent on the main topic, and a less complex final tone. In addition to these acoustic measures, the stimuli were rated by 10 typical individuals for accuracy of expressed affect, confirming that the selected sentences clearly and naturally portrayed the target emotion without exaggeration. All sentences contained a final phrase indicating a label of the target emotion, for example, "When the kids tease Sue, she's upset," which was designed to enable participants to rely heavily on semantic cues for affect determination in the unfiltered task and thereby provide a significant contrast in processing strategy to the filtered sentences, which had no verbal content at all. (A complete list of unfiltered stimuli is provided in Appendix A.) All 18 sentences were of approximately equal length and followed the same syntactic structure. The sentences were then low-pass filtered to delete speech frequencies above 100 to 150 Hz. We chose this cutoff point to ensure that all semantic content was completely removed while fully preserving prosodic contours. We pseudorandomized the

Table 1. Descriptive characteristics of participant groups.

Characteristic	HFA[a]			TD[b]		
	M	*SD*	Range	*M*	*SD*	Range
Age[c]	12;4	2;3	7;6–17	12;7	3;1	7;6–18
Full-Scale IQ	106.7	10.6	87–123	108.9	11.3	87–123
Verbal IQ	101.2	14.3	83–127	108.1	14.6	81–127
Nonverbal IQ	109.6	19.1	94–127	106.7	9.8	85–116
PPVT–R	107.0	15.4	79–138	111.3	15.3	79–139

Note. HFA = high-functioning autism; TD = typically developing; PPVT–R = Peabody Picture Vocabulary Test—Revised (Dunn & Dunn, 1981).

[a]n = 16. [b]n = 15. [c]Years; months.

18 filtered sentences into three different counterbalanced sequences for the *filtered task,* and administration of those sequences was alternated between participants. We also pseudorandomized and counterbalanced the 18 unfiltered sentences for the *content task.* Additional sentences were used for training. Each participant was given the filtered task before the content task. All stimuli were presented by means of a PC with standard speakers. Responses were recorded using a Cedrus button box.

Procedure

All participants completed a short training session containing only filtered sentences to ensure that they understood the task. Participants were instructed that they would hear "strange-sounding" sentences and were asked to determine how the speaker was feeling, with the available choices being "happy," "sad," or "neutral." They were further explicitly instructed not to try and decipher what the speaker was saying. Responses were made on a button box, with the leftmost button representing "happy," the second button "sad," and the third button "neutral." Buttons were marked with line drawings of faces representing the target emotion. At the beginning of the experimental task, participants were reminded to listen to how the speaker was feeling and not what she was saying. After completing the filtered task, each participant was presented with the content task, in which the same design and procedures were used. The unfiltered content stimuli were included to provide a control task for participants to demonstrate their ability to follow the task instructions and process meaningful language stimuli. Our aim was to contrast performance on this control task with performance on the filtered task, to determine differences in emotion-processing abilities for prosody-only stimuli as opposed to those carrying full verbal and vocal language information.

Results

Mean accuracy scores presented as percentages correct for the filtered and content tasks are shown in Table 2. Both group's accuracy scores were significantly above the 33.3% chance level for the filtered task, $t(15) = 9.03$, $p < .001$, in the HFA group, and $t(14) = 10.2$, $p < .001$, in the TD group, and the content task, $t(15) = 19.7$, $p < .001$, in the HFA group, and $t(14) = 23.7$, $p < .001$, in the TD group, based on paired samples t tests. We conducted a 2 (Group) × 2 (Task) × 3 (Emotion) repeated measures ANOVA (data were normally distributed) with percentage correct as the dependent variable and found a significant main effect of task, $F(1, 29) = 16.6$, $p < .001$, partial $\eta^2 = .36$, with performance better on the content task than the filtered task in both groups, $t(15) = 3.4$, $p < .001$, Cohen's $d = 1.04$, for the HFA group, and $t(14) = 2.4$,

Table 2. Experiment 1: Percentage correct for perception of affective prosody.

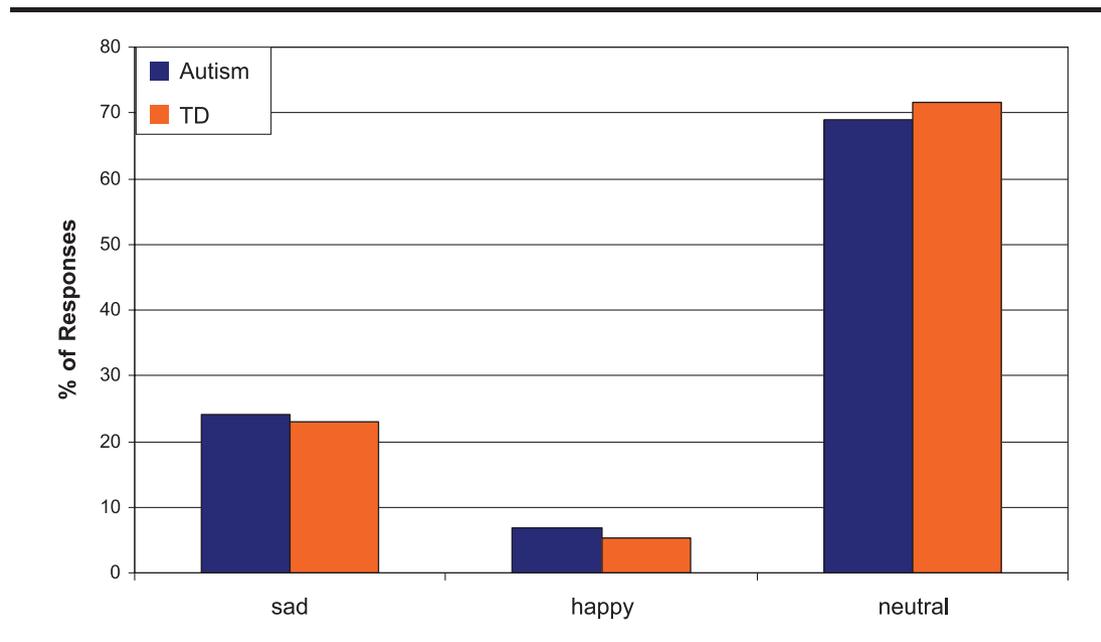| Stimulus	HFA[a]		TD[b]	
	M	SD	M	SD
Filtered happy	72.9	20.9	86.7	9.3
Content happy	89.6	13.4	95.6	7.6
Filtered sad	86.5	18.5	82.2	23.1
Content sad	89.6	16.0	95.6	9.9
Filtered neutral	61.5	24.9	65.6	27.1
Content neutral	83.3	18.3	74.4	26.6
Filtered overall	73.6	17.9	78.1	17.1
Content overall	87.5	11.0	88.5	9.0

[a]$n = 16$. [b]$n = 15$.

$p = .03$, Cohen's $d = 0.76$, for the TD group, based on paired samples t tests. There was also a main effect of emotion, $F(2, 58) = 20.26$, $p < .001$, partial $\eta^2 = .41$, with lowest performance on the neutral sentences in both tasks for both groups. An error analysis of neutral sentences showed that both groups were more than 68% accurate in categorizing neutral sentences, but when they did make mistakes they were more likely to incorrectly label neutral expressions as sad than happy (see Figure 1). There was no main effect of group, $F(1, 29) = 0.44$, $p = .51$, partial $\eta^2 = .02$, or a significant Group × Task, $F(1, 29) = 0.35$, $p = .56$, partial $\eta^2 = .01$, or Group × Emotion, $F(1, 29) = 2.3$, $p = .11$, partial $\eta^2 = .07$, interaction.

We also conducted a 2 (Group) × 3 (Emotion) repeated measures ANOVA within each task (content and filtered), which revealed a main effect of emotion in the content task, $F(2, 58) = 7.6$, $p = .001$, partial $\eta^2 = .21$, and the filtered task, $F(2, 58) = 16.2$, $p < .001$, partial $\eta^2 = .36$. In both tasks the neutral sentences achieved less accuracy than the sentences containing either happy or sad affective prosody. There was no significant difference between groups, $F(1, 29) = 0.52$, $p = .47$, partial $\eta^2 = .02$, for filtered stimuli, and $F(2, 58) = 0.01$, $p = .78$, partial $\eta^2 = .003$, for content stimuli, and no Group × Emotion interaction, $F(2, 58) = 2.7$, $p = .07$, partial $\eta^2 = .09$, for filtered stimuli, and $F(2, 58) = 2.24$, $p = .12$, partial $\eta^2 = .07$ for content stimuli.

Discussion

As expected, both groups were significantly more accurate on the content task than on the filtered task. One surprising aspect of this study was that the TD group scored relatively poorly on the filtered task. It is possible that the large age range of our participants (7;6 [years; months]–18, $M = 12;7$) meant that some members of the

Figure 1. Responses for filtered neutral sentences in all response categories, Experiment 1. TD = typically developing group.

TD group were too young to successfully perform this task, thereby lowering the group's accuracy scores. However, both groups' mean accuracy scores were significantly above chance, indicating that overall, they understood the task and were able to interpret the stimuli accurately. The filtered sentences are an unnatural presentation of language, providing affect information exclusively in the voice, and therefore are much more difficult to interpret than natural sentences with intact verbal content, resulting in lower accuracy scores across groups. However, the results of the filtered speech task do show that both participant groups were equally able to discern happy, sad, and neutral emotions in sentences using only prosodic cues, thereby confirming our main hypothesis. Both groups were also more accurate in categorizing stimuli with an affective overlay than neutral sentences, paralleling the results found by Plesa Skwerer et al. (2007), who used the same stimuli with a group of adolescents and adults with Williams syndrome, learning or intellectual disabilities, and TD control participants. These data suggest that the absence of affective information may be more difficult to interpret for TD individuals as well as participants with a range of communication and/or cognitive deficits. Further studies are required to determine whether this neutral affect disadvantage is robust across different methodologies and stimulus types.

The data also confirm our secondary hypothesis that the children with HFA would be as able as their TD peers to use the addition of semantic information to perform more accurately on the unfiltered sentences than on the filtered sentences. These data are consistent with the results of prior studies showing that individuals with HFA are able to label basic emotions using prosodic cues of word lists (e.g., days of the week; Boucher et al., 2000) and full sentences (Paul, Augustyn, et al., 2005), but they contradict data indicating deficits in single-word affective prosody processing among children with ASD (McCann et al., 2007; Peppé et al., 2007). In contrast to these latter studies, the filtered and unfiltered sentences in this study provided additional prosodic cues sustained over longer time periods than the comparatively short duration of a single spoken word. Our data suggest that the added complexity and prosodic information of sentences may assist participants with HFA with the decoding of affective prosody. It is also important to note that the participants in our studies were chronologically a few years older than the participants in McCann et al.'s (2007) and Peppé et al.'s (2007) studies, whose participants were, on average, 9;4 and 9;10, respectively. Although the mean age of our participants is only 3 years older, this difference is meaningful because it implies that most of the participants in our study were expected to have reached maturity in prosody processing ability. Furthermore, the participants with ASD included here all had verbal IQ and language scores within 2 *SD*s of the mean, meaning that—in contrast to McCann et al.'s; Peppé et al.'s; and Paul, Augustyn, et al.'s (2005) studies—our results are based on a cohort of adolescents with HFA who did not at this time exhibit measurable language impairment.

Our data conflict with the results of Golan et al. (2007), who also used full sentences but found deficits in affective prosody recognition among adults with ASD. This difference can be explained by the fact that we included only two basic emotions that contrasted in valence (happy, sad), whereas Golan et al. (2007) used complex emotional states, such as "defensive," "joking," "unconcerned," or "indignant," and a choice among four, rather than three, possible responses. It may well be that individuals with autism have a preserved ability to discern basic emotions from affective prosodic cues alone but fail in their interpretation of affective prosody for more complex and subtle emotions such as those Golan et al. tested.

Overall, the children and adolescents with HFA in our study labeled happy, sad, and neutral emotions in sentence-length, filtered speech stimuli as well as their TD peers. Further studies using more emotion types and larger participant groups are necessary to determine whether there are group differences that were not fully expressed in our study, most likely because of the small sample sizes.

Experiment 2: Perception of Lexical Prosody

The purpose of the receptive lexical prosody task was to assess the ability of children and adolescents with HFA to use lexical stress alone to determine the meaning of the same ambiguous word pairs originally designed by Atkinson-King (1973) and adapted by Plesa Skwerer et al. (2007) for use with adolescents and adults with Williams syndrome and learning or intellectual disabilities. Our aim was to expand the application of these proven stimuli to assess the comprehension of lexical stress in children with HFA.
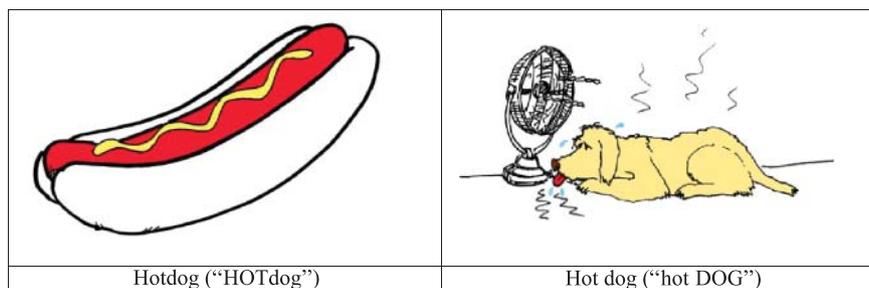
Method

Participants. Participants were the same sample of children and adolescents as in Experiment 1.

Materials. The lexical ambiguity task contained 22 two-syllable word pairs, 11 of them with ambiguous meaning, for example, *PICKup* (a type of truck) and *pick UP* (taking an item off the floor) and 11 foil words with only one possible meaning, for example, *T-shirt*. (The full list of stimuli is presented in Appendix B.) The target words were recorded by a female native English speaker at least three times each with stress on the first syllable, stress on the second syllable, and equal stress on both syllables. This last stress pattern was a control condition, because equal stress is not a valid form of lexical stress in American English. The first- and last-syllable stress productions determined the two possible meanings of the ambiguous word pairs (e.g., *HOTdog* vs. *hot DOG*). To ensure that each word pair was produced correctly, Plesa Skwerer et al. (2007) acoustically analyzed each stimulus recording for duration, mean intensity, and average pitch (F0) using Praat software. In accordance with established guidelines for acoustic properties of stressed syllables (Klatt, 1976; Lieberman, 1960), Plesa Skwerer et al. verified that all samples in the stimulus list adhered to the criteria of longer duration, higher pitch, and greater intensity of the stressed syllable. Acoustic analyses done with Praat confirmed that all stimuli displayed this distinction in both the first-syllable and last-syllable stress items, but—appropriately—not the equal stress items. The acoustically verified stimuli were then presented to 10 raters, and items that incurred consistent errors were eliminated from the stimulus list. The 11 foil words were recorded by the same speaker with their correct first-syllable stress pattern as well as an equal stress pattern, to mask the experimental manipulation.

Two picture stimuli were created by an amateur artist for each of the two-syllable words. Every target word was combined with one picture representing the correct meaning and another indicating the meaning resulting from the opposite stress pattern (e.g., *hotdog* vs. *hot dog*; see Figure 2). The foil words were paired with one picture representing the correct meaning of the word and another picture representing a feature of the foil item without matching the meaning of the target stimulus

Figure 2. Sample images showing the meaning of compound nouns and noun phrases for lexical stress perception, Experiment 2.

| Hotdog ("HOTdog") | Hot dog ("hot DOG") |

(e.g., *rainbow* was paired with a picture of a rainbow and a picture of a hair bow). The pictures were piloted in an open-ended task to ensure that they conveyed their intended meaning clearly. The experimental task was presented in two segments separated by a break. The words were pseudorandomized and counterbalanced so that each word pair was presented in all three stress patterns during each half of the experiment, for a total of 66 experimental trials. The 11 foil words were presented at least once in each stress pattern for a total of 29 foil trials. We created two pseudorandomized sequences of the experiment to ensure that the same word pair did not appear twice in a row, and we counterbalanced them across participants. Participant responses were recorded via a Cedrus button box.

Procedure. Before beginning the experimental trials, all participants completed a three-item training session to ensure that they understood the task. Participants saw two pictures side by side and heard a sentence via speakers containing the target word (e.g., "Take a copy of the *HANDout*"). This was followed by a repetition of the two-syllable target (*HANDout*) in isolation. We then asked participants to select the picture that matched the meaning of the target word by pressing the left or right button on the button box, corresponding to the picture on the left or right of the screen. When participants demonstrated competence on the task in training, they moved on to complete the experimental task. The experimental stimuli consisted of prerecorded presentations of the ambiguous stimulus words/noun phrases without contextual sentences, to ensure that prosodic stress assignment was the only cue participants could use to accurately determine which picture to choose. The prosody of each stimulus presentation clearly stressed one syllable over the other, making the intended meaning of the target word clear, meaning that the participant's choice between the two pictures was determined exclusively by attending to the lexical stress pattern of the presented target word. We reminded participants to listen carefully to how the words were spoken and choose the picture accordingly. Corrective feedback was given during training but not during the experimental task.

Results

All participants were at least 80% accurate on the foil items. We excluded all equal-stress items from the final analysis, because they were not accurate representations of lexical stress in American English but instead served as a prosodic foil. Mean accuracy levels of experimental trials in percentage-correct form for both stress patterns are presented in Table 3. Both groups' accuracy levels for first- and last-syllable stress items were significantly above the 50% chance accuracy level:

Table 3. Experiment 2: Percentage correct for perception of lexical prosody.

Type of stimulus	HFA[a]		TD[b]	
	M	SD	M	SD
First-syllable stress	79.8	14.77	69.1	13.4
Last-syllable stress	59.9	18.2	58.8	12.0

[a]$n = 16$. [b]$n = 15$.

$t(15) = 8.1, p < .001$, for first-syllable samples in the HFA group; $t(15) = 2.2, p = .045$, for second-syllable samples in the HFA group; $t(14) = 5.5, p < .001$, for first-syllable samples in the TD group; and $t(14) = 2.9, p < .01$, for second-syllable samples in the TD group, based on paired sample t tests. To compare accuracy for each stress condition, we conducted a 2 (Group) × 2 (Stress) ANOVA with percentage correct as the dependent variable and found a main effect of stress, $F(1, 29) = 21.4, p < .001$, partial $\eta^2 = .43$ (data were normally distributed). Pairwise comparisons showed that, as predicted, both participant groups were more accurate on the first-syllable stress patterns than the last-syllable stress items, $t(15) = 3.6, p = .003$, Cohen's $d = 1.2$, for the HFA group, and $t(14) = 3.1, p = .008$, Cohen's $d = 0.8$, for the TD group, based on paired samples t tests. There was no significant main effect of group, $F(1, 29) = 2.0, p = .17$, partial $\eta^2 = .07$, and no Group × Stress interaction, $F(1, 29) = 2.16, p = .15$, partial $\eta^2 = .07$.

Discussion

There were no significant differences in performance between the two participant groups, indicating that the children and adolescents with HFA had the same competence level for this task as their TD peers. The lack of group differences in this task supports our hypothesis that participants with HFA would be able to use their relatively preserved language skills, as evidenced by normal-level verbal IQ and receptive vocabulary scores, to tap into the grammatical nature of these stimuli and successfully use lexical prosody to disambiguate compound nouns (e.g., *greenhouse*) from noun phrases (e.g., *green house*). The results also show that both participant groups were significantly better at categorizing first-syllable stressed items, which is the predominant stress pattern for nouns in American English, than at categorizing the rarely used final-syllable stress pattern (Arciuli & Cupples, 2006; Jusczyk et al., 1993).

An alternative explanation for the equal performance levels across the HFA and TD participant groups is that the TD group's accuracy levels were surprisingly low. This binary forced-choice task of basic lexical stress

patterns should not have presented a great obstacle for TD children and adolescents, especially because the mean age for both groups (12;7 for the TD group and 12;4 for the HFA group) was above the age of lexical prosody perception competence established by Cruttenden (1974) and Atkinson-King (1973). However, the age range of our participants was relatively large (7;6–17;0 in the HFA group, 7;6–18;0 in the TD group), including children who were well below the age of competence for this type of task. The relatively small size of our groups and specific age distributions did not allow for statistical comparison of age-based subgroups, but it is possible that the younger members of the TD group accounted for the surprisingly low accuracy levels in this task. An older group of TD individuals (12;8–32;0) who participated in Plesa Skwerer et al.'s (2007) study, in which the same stimuli were used, achieved accuracy scores of 95% on first-syllable stress items and 80% on second-syllable stress items, indicating that age might have been the biggest contributing factor to the poor accuracy results of our TD group. Further investigations using this type of methodology are required to definitively ascertain whether the low accuracy scores among the TD participants were a function of age distribution.

Our results agree with prior data showing that individuals with ASD have preserved competence recalling stressed words (Fine et al., 1991) and are able to disassociate nouns from homophone verbs, which are distinguished only through first- or second-syllable lexical stress (e.g., *REcall* vs. *reCALL*; Paul, Augustyn, et al., 2005). Overall, both groups were equally able to distinguish the two types of lexical stress and, as expected, both groups were more accurate at recognizing first-syllable stress patterns, which is the prevalent and primary stress pattern of American English.

Experiment 3: Production of Lexical Prosody

The task we used in Experiment 3 was designed to expand on past research of prosody production in individuals with ASD by using objective acoustic measures of lexical stress (pitch, intensity, and duration) instead of a subjective rating. By using these measures, we hoped to discern whether the frequently perceived inappropriate prosody of individuals with autism (Baltaxe, 1984; Baltaxe & Guthrie, 1987; Fine et al., 1991; Paul, Shriberg, et al., 2005; Shriberg et al., 2001) could be captured in a more objective manner for elicited lexical stress items. Drawing on the same set of stimuli used in Experiment 2, we designed the task to test the ability of children and adolescents with HFA to produce lexical stress to disambiguate compound nouns from homophone noun phrases.

Method

Participants. Participants were the same groups of children and adolescents as in Experiments 1 and 2.

Materials. We created 22 two-sentence narratives to elicit the two possible meanings of 11 ambiguous word pairs. Each contextual framework was composed of a short sentence (5–9 words) to determine the context, followed by a longer sentence (11–14 words) that concluded with a picture illustrating the target word, paired with the target word in writing, for example, "Kate calls Tom on his cell phone. When Tom doesn't answer, Kate wishes he would (pick up)," paired with an illustration of Tom picking up the phone. (The full list of stimuli is presented in Appendix C.) We maintained the same sentence-final position for all stimuli to ensure consistency across target word context. A sentence-final position also ensured that participants obtained all contextual information relevant for defining the target's stress assignment before having to produce it, which made the elicited word/noun phrase more naturally embedded within the sentence. We also created contextual sentences for three foil compound nouns with first-syllable stress (e.g., *DOORmat*) and two noun phrases with second-syllable stress (e.g., *small FISH*). The ambiguous words and foils were a subset of the ones used in Experiment 2, with the addition of two last-syllable stress foils, which were used to elicit this less common version of lexical stress. The pictures used to elicit the target words were either taken with a digital camera or obtained through the Internet and were intended to look realistic and minimize apparent similarities between the tasks. Several versions of each contextual framework were recorded by a native English speaker. Two independent judges listened to the sentences and selected the most natural-sounding example of each to be used in the task. We created two different pseudorandomized and counterbalanced sequences of the sentences and randomized them across groups.

Procedure. Participants heard each of the contextual sentences via standard PC speakers attached to a portable CD player. The pictures were presented to the participants in a notebook that was placed in front of them. Before beginning the task all participants were told that they were going to hear a story about Tom and Kate and their summer vacation and that they were to say the missing word(s), which were illustrated and written in the notebook in front of them. They were then shown a picture of a practice word (*snowman*) and asked to say it into the microphone to ensure that they could be recorded clearly. The participants heard the 27 sentences (11 ambiguous word sentences with first-syllable stress, 11 ambiguous word sentences with last-syllable stress, and 5 foil sentences) with a pause after each sentence to allow them to say the missing word and turn to the page showing the next picture and word.

Participants completed the receptive (Experiment 2) and expressive tasks in two separate testing sessions that were divided by approximately 1.5 hr of other, non-related activities, with Experiment 2 always preceding Experiment 3.

Results

To determine the accuracy of differential lexical stress production for each participant, we conducted analyses of mean pitch and intensity using Praat software to determine whether there were quantifiable group differences in the manner in which stress was assigned and produced. Also using Praat, we measured the whole-word duration of every utterance. We chose this measure because a longer whole-word duration is indicative of a last-syllable stress pattern, whereas a shorter duration is indicative of a first-syllable stress pattern (Smith & Robb, 2005; Y.-T. Wang, Kent, Duffy, & Thomas, 2005), thereby providing us with a good measure of whether participants were able to produce the targeted stress patterns accurately. We established reliability for the coding of whole-word duration by having 15% of the productions rated by a second, independent coder. The whole-word duration determined by the second coder was then subtracted from the original measurement of the child's production, and the mean absolute difference between the two measurements was calculated. Reliability was defined as a mean absolute difference in length of less than 5 ms. The mean absolute intercoder difference in whole-word duration for 15% of the total sample was 4 ms, thereby meeting the criterion for reliability.

The sound quality of recordings for 5 participants with HFA and 6 TD participants contained too much static to reliably analyze for pitch and intensity, so we conducted a one-way ANOVA (in which the data were normally distributed) for mean pitch and intensity on the productions of 11 participants with HFA and 9 TD control participants. There were no significant group differences for intensity on first-syllable stress items, $F(1, 18) = 2.56, p = .13$, or second-syllable stress items, $F(1, 18) = 2.44, p = .14$, and no significant group differences for pitch on first-syllable stress items, $F(1, 18) = 0.54, p = .51$, or second-syllable stress items, $F(1, 18) = 0.27, p = .61$, showing that the mean intensity and pitch levels of the first- and second-syllable stressed productions were comparable across groups. However, we should point out that these results are based on a very small subset of our overall data and that technical difficulties with our digital recordings may have influenced the ability of Praat to extract accurate pitch and intensity values from these productions.

Mean word length results for each group are shown in Table 4. We conducted a 2 (Group) × 2 (Stress) ANOVA

Table 4. Experiment 3: Whole-word duration (in seconds) for production of lexical stress.

Type of stimulus	HFA[a]		TD[b]	
	M	SD	M	SD
First-syllable stress	0.82	0.15	0.68	0.19
Last-syllable stress	0.98	0.19	0.83	0.21

[a]$n = 16$. [b]$n = 15$.

(in which the data were normally distributed) with mean word duration as the dependent variable. The analysis shows a significant main effect of stress, $F(1, 29) = 33.04$, $p < .001$, partial $\eta^2 = .53$, with the first-syllable stress words produced significantly shorter than the last-syllable stress items in both groups, $t(15) = 3.5, p = .003$, Cohen's $d = -0.9$, for the HFA group, and $t(14) = 6.0, p < .001$, Cohen's $d = -0.8$, for the TD group, based on paired samples t tests. There was also a significant group effect, $F(1, 29) = 5.62, p = .024$, partial $\eta^2 = .16$, with participants in the HFA group having longer overall productions than participants in the control group, $F(1, 29) = 5.1, p = .032$, Cohen's $d = 0.82$, for first-syllable stress items, and $F(1, 29) = 4.4, p = .044$, Cohen's $d = 0.75$, for last-syllable stress items, based on a one-way ANOVA. This difference in overall utterance length is apparent despite the fact that the HFA group followed the correct pattern of American English by producing shorter first-syllable stressed words and longer last-syllable stressed items. There was no Stress × Group interaction, $F(1, 29) = 0.072, p = .79$, partial $\eta^2 = .002$.

Discussion

The lexical prosody production task revealed two interesting results, both confirming and not confirming our hypotheses. Contrary to our expectation, children and adolescents with HFA were able to appropriately disambiguate word pairs through differentiated production of first-syllable and last-syllable stress patterns; however, as predicted, all of their productions were measured to be significantly different (i.e., longer) than those of their TD peers, despite the fact that we observed no significant group differences in pitch or intensity allocation for the two different syllable-stress elicitations.

Listening to individual productions by adolescents with HFA, we noticed that participants often produced exaggerated pauses between the syllables, especially for second-syllable stress items. The labored and slow enunciations of the HFA group stood in stark contrast to the recordings of the TD control participants, who produced effortless enunciations that subjectively appeared briefer, less labored, and more fluid in their transitions between

syllables. It is possible that the participants with HFA were able to use their largely preserved grammatical skills as a compensatory strategy in this task, creating the grammatical distinction between these two stimulus types (e.g., the compound noun *PICKup* vs. the noun phrase *pick UP*) through purposeful pausing between the nouns of the two-word phrase instead of through the production of the corresponding prosodic contours. It is important to point out, however, that the HFA group's enunciations were longer than those of the TD group for both first- and second-syllable stress items. This indicates that the abnormal prosodic lengthening we recorded cannot be explained solely by the HFA group's attempt to emphasize the grammatical class of two-word phrases through abnormally long pauses between the nouns but must also reflect a more generalized expressive prosody difference, specifically in the rhythm/timing domain. Overall, these results show quantifiable differences in prosody productions between participants with HFA and their TD peers and confirm that whole-word duration is an appropriate measure for capturing this group difference, because syllable duration would not have been sensitive to a pause between syllables. This quantifiable duration difference may help explain the prosody production deficits found in previous studies (McCann et al., 2007; Paul, Augustyn, et al., 2005; Peppé et al., 2007; Shriberg et al., 2001).

The technically accurate production of lexical stress in the HFA group may be explained by the narrowly defined task presented in this study, which asked for the production of only two specific stress patterns in two-syllable word pairs. A. T. Wang, Lee, Sigman, and Dapretto (2007) showed that overtly asking participants with ASD to attend to specific aspects of prosody enhanced their performance on a neurophysiology level. This indicates that task instructions can have a significant effect on the performance of this group on tasks of prosody. It is possible that our very narrowly defined task design and specific instructions helped participants with HFA produce accurate lexical stress distinctions. It is also possible that the competence level shown by the children and adolescents with HFA in the lexical stress perception task (Experiment 2) informed their productions of lexical stress for the same, or at least similar, stimuli.

It is important to reiterate, however, that despite being able to produce two appropriately differentiated stress patterns, the prosody productions of the children and adolescents with HFA were nevertheless acoustically, and thereby objectively, different from the productions of their TD peers. This duration difference was consistent across first-syllable and last-syllable stressed items and amounted to roughly 100 ms. Although this difference may appear small, it is significant in the context of mean word productions that are only 750 ms long

for first-syllable stressed words and 900 ms for second-syllable stressed items. The HFA group lengthened these productions by 13% and 11%, respectively, constituting a significant change in production and resulting in a significant group difference.

There is evidence to suggest a developmental component to sentence duration in the productions of very young children, with preschoolers producing slower speech that may also be more varied in speed, than the speech of school-age children. However, this variability in articulatory lengthening appears to stabilize and developmentally plateau in children over the age of 6 years (Kent & Forner, 1980; Smith & Zelaznik, 2004), thereby making it unlikely that the production lengthening of the HFA group was caused by this aspect of developmental speech rate.

The interesting finding in Experiment 3 is that the participants with HFA were *accurate* in their productions of lexical stress but still quantifiably *different* from their TD peers. These results are supported by evidence from other analyses of nonverbal language use in ASD. Grossman et al. (2008) showed that adolescents with ASD were able to produce prosodic and facial expression modulations that effectively communicated affect but were still judged to be awkward in their execution. Similarly, in the present study, participants with HFA successfully distinguished between items requiring first-syllable and second-syllable stress, but both objective acoustic analyses and informal subjective listening were able to perceive the overall lengthening and atypical midword pausing of the productions as a distinct group difference. Further studies using more refined acoustic measures, as well as subjective coding of naturalness of production, are required to better distinguish between the ability to produce a distinction between different verbal constructs through the accurate use of lexical stress and the ability to produce overall prosody that is perceived and measured to be within the normal range of expectation for natural conversation.

General Discussion and Conclusion

The aim of this study was to enhance the current knowledge base of prosody competence in high-functioning, linguistically able individuals with ASD by investigating the perception and production of lexical stress and the ability to categorize affective prosody in sentence-length stimuli. Our findings indicate that children and adolescents with HFA were equal to their TD peers in the ability to perceive affective prosody in sentences filtered to eliminate semantic content. However, we found surprisingly poor results in the TD group for this task that may explain the lack of a significant group difference. Both participant groups showed the same pattern of

being more successful at categorizing sentences containing sad or happy affect than neutral sentences, indicating that their approaches to affective prosody interpretation at the sentence level are similar. The relatively simplistic nature of the task (forced choice, three options only) and the selection of two basic emotions (happy and sad) as stimuli cannot be overlooked. The results clearly indicate that, within the confines of such a task, and using only basic emotions, adolescents with HFA are not significantly different in their accuracy and error types from their TD peers, but we cannot say that this ability would necessarily translate to more complex emotions or a more open-ended task. Nevertheless, the filtered-affect task (Experiment 1) did show that individuals with HFA are able to use sentence-level prosodic cues in isolation, without any semantic content to determine basic emotions of a speaker. These data clearly speak to the fact that individuals with HFA have at least basic affective prosody interpretation abilities.

The HFA group also was equal to their TD peers in the ability to perceptually disambiguate compound nouns from noun phrases using only prosodic cues, suggesting a preserved ability to decode grammatical prosody in children and adolescents with HFA, again within the confines of a binary forced-choice task. Our results also suggest that children and adolescents with HFA were able to accurately produce stress patterns to disambiguate compound nouns from noun phrases in the context of a sentence completion task, despite abnormally long word productions. These findings regarding lexical stress suggest that children and adolescents with HFA can use the grammatical and prosodic rules underlying lexical stress to differentiate between compound nouns and noun phrases as well as their TD peers. The only group difference in this set of experiments was found in the significantly longer whole-word duration of lexical stress productions by the HFA group in Experiment 3. It is important to emphasize that their productions accurately differentiated the two stress patterns, showing a basic awareness of lexical stress rules and their expressive applications. The distinction between the two groups is therefore tied not to an inability of the participants with HFA to perform the lexical prosody production task but instead to their unique speech patterns captured by quantifiable differences in acoustic measurements of length and a subjectively noticeable awkwardness. Overall, the participants with HFA demonstrated the same levels of competence for all three tasks in both modalities as their TD peers and were different only in the overall length of their prosodic stress productions.

Although some studies have found impaired perception of prosodic stress in individuals with ASD, Shriberg et al. (2001) pointed out that most of these deficits were reported on the level of full sentences, such as the placement of stress on inappropriate words within a phrase or sentence, instead of the use of lexical stress within single words. This might explain why the participants with HFA in our study were able to correctly differentiate the meanings of two-syllable words and were able to produce the appropriate first- and last-syllable stress patterns. Our data appear to be consistent with other studies that have found deficits in stress *production* in participants with HFA when compared with their TD peers despite being able to correctly *differentiate* stress patterns perceptually (Foreman, 2002; Paul, Augustyn, et al., 2005). McCann et al. (2007) described atypical prosody productions in a group of individuals with ASD that could not be explained by a developmental delay but instead appeared to be symptomatic of a disorder encompassing prosodic skill. This interpretation captures well the prosodic productions we recorded in Experiment 3. The children and adolescents with HFA did not show deficits, or immaturity, in producing differential stress patterns but rather an overall atypical pattern of two-syllable word productions that was significantly different from the utterances of their TD peers.

In all three experiments, participants had to rely on prosodic information to perform the task accurately. The filtered sentence affect task excluded all semantic information, leaving only prosody as the determining factor. The lexical stress ambiguous word pairs in the receptive task were differentiated exclusively through the perception of prosodic emphasis, which determined their grammatical class as a compound noun or noun phrase. Accuracy in the lexical stress production task was achieved by correctly assigning prosodic stress to the first or second syllable of the target, on the basis of the interpretation of contextual cues. It has previously been suggested that individuals with ASD rely overly on verbal content and ignore prosodic cues (Lindner & Rosén, 2006). Our data show that children and adolescents with HFA are able to apply basic knowledge of lexical and affective prosody when it is the main focus of their attention. These results are based on a relatively small sample size, and further work comparing different methodological approaches to prosody perception, such as contrasting elicited versus spontaneous productions, varying the specificity of task instructions, and using more refined objective measures of prosody production, would be helpful in determining the underlying prosodic abilities of individuals with autism.

Acknowledgments

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental of mental disorders* (4th ed.). Washington, DC: Author.

Arciuli, J., & Cupples, L. (2006). The processing of lexical stress during visual word recognition: Typicality effects and orthographic correlates. *Quarterly Journal of Experimental Psychology, 59,* 920–948.

Atkinson-King, K. (1973). *Children's acquisition of phonological stress contrasts* (Unpublished doctoral dissertation). University of California, Los Angeles.

Bachorowski, J.-A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science, 6,* 219–224.

Baltaxe, C. A. M. (1984). Use of contrastive stress in normal, aphasic, and autistic children. *Journal of Speech and Hearing Research, 27,* 97–105.

Baltaxe, C. A. M., & Guthrie, D. (1987). The use of primary sentence stress by normal, aphasic and autistic children. *Journal of Autism and Developmental Disorders, 17,* 255–271.

Baltaxe, C. A. M., & Simmons, J. (1985). Prosodic development in normal and autistic children. In E. Schopler & G. Mesibov (Eds.), *Communication problems in autism* (pp. 95–125). New York: Plenum Press.

Baltaxe, C., & Simmons, J. Q. (1992). A comparison of language issues in high-functioning autism and related disorders with onset in childhood and adolescence. In E. Schopler & G. Mesibov (Eds.), *High-functioning individuals with autism* (pp. 201–225). New York, NY: Plenum.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 170,* 614–636.

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer (Version 4.3.14) [Computer program]. Retrieved from http://www.praat.org.

Boucher, J., Lewis, V., & Collis, G. M. (2000). Voice processing abilities in children with autism, children with specific language impairments, and young typically developing children. *Journal of Child Psychology and Psychiatry, 41,* 847–857.

Cosmides, L. (1983). Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 864–881.

Cruttenden, A. (1974). An experiment involving comprehension in children from 7–10. *Journal of Child Language, 1,* 221–231.

Cutler, A., & Swinney, D. (1987). Prosody and the development of comprehension. *Journal of Child Language, 14,* 145–167.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised*. Circle Pines, MN: AGS.

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

Fine, J., Bartolucci, G., Ginsberg, G., & Szatmari, P. (1991). The use of intonation to communicate in pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, 32,* 771–782.

Foreman, C. (2002). The use of contrastive focus by high-functioning children with autism. *Dissertation Abstracts International, 62,* 3759.

Fosnot, S. M., & Jun, S. (1999). Prosodic characteristics in children with stuttering or autism during reading and imitation. In J. J. Ohala & Y. Hasegawa (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 103–115). Berkeley, CA: University of California Press.

Gerken, L., & McGregor, K. K. (1998). An overview of prosody and its role in normal and disordered child language. *American Journal of Speech-Language Pathology, 7,* 38–48.

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2007). The Reading the Mind in the Voice Test—Revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders, 37,* 1096–1106.

Grossman, R. B., Edelson, L., Rubinstein, L. B., Lomibao, J., Borawski, S., & Tager-Flusberg, H. (2008, May). *Production of emotional prosody and facial expressions in adolescents with autism*. Poster presented at the International Meeting for Autism Research, London, England.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64,* 675–687.

Kanner, L. (1968). Autistic disturbances of affective contact. *Acta Paedopsychiatrica, 35,* 100–136. (Original work published 1943)

Kaufman, A., & Kaufman, N. (2004). *Manual for the Kaufman Brief Intelligence Test, Second Edition*. Circle Pines, MN: AGS.

Kent, R. D., & Forner, L. L. (1980). Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics, 8,* 157–168.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America, 59,* 1208–1221.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America, 32,* 451–454.

Lindner, J. L., & Rosén, L. A. (2006). Decoding of emotion through facial expression, prosody and verbal content in children and adolescents with Asperger's syndrome. *Journal of Autism and Developmental Disorders, 36,* 769–777.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The Autism Diagnostic Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30,* 205–223.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule—WPS (ADOS–WPS)*. Los Angeles, CA: Western Psychological Services.

Lord, C., Rutter, M., & LeCouteur, A. (1994). Autism Diagnostic Interview—Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24,* 659–685.

McCaleb, P., & Prizant, B. (1985). Encoding of new versus old information by autistic children. *Journal of Speech and Hearing Disorders, 50,* 226–230.

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language and Communication Disorders, 38,* 325–350.

McCann, J., Peppé, S., Gibbon, F. E., O'Hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *International Journal of Language and Communication Disorders, 42,* 682–702.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29,* 143–178.

Merewether, F. C., & Alpern, M. (1990). The components and neuroanatomic bases of prosody. *Journal of Communication Disorders, 23,* 325–336.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America, 93,* 1097–1108.

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 35,* 205–220.

Paul, R., Shriberg, L. D., McSweeny, J., Cicchetti, D., Klin, A., & Volkmar, F. R. (2005). Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 35,* 861–869.

Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS–C test and the revised version. *Clinical Linguistics & Phonetics, 17,* 345–354.

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2006). Assessing prosodic and pragmatic ability in children with high-functioning autism. *Journal of Pragmatics, 38,* 1776–1792.

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

Plesa Skwerer, D., Schofield, C., Verbalis, A., Faja, S., & Tager-Flusberg, H. (2007). Receptive prosody in adolescents and adults with Williams Syndrome. *Language and Cognitive Processes, 22,* 247–271.

Rapin, I., & Dunn, M. (1997). Language disorders in children with autism. *Seminars in Pediatric Neurology, 4,* 86–92.

Shriberg, L., & Kent, R. (2003). *Clinical phonetics* (3rd ed.). Boston, MA: Allyn & Bacon.

Shriberg, L. D., Kwiatkowski, J., & Rasmussen, C. (1990). *The Prosody-Voice Screening Profile*. Tucson, AZ: Communication Skill Builders.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research, 44,* 1097–1115.

Smith, A., & Robb, M. (2005). Durational characteristics of newly-learned trochees and iambs in children with and without speech delays. *Clinical Linguistics & Phonetics, 19,* 1–14.

Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology, 45,* 22–33.

Sparrow, S., Balla, D., & Cicchetti, D. (1984). *The Vineland Adaptive Behavior Scales (Survey Form)*. Circle Pines, MN: AGS.

Vogel, I., & Raimy, E. (2002). The acquisition of compound vs. phrasal stress: The role of prosodic constituents. *Journal of Child Language, 29,* 225–250.

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2007). Reading affect in the face and voice. *Archives of General Psychiatry, 46,* 698–708.

Wang, Y.-T., Kent, R. D., Duffy, J. R., & Thomas, J. E. (2005). Dysarthria in traumatic brain injury: A breath group and intonational analysis. *Folio Phoniatrica et Logopaedica, 57,* 59–89.

Appendix A. List of stimulus sentences for the affective discrimination task.

Happy
- When Mike pets the puppy, it's wagging its tail.
- When the kids invite Mike, he's excited.
- Whenever Sue calls Mike, he's glad.
- When Sue hugs the cat, it's purring.
- When Sue babysits the kids, they're cheering.
- Whenever the teacher praises Sue, she's thrilled.

Sad
- When Mike hits the puppy, it's whining.
- When Mike sees his mother, she's crying.
- When nobody visits Sue, she's lonely.
- When Sue leaves her cat, it's howling.
- When the kids tease Sue, she's upset.
- If Mike doesn't call Sue, she's unhappy.

Neutral
- When Mike pets the puppy, it's sleeping.
- When Mike calls his mother, she's in the kitchen.
- When Mike buys the candy, it's chocolate.
- Whenever Sue draws a cat, it's grey.
- When Sue leaves the class, it's noisy.
- When Sue closes the store, it's dark.

Appendix B. Word list for the lexical ambiguity reception task.

Ambiguous stimuli	Foils
Blackboard	Bluebird
Bulls-eye	Bookcase
Greenhouse	Doormat
Highlight	Driveway
Holdup	Headphones
Hotdog	Mailbox
Makeup	Soft drink
Pickup	Top shelf
Takeoff	Tree house
Top hat	T-shirt
Wetsuit	Woodpile

Appendix C. List of sentences to elicit utterances for lexical ambiguity production task. Words in parentheses are the correct utterance to be produced.

Tom helps David build a birdhouse. David tells him: "I need three nails and a (black board)."
Tom and Kate visit an old schoolhouse. In the school there are desks and a (blackboard).
Tom and Kate go to a farm. One cow comes so close to them, they can see right into the (bull's eye).
Kate is learning to use a bow and arrow. When she takes lessons, the teacher tells her to aim for the (bulls-eye).
Tom has sand on his shoes. When he comes home Kate tells him to wipe his feet on the (doormat).
Tom likes learning about plants, so when Kate goes out shopping he decides to visit a (greenhouse).
Tom and Kate will go visit their friend David. David lives on top of a hill in a (green house).
Kate wants to mark her guidebook. Tom gives her a marker and tells her: "Use it to (highlight)."
Kate is walking down Main Street. She looks up at the tall streetlight and says: "That is a (high light)."
David wants to meet Kate on a beach. So that Kate will find him he writes her name on a sign to (hold up).
Tom and Kate go to a movie. In the movie a man says: "Stop where you are. This is a (holdup)."
Kate takes her dog for a walk, so when she comes back he is a (hot dog).
Tom and Kate go to a restaurant. Kate is really hungry, so she orders a hamburger and a (hotdog).
Kate has written some postcards for her friends. She puts a stamp on each card and takes them to a (mailbox).
Kate is going to a movie. Before she leaves she puts on her ring and her (makeup).
Tom is angry that Kate broke his souvenir. Later Kate tells Tom that she's sorry and they (make up).
Kate calls Tom on his cell phone. When Tom doesn't answer, Kate wishes he would (pick up).
Tom and Kate rent a truck. While they are driving, Tom says: "Wow, I like this (pickup)."
Kate goes to the aquarium. She says to Tom: "Look at this (small fish)."
Tom wants to send a gift to his mother. He wraps the gift in paper and puts it in a (square box).
Tom and Kate board the airplane. The pilot says: "Fasten your seatbelts, it's time for (takeoff)."
Because it is hot outside, Tom is glad he wore a jacket that he can (take off).
Kate wants to buy something for Tom. She sees a pile of things on a high shelf and tells the salesperson she wants the (top hat).
Tom and Kate go to the opera. Tom decides that he will wear a suit and a (tophat).
Tom buys an ice cream cone. On the way home he drops his ice cream on his (t-shirt).
Tom decides he wants to go diving. So he puts on his diving mask and his (wetsuit).
Tom didn't bring his umbrella. When it starts raining, he comes home with a (wet suit).