4-18-2011

Unusual Prosodic Descriptors in Young, Verbal Children with Autism Spectrum Disorders

Rachel Diane Fields

UNUSUAL PROSODIC DESCRIPTORS IN YOUNG, VERBAL CHILDREN WITH AUTISM SPECTRUM DISORDERS

by

Rachel Diane Fields

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

Major: Speech-Language Pathology

The University of Memphis

Acknowledgments

I would like to extend my appreciation to the members of my committee: Dr. Julie E. Cleary, Dr. Eugene H. Buder, and Dr. D. Kimbrough Oller. I feel honored to have had the privilege to work with each of them during my two years at the University of Memphis. I am thankful for Dr. Cleary’s efforts as my thesis advisor, her patience with me as well as her guidance and encouragement throughout the process. I am also grateful to Dr. Buder, who spent hours teaching and re-teaching pitch analysis techniques and who was always available for questions. I would also like to extend a heartfelt thank you to my family and friends who were supportive of my research endeavors.

Abstract

Fields, Rachel Diane. M.A. The University of Memphis. May/2011. Unusual Prosodic Descriptors in Young, Verbal Children with Autism Spectrum Disorders. Major Professor: Dr. Julie E. Cleary.

This study aimed to determine which prosodic descriptors best characterized the speech of children with autism spectrum disorders (ASD) and whether these descriptors (e.g., sing-song and monotone) are acoustically different. Two listeners' auditory perceptions of the speech of the children with ASD and the pitch of the speech samples were analyzed. The results suggest that individual children are characterized by a variety of prosodic descriptors. Some thought groups were described as both sing-song and monotone, however, most children appear to be either more monotone or more sing-song. Furthermore, the subjective and acoustic data suggest a strong relationship between atypical intonation and sing-song perceptions as well as atypical rhythm and monotone perceptions. Implications for an earlier diagnosis of ASD and for the development of therapy tasks to target these deficits are discussed.

Table of Contents

List of Tables

Chapter 1

Introduction

Autism Spectrum Disorders (ASD) are developmental disorders that present with qualitative impairment in social interaction (e.g., failure to develop appropriate peer relationships), communication (e.g., delay of language), and restricted repetitive and stereotyped patterns of behavior, interests, and activities (e.g., hand flapping; APA, 2000). Unfortunately, these diagnostic criteria cannot be easily applied to preschool children with ASDs as social interaction and communication delays are difficult to parse at such a young age. The proposed revision of the DSM-V will place deficits in social interaction and communication into one criterion, which will include deficits in both nonverbal and verbal communication (APA, 2010). Both, understanding earlier developing symptoms of ASD and adding diagnostic criteria that are more applicable to younger children, will make an earlier diagnosis more feasible.

The Center for Disease Control and Prevention (CDC) estimated that 1 in 110 children in the United States have an Autism Spectrum Disorder (CDC, 2010). Early identification of ASD will be possible only when early developing characteristics of ASD are more fully understood. This report from the CDC emphasized the importance of identifying specific characteristics, within social interaction, communication, and behavior, which differentiate young children with ASD from young, typically developing children. Researchers have conducted a substantial number of studies regarding nonverbal social communication in young children with ASD (Stone, Ousley, Yoder, Hogan & Hepburn, 1997; Wetherby & Prutting, 1984); yet, there is less understanding of verbal communication in these children. Furthermore, studies examining verbal aspects of communication tend to assess the quantity of verbalizations, function of verbalizations, and phonemic or vocal quality of verbalizations (Sheinkopf, Mundy, Oller, & Steffens, 2000; Wetherby & Prutting, 1984; Wetherby, Yonclas & Bryan, 1989). In particular, few studies have investigated if and how the prosodic characteristics of young children with ASD differ from prosodic characteristics of children who are typically developing. Studies that have investigated prosodic differences in speech of children with ASD tend to focus on older children or adolescents (Shriberg, Paul, McSweeny, Klin, Cohen, & Volkmar, 2001), use ill-defined subjective judgments to measure the degree of difference (Hubbard & Trauner, 2007), and often use unnatural elicitation techniques such as imitation or prompting (Peppe, McCann, Gibbon, O'Hare, & Rutherford, 2007).

Chapter 2

Literature Review

Early Characteristics in ASD

Several researchers have investigated early characteristics of ASD in efforts to make an early diagnosis more feasible. The evaluation of communication in toddlers at risk for ASD typically assesses: rate of communication, use of gaze and gestures, responsiveness to speech and gestures, communicative functions expressed, play schemes, and quality of vocalizations (Paul, 2005). Wetherby et al. (2004) compared children with ASD, developmental delay (DD) and typical development (TD) to establish predictors of ASD during the second year of life. These investigators selected from a large pool of children (see Wetherby et al., 2004 for more information about selection process). Behavior samples and parent checklists were administered while the children were less than 2 years of age. The three groups of 18 children, between the ages of 30 months and 5 years of age, were contacted for a follow-up. A *best estimate diagnosis* was made from the results of the Mullen Scales of Early Learning (MSEL; Mullen, 1995), Vineland Adaptive Behavior Scales (VABS; Sparrow, Balla, & Cicchetti, 1984) and Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000). Additionally, the Social Communication Questionnaire (SCQ; Rutter, Bailey, Berument, Lord, & Pickles, 2001) was mailed to the families of the children who were TD. The Communication and Symbolic Behavior Scales Developmental Profile (CSBS DP; Wetherby & Prizant, 2002) was also completed. It is an in-depth tool that is organized into a social composite, a communication composite, and a symbolic composite. The assessment consists of a 24-item checklist completed by the parent or caregiver regarding acquisition of developmental milestones, a follow-up caregiver questionnaire, and a behavior sample of the child with the parent and clinician. The authors found significant differences between the ASD group and the DD group on nine items, including: lack of appropriate gaze; lack of warm expression with gaze; lack of sharing enjoyment or interest; lack of response to name; lack of coordination of gaze, facial expression, gesture, and sound; lack of showing; unusual prosody; repetitive movements or posturing of his or her body; and repetitive movements with objects. Wetherby and colleagues (2004) also found significant differences between only the ASD group and the TD group on four other items, including: lack of response to contextual cues, lack of pointing, lack of vocalization with consonants, and lack of playing with a variety of toys consistently. Because the majority of these red flags include a communicative characteristic that children with ASD are missing, further study is needed to determine what is present in the communication of children with ASD. The red flag of interest for the current study – unusual prosody – has seldom been investigated, but might provide characteristics of speech that are present rather than absent in this population.

Wetherby, Watt, Morgan, and Shumway (2007) conducted a similar study with 123 participants: 50 with ASD, 23 with DD, and 50 who were TD. Over half of the children were considered to have high-functioning autism (HFA). The authors used the CSBS DP 24-item check list, recorded a behavior sample, and completed the MSEL. This study included 54 participants in the previous study by Wetherby and colleagues (2004) and used the same procedures to determine a diagnosis. The experimenters compared social communication measured late in the second year of life across groups and investigated which of these social communication characteristics was predictive of the child's developmental level at 3 years of age. The authors found that *understanding*, which was measured by how many single-words the children understood without gestural cues, had a largely significant correlation with both the nonverbal and verbal developmental quotient. Therefore, the authors controlled for the age of understanding and found that the following were predictive of nonverbal DQ: rate of communicating, acts for behavior regulation, inventory of gestures, inventory of play actions, and stacking blocks. The verbal DQ could be predicted from: acts for behavior regulation and inventory of consonants.

Shumway and Wetherby (2009) conducted a follow-up study to the study by Wetherby and colleagues (2007). They examined 125 children (123 from the previous study and 2 more) to understand rate, function, and means of communication in children with ASD. Gestures used to establish reference or call attention to an object (i.e. deictic gestures) were the most delayed in children with ASD. The authors also found a significantly lower rate of communicative acts in these children compared to children with DD or children who were TD. This appears to reflect a lack of social motivation, lack of desire to initiate and lack of response to other's initiations. The children with ASD also engaged more often in behavior regulation as a communicative function than in social interaction as a communicative function, and they engaged in joint attention least often as a communicative function. These data suggest that young children with ASD are requesting and protesting, but may not be purposefully seeking out others for interaction.

Other studies have examined some of the earliest purposeful communication in infants such as the quality of their crying and preverbal vocalizations (Esposito & Venuti, 2009). Espositio and Venuti (2009) conducted a retrospective study which investigated the crying behavior in infants with ASD. Crying is understood as one of the earliest ways that a child expresses his or her needs and is certainly an appropriate measure of early communication. In this study, the Cry Observation Codes (COC) assessment was used to code infants with ASD, TD infants, and infants with DD. This coding system includes the following three categories of coding: infant acoustic production, infant movement production, and mother behavior during an episode of crying. Within the first category, infants with ASD were found to have significantly longer screams, less proportional duration of pause time and less rhythmic cry durations than both typically developing children and children with developmental delays. Within the second category, children with ASD were found to have a significantly greater amount of stereotypy or repetitive movements when compared with the other groups. Within the final category, mothers of children with ASD were more likely to verbally soothe their infant, while mothers of children in the other groups were more likely to use tactile or vestibular stimulation to soothe their infants. Espositio and Venuti (2009) suggested that because children with ASD have a less typical cry pattern and communication between mother and infant is bidirectional, it is possible that mothers of children with ASD have difficulty discriminating their children's cries. Therefore, the authors suggested that these mothers are likely unsure as to what form of soothing their infant needs. This study presents early evidence that the sounds of communication in ASD may be perceived differently than those in children who are developing typically.

Sheinkopf and colleagues (2000) studied preverbal vocal development of young children with ASD to identify positive symptom markers for ASD. Sheinkopf and colleagues (2000) were interested in describing vocal qualities specific to preverbal infants with ASD. The study included one group of 11 boys with ASD and another group of 13 boys with DD. All of the participants were preverbal or produced fewer than 5 words. These authors used the Early Social and Communication Scale (ESCS; Mundy et al. 2003) to code initiations of and responses to joint attention, behavior regulation, and social interaction acts. Their focus was on the vocal variables of these two groups. Sheinkopf and colleagues (2000) assessed the infant's ability to produce complex canonical syllables, which did not differ between the groups. The authors further assessed the infant's vocal quality during phonation, which did differ between groups. These abnormal vocalizations were coded as squeals, growls, and yells. The authors found that 9 of the 15 children with ASD produced greater than 20% of their syllables with atypical vocal quality, whereas only 2 of the 11 children with developmental delays produced greater than 20% of their syllables with atypical vocal quality.

Similar to other studies (Chiang, Soong, Lin, & Rogers, 2008; Loveland, Landry, Hughes, Hall, & McEvoy, 1988; Shumway & Wetherby, 2009), Sheinkopf and colleagues (2000) found that children with ASD initiated fewer joint attention acts and fewer social interaction acts compared to children with DD. Correlation analyses in this study concluded that the lack of joint attention and atypical vocal quality were independent of one another. This study suggests that the quality of vocalizations in preverbal infants is unrelated to pragmatic communicative acts (e.g., joint attention), which further suggests that atypical prosody may be present in verbal children with ASD regardless of the communicative function. This is strikingly different from children who are developing typically according to a study by Furrow (1984) who found differences in the quality of verbalizations in infants according to communicative function. These atypical vocalizations are supported by other research that used similar subjective measures of analysis (Asperger, 1952/1991; Kanner, 1943), but few studies have used acoustic analysis to support these findings. Although the study by Wetherby and colleagues (2004) found that six of the nine red flags that differentiated children with ASD from TD children were negative symptoms (e.g., the lack of a certain skill), the aforementioned studies represent attempts to identify positive symptoms of ASD (e.g., the presence of a specific characteristic) in young, preverbal children. Sheinkopf et al. (2000) provides information regarding early vocalizations of young, preverbal children with ASD; however, much of the research on this population has focused on nonverbal communication.

Nonverbal communication in autism spectrum disorders. Wetherby and Prutting (1984) studied four children with ASD (age 6:11 – 11:10 years) and four typically developing children (ages 1:0 – 2:2 years). The children with ASD were in the prelinguistic and early linguistic stages of language development, and the children who were typically developing were matched on their stage of language development. The authors examined videotaped samples of communicative and play behavior in an environment that was familiar to the child. They also examined a structured communication condition during which each child was engaged in a series of eight situations (e.g., the tester would eat an item of food that the child likes in front of the child without offering him or her any of the food). The authors assessed cognitive-social abilities, language comprehension abilities, and communicative behavior. The authors found that the TD children used significantly more vocalizations than gestures; although, children with ASD used significantly less vocalizations than gestures. Children with ASD protested, requested an action, and requested an object significantly more often than typically developing children. However, children with ASD did not request information, label, comment, show - off, perform, self-regulate, or acknowledge others as often as children who are TD engaged in these behaviors. Compared to children who are TD, the children with ASD engaged in a greater percentage of interactions that resulted in an environmental response, rather than interactions that resulted in a social response. These findings are supported by the results of other studies (Chiang et al., 2008; Shumway & Wetherby, 2009). This study suggests that children with ASD engage in requesting more often than other communicative functions; yet, requesting may serve a more narrow function in children with ASD than in children developing typically. Although this study demonstrated a difference in the frequency of vocalizations in children with ASD and children who are TD, the research described above focused on function of communicative acts rather than means of communication.

Nonverbal communication in children with ASD has significant implications for early diagnosis and early intervention. Requests and protests occur significantly more often than comments in these children which may reflect a greater desire for environmental interaction and less of a desire for social interaction. However, fewer studies have investigated the verbal communicative acts in children with ASD and whether the vocal qualities of these communicative acts differ from children who are TD. Wetherby and Prutting (1984) investigated gestures and vocalizations, but did not draw conclusions about the quality of these vocalizations. Due to the high number of nonverbal children with ASD, research regarding verbal means of communication in ASD is often focused on the presence or absence of that verbalization, rather than the nature of the verbal communication.

Verbal Communication in Typical Development

It is first necessary to understand how typical children are using vocalizations and gestures to convey different communicative functions, before researchers can draw conclusions regarding how children with ASD convey these communicative functions. Wetherby, Cain, Yonclas, and Walker (1988) studied 15 typical infants who were 11-14 months at the time of the initial testing. Multiple evaluations were conducted with these children within their second year of life. The authors investigated the infants' intentional communication by assessing the number of gestural communicative acts and the number of vocal communicative acts they produced. They collected structured and unstructured communicative samples using a similar method to that in the Wetherby and Prutting (1984) study, and they collected these over a 12 month period during the prelinguistic stage, one-word stage, and multiword stage. The authors found that 14 of the 15 typical children were using behavior regulation, joint attention, and social interaction communicative functions. The proportion of intelligible words used to communicate increased from 0% to 53.6% from the prelinguistic stage to the one-word stage and further increased to 94.0% in the multiword stage. The authors noted that the proportion of consonants in their vocalizations increase from the prelinguistic stage to the one-word stage to the multiword stage. These typically developing children were using words more often than gestures to communicate at the end of the 12-month assessment. This study indicates that children should be using words more often than gestures to communicate at one-year, however, the research on children with ASD at 2- and 3-years of age has focused almost entirely on nonverbal means of communication which makes it difficult to compare their verbal development to their typical peers. It is true that many children with ASD are using nonverbal means of communication at 2- and 3-years of age; yet, for high functioning children with ASD who are verbal, very little is known about their use of verbal language.

Verbal Communication in Autism Spectrum Disorders

Similar to the Wetherby and colleagues (1988) study mentioned above, Wetherby, Yonclas, and Bryan (1989) conducted a study to examine the communication profiles of 11 preschool children with Down syndrome, specific language impairment, or ASD who were functioning in the prelinguistic and one-word stage. The authors collected a 30-min sample of each participant's communicative behavior as described by Wetherby and Prutting (1984) and Wetherby and colleagues (1988). Each of the preschoolers with ASD was judged to be in the prelinguistic stage when compared to the study of typically developing infants by Wetherby and colleagues (1988). Each of the participants with ASD showed fewer joint attention acts, fewer vocalizations, and a smaller proportion of vocal acts including consonants. Vocalizations that were not intended for interaction were not included in the analysis because the authors were interested in the quality of intentional communication rather than the quality of all vocal output. This study provides information about the number of vocalizations, function of vocalizations and phonemic quality of vocalizations in children with ASD regarding the absence of certain sounds (i.e. consonants); however, it does not address the suprasegmental quality (e.g. prosody) of the speech in these children.

One of the earliest descriptions of the vocal qualities of children with ASD was provided by Kanner in 1943. He described 11 children with “autistic disturbances”; eight of these children were verbal. Kanner described one child as having clear enunciation, engaging in spontaneous squealing, and being overly literal in his interpretation of other’s speech. Kanner reported that another child in the study produced “short, staccato, forceful sounds” (p. 226). A different child was described as having an unmodulated, hoarse voice and speaking her words in an “abrupt manner” (p. 241). One of the children began to speak at 5-years of age and used simple, mechanical sentences. This child could speak about almost any topic, but did so with an “odd intonation” (p. 241). Kanner described these children as having inflexible grammar or using phrases only in the manner in which the phrases were heard. He wrote that the children he observed had demonstrated an inability to relate themselves to other people and to other situations. Kanner found no difference between the communicative functions used between the verbal children and nonverbal children.

More recent studies (e.g., see above Wetherby, Yonclas, & Bryan, 1989) have compared not only the communicative functions of verbal acts in ASD with TD children, but also the number of verbal acts in ASD versus TD children. Loveland and colleagues (1988) studied verbal and gestural language use in children with ASD through interactions with parent and child. The authors studied 12 children with ASD, 12 children with DD, and 13 children who were TD and all matched at a 3- to 5-year mental age. Loveland and colleagues (1988) coded only utterances that were interactive in nature (i.e. utterances or gestures that involved manipulating objects were not coded). In this study, children with ASD used less initiating communicative acts and more responsive communicative acts. These results are supported by Wetherby and Prutting (1984) and Shumway and Wetherby (2009). There was not a significant difference in the number of verbalizations used between the two groups. This suggests that the function of verbal communicative acts in children with ASD is similar to the function of nonverbal communicative acts in children with ASD. It is important to note that Koegel, Koegel, Green-Hopkins, and Barnes (2010) found that children with ASD who lack question-asking could be taught the appropriate use of the question “where is it?” if child-preferred items were used as motivation in a natural setting. Interestingly, Loveland and colleagues (1988) found that the number of verbal communicative acts in children with ASD was not significantly different than the number of verbal communicative acts in TD children. This is different from the Wetherby et al. (1989) study which found that preschool children with ASD used fewer vocalizations than their TD peers. These authors did not investigate the differences in the quality of verbalizations in children with ASD, children with DD, and TD children; in other words, they did not explore the speech of children with ASD to determine what creates this “odd intonation” described by Kanner (1943).

Volden and Lord (1991) found that speakers with ASD tend to increase in their use of unnatural sounding language as the amount of speech they use increases. Their study included 80 children with ASD, and they found that as the children’s language became more complex, they tended to use words and phrases in more unusual ways. Schoen, Paul, Berkovits and Volkmar (2010) studied children with HFA and TD children to compare production and perception of prosody in these groups. These authors concluded that children with HFA who have age appropriate language skills continue to struggle with production and perception of prosody. In other words, prosody often continues to be a deficit in children with ASD, even if language improves. McCann, Peppe and colleagues (2007) suggest that this is due to parents' natural tendencies to correct grammar, lexical choice, and pronunciation. For example, parents tend to automatically correct a child's statement, "We goed to the store", by responding with, "Yes, we *went* to the store." This grammatical correction is natural for parents. However, parents do not typically correct a child's stress, rate or pitch during speech (McCann et al., 2007). It is logical that sentence structure and vocabulary might improve more quickly than prosodic abilities.

Prosody Overview

The *segmental* aspects of speech include the phonemes, or speech sounds, of a language. Prosody is generally described as the *suprasegmental* aspects of speech production. The suprasegmental aspects of speech exist above the level of the speech sound; they connect the segmental aspects of speech (Shriberg & Kent, 2003). Prosody often has been described as *the music of speech.* Prosody is modulated by changing fundamental frequency, intensity, and duration of the acoustic signal of speech. When these aspects of the speech signal are changed, they will modulate and enhance the meaning of the acoustic signal (Paul, Augustyn, Klin, & Volkmar, 2005; Shriberg & Kent, 2003). The psychological correlations to these signal modifications are pitch, loudness, and length, respectively. However, it is difficult to ascertain that each acoustic change is always matched with the same psychological correlate. For example, increased stress is correlated with longer duration, greater intensity, and an increase in fundamental frequency (Shriberg & Kent, 2003). In fact, Hubbard and Trauner (2007) found that subjective perceptions of pitch variations are not always significantly correlated with variations in fundamental frequency.

Few studies support the use of a particular instrument over another to measure prosody functions (Paul et al., 2005). Most researchers have adhered to subjective interpretation of pitch, loudness and length, rather than extracting measurements of fundamental frequency, amplitude, and duration. More research is needed to develop highly valid methods of measurement and to assure that the methods currently being used are reliable.

Prosodic functions. It is generally agreed that elements of prosody can be modulated to affect three prosodic functions: grammatical, pragmatic and affective. It is notable that these classifications vary according to author; For example, Paul, Augustyn, and colleagues (2005) suggest that pragmatic and affective categories are less exclusive from one another than from grammatical prosody. Furthermore, these authors also describe prosodic elements (e.g., stress, intonation, and phrasing), which they define as the parts of prosody that are adjusted to change the meaning according to one of the functions (grammatical, pragmatic or affective).

Prosody is typically important in dissecting the underlying meaning of a message, but it is not always be necessary when interpreting speech. Several studies have analyzed receptive prosody in ASD to better understand their processing of non-literal aspects of language (Peppe et al., 2007). The contextual cues surrounding an utterance are often enough information to dissect the intended meaning. For example, the sentence “I gave you a present” makes it easy to discern whether *present* is a noun or a verb even when prosody is ambiguous. This is more often the case in the grammatical function of prosody which is described below.

Grammatical prosody includes how words are grouped into appropriate prosodic phrases (Peppe et al. 2007). As explained by Shriberg and colleagues (2001), grammatical prosody often refers to how stress delineates a single word as a noun (PREsent) or a verb (presENT), or how modulating pitch can signify a question (rising) or statement (falling).

Pragmatic prosody serves a social function. Stress can be used to place emphasis on one aspect of the utterance that is especially significant. For example, one item on the Profiling Elements of Prosodic Systems in Children (PEPS-C; Peppe & McCann, 2003) assessment instructs the speaker to say, "I wanted BLUE and black socks" and the child is asked to judge which color socks the speaker must be missing. Because the speaker emphasized the word BLUE, he or she has drawn the listener's attention to that word and the speaker has conveyed inexplicit social information. Perhaps the speaker asked for blue and black socks and the listener brought green and black socks, or perhaps the speaker was disappointed with the color of socks she had received as a gift. Perhaps the speaker realized she had purchased the wrong socks after arriving at home (Peppe et al., 2007). This pragmatic use of stress is also referred to as *contrastive* or *emphatic stress* (Paul, Augustyn, et al. 2005). Pragmatic prosody is essential when conveying the true meaning of figurative language (e.g., sarcasm) as the segmental aspect is usually insufficient (McCann et al., 2007).

Affective prosody is used to convey emotion in different situations; it serves more global functions. It typically includes changes in register according to the speaker's social environment. A higher pitch typically conveys a more positive affect and a lower pitch typically conveys a more negative affect. Paul, Augustyn, and colleagues (2005) provide an example of the phrase “My husband’s plane hasn’t landed yet”. In this sentence, the wife may feel nervous that the plane should have landed an hour before, or she may feel relieved that she does not have to leave for the airport yet. Affective prosody would differentiate which of these truly represent this wife’s current feeling.

As mentioned above, explanations and definitions of prosody vary according to author. Shriberg and Kent (2003) described intonation as a subarea of prosody that includes rhythm, stress, tone and pitch; whereas, Peppe and colleagues (2007) use intonation and prosody synonymously. The following review of intonation will adopt the definition of intonation from Snow and Balog (2002) who define intonation as the pattern of pitch changes within an utterance.

Intonation Parameters. Before continuous speech is analyzed, it must be divided into units so that it can be analyzed in smaller chunks; these divisions are called *intonation-groups* (Snow & Balog, 2002). Intonation-groups may be defined by pause time, respiration, or other criteria. *Nuclear tones* are defined as sub-components of intonation that occur at the end of the intonation-group. They comprise the time from the last accented syllable through the end of the intonation-group. Thus, intonation-groups describe the entire utterance to be analyzed, while nuclear tones describe the conclusion of that utterance. Snow and Balog (2002) discuss the nuclear tone approach which is most often used to analyze the intonation of children, as well as an approach called the autosegmental theory which represents intonation through low and high pitch targets. The nuclear tone approach defines the following terms: declination, register, key, direction, and accent range. *Declination* refers to rises and falls in pitch across utterances. *Register* refers to whether the intonation contour is high or low in the speaker's typical fundamental frequency range; this relative pitch height is thought to express the emotion of the speaker. For example, high pitch is associated with fear, anxiety and distress. *Key* describes the width of pitch changes over a given time frame. *Direction* is used to describe rising or falling pitch in one nuclear tone. *Accent range* is used to describe the difference between the highest and lowest values of the intonation contour for a certain tone. The nuclear tones tend to have a pragmatic function; these tones communicate the speakers' intent. A falling contour is associated with statements, commands and Wh-questions. A rising contour is associated with uncertainty and yes/no questions (Snow & Balog, 2002). However, some of these pragmatic categories are ambiguous. For example, a tag question (e.g. "I liked it, didn't you?") may be produced with rising intonation or falling intonation and both require a response. The question produced with rising intonation may sound unsure, and the falling intonation may sound more expectant. Nuclear tones may also have an affective function, for instance, greater certainty regarding an utterance is associated with falling contour, and less certainty regarding an utterance is associated with rising contour.

It is through these parameters (i.e., declination, register, key, direction, and accent range) that a listener interprets the underlying intentions of a speaker's utterances. It is clear that adherence to these principles is essential to appropriately manipulate one's grammatical, pragmatic and affective prosody.

Development of Intonation

Many believe that children develop intonation before they produce their first words (Halle, Boysson-Bardies & Vihman, 1991; Jusczyk, 2002; Snow & Balog, 2002). However, some research contradicts this hypothesis (Snow, 2006). Snow and Balog (2002) reviewed studies of the development of intonation in children through 25 months of age as they related to grammatical, pragmatic and affective/attitudinal functions of pitch. They aimed to find support or lack of support for the notion that children develop intonation of speech before words. The authors found that in early speech, children typically use a falling pitch and lower fundamental frequency on the final word of an utterance to signify the utterance's conclusion. Snow and Balog concluded that accent range in falling intonation contours develops alongside the first 50 words, and the accent range of rising intonation develops later.

Snow (2006) studied 60 infants ranging in age from 6-23 months. The infants played with their mothers and the interactions were recorded. The experimenters performed an acoustic analysis on monosyllabic utterances. Snow (2006) concluded that accent range is significantly developing around 18-months of age. The authors suggested a U-shaped pattern of development. In other words, there was a regression of accent range from 9 to 11 months, and a rapid growth from 18 months to 20 months. These findings do not support the hypothesis that intonation is established before speech begins; instead, Snow (2006) suggested that intonation develops congruently with speech and before the production of word combinations. Snow (2007) conducted a follow-up study on the same group of children and analyzed polysyllabic vocalizations. This study revealed that falling contours had a greater accent range than rising contours. These data provided less support for a U-shaped pattern of intonation development, suggesting a less extreme regression of accent range from 9 to 11 months and a more linear development. Furthermore, Snow (2007) proposed that the development of intonation is a result of physiological constraints on the system rather than influences of the infant's native language.

However, Halle and colleagues (1991) found intonation development to be a result of the native language structure and to develop before language. They evaluated intonation patterns in four 18-month old children from French speaking homes and four 18-month old children from Japanese speaking homes. Each of these children used approximately 50 words. The French children produced largely rising intonation contours in babbling and in speech, and the Japanese children produced largely falling intonation contours in babbling and in speech. These results support the hypothesis that children have developed intonation systems before they begin producing speech (unlike the findings of Snow, 2006).

Furrow (1984) compared pitch height, pitch range, and loudness to social behaviors in 12 typical children ranging in age from 1:11 – 2:1 years of age. The authors analyzed free play sessions in the child's home with the child's parent. First, audiotapes were played in the absence of video to assure that the judges were focusing only on prosody. Second, the videotapes were analyzed to assign social behaviors (e.g., eye contact, private speech, etc.). Furrow used a rating scale to analyze the three parameters of the speech signal. The author's ratings for each utterance ranged from 3 (quiet, flat, low pitched voice) to 9 (loud, higher pitched, and exaggerated contour). The results suggested that children around 2-years of age vary prosody according to the context of their utterances. For example, utterances with eye contact typically correlate with louder, higher, more variable speech. Utterances that are impersonal are less variable, quieter, and lower in pitch. Perhaps this is because typical children understand a highly variable and louder utterance will more likely elicit a response from the conversational partner than a less variable and quieter utterance. Interestingly, the study by Sheinkopf and colleagues (2000) suggested that this is untrue in regards to the relationship between communicative function and vocal quality in the autistic population. The study by Furrow (1984) differs from other studies that measure rising and falling pitch as a function of communicative intent (questions, statements, etc.) because Furrow suggests a broader, holistic function of prosody in children; a greater fluctuation in pitch is used more often for social interactions. At lower mean lengths of utterances (MLUs), the differences were less significant, but became more significant with linguistic development. This suggests that typical children use prosody differentially for social and non-social communication.

Prosody in language development. It is evident that different researchers have found support for a variety of hypotheses regarding whether language or prosody develops first. Yet, most researchers agree that prosody and language are intimately related. Prosody likely plays a significant role in early language development. Infants might learn the sound patterns of a language before associating those patterns with meaningful words. Within the first few months of life, children engage in categorical perception (i.e., they can differentiate between an infinite number of phonemes; Jusczyk, 2002). However, as infants age they become more sensitive to their native language. Jusczyk (2002) suggested that infants learn the prosodic organization of their native language around 4 -5 months of age and learn specific phonemes of their native language around 8-10 months. Infants use this understanding of the sounds of their native language to segment words in running speech. Infants must differentiate allophonic variations of phonemes to correctly identify a word. Jusczyk provided the following example: the allophones /t/ and /r/ in “nightrate” are different than the allophones /t/ and /r/ in “night rate”. An infant must recognize these differences before he or she can delineate one word from two words. Jusczyk (2002) proposed that by 10.5 months, infants learning English separate these words using allophonic cues.

Bedore and Leonard (1995) suggest that pauses, durations and fundamental frequency changes are the most salient indicators of linguistic boundaries. It is believed that child-directed speech (CDS) provides cues for this prosodic segmentation because CDS uses pauses and changes in intonation that keep the child’s attention. CDS is characterized by a higher fundamental frequency, wider range of frequencies, shorter vocalizations, and repetition. The prosodic bootstrapping hypothesis proposes that CDS is the key to beginning language acquisition. Fernald (1989) filtered CDS and adult-directed speech to render utterances semantically unintelligible, while intonation contours were unaffected. The goal was to determine if CDS more effectively conveyed the communicative intent of the message through its prosodic characteristics compared to adult-directed speech. Eighty adult subjects participated in the study and were told they would hear speech that sounded as though it were “heard through a wall” (Fernald, 1989, p. 1503). The participants were significantly better at judging the communicative intent in CDS compared to adult-directed speech, suggesting that infants learn communicative intentions through intonation contours before they acquire language. This is different from what Snow (2006) suggested. There are four proposed steps included in the prosodic bootstrapping hypothesis (Fernald, 1994). First, CDS evokes innate, biological responses from the infant. For example, “No!” is typically spoken with low pitch, high intensity, and short Fo-contours which typically startles the infant and stops his or her movement (Fernald, 1989). Second, CDS is used to modulate the attention, arousal and affect of the infant. CDS tends to increase sustained joint attention between the adult and child. Third, the child starts to recognize the emotion of the caregiver and can engage in shared experiences with the caregiver. Fourth, CDS takes on a linguistic function and acoustically highlights words so that the child can discriminate single words from connected speech. Overall, this hypothesis states that infants have a biological response to certain stimuli in the mother’s or father’s voice and eventually these stimuli will become meaningful. One might propose that children who cannot perceive these stimuli will struggle to acquire language quickly and that the language they do acquire may be less meaningful.

Asperger (1952/1991) described a similar phenomenon, although he did not define it as the prosodic bootstrapping hypothesis. He suggested that the content of parents’ words is not what encourages the child to comply, rather the affect with which the caregiver produces the words is more important. He proposed that the affect of a request could be understood by infants, foreigners, or animals; although, each group would be unable to process the literal meaning of the language. Asperger proposed that children with ASD interpret others’ affect in an atypical way and the result is that they themselves tend to use an “unnatural” sounding voice (p. 70).

It is reasonable to expect that people with ASD will have difficulties with prosody. The widely accepted Theory of Mind hypothesis states that various behaviors in

ASD are due to the inability to perceive cognitive states of others. Across a variety of studies, children with ASD have performed significantly worse than their peers on Theory of Mind tasks (e.g., Tager-Flusberg, 1999).

Relationship between Theory of Mind and prosody. Peppe and colleagues (2007) proposed that when children with ASD are asked to judge photos according to the preferences of a recording (e.g., what the person on the computer likes or dislikes) they tend to judge the pictures according to their own preferences instead. For example, although the acoustic cue of a rising intonation suggests a "liking" quality, a child with ASD may hear a rising intonation for the word *mushrooms* and judge that the speaker dislikes mushrooms because the child dislikes mushrooms. Peppe, McCann, Gibbon, O'Hare, and Rutherford (2006) suggested that disordered comprehension of prosody may be responsible for the social difficulties that are characteristics of ASD. They proposed that disordered understanding of prosody may prevent children with ASD from understanding figurative language. Tager-Flusberg (1999) also suggested that people with HFA have difficulties grasping the intended meaning of a message rather than the literal meaning. Paul, Augustyn, and colleagues (2005) suggested that the majority of prosodic deficits identified in people with ASD are pragmatic or affective in nature, providing support for the Theory of Mind hypothesis.

In addition, Paul, Orlovski, Marcinko, and Volkmar (2009) mentioned that participants with HFA/PDD-NOS had deficits in topic maintenance, providing the correct amount of needed information, and reciprocity of conversation. These deficits may be related to a lack of presupposition skills. Presupposition refers to a person's ability to recognize the knowledge that another person currently has, which is similar to the Theory of Mind. The authors present this deficit as a separate issue from unusual intonation; in fact, they state that presupposition skills seem unrelated to intonation differences. However, in order to emphasize new information through prosodic changes, a person must be aware of what the listener considers to be new information. It seems that presupposition skills may be very important for the development of pragmatic prosody.

Assessing Prosody in ASD

People with ASD are consistently described as having an inappropriate tone of voice (Hubbard & Trauner, 2007) and yet prosody is one of the few characteristics seldom researched within this population (McCann et al., 2007). Peppe and colleagues (2007) conducted a literature review and found a variety of words used to describe speakers with ASD, including: dull, wooden, robotic, bizarre, sing-song, over-precise, and stilted. Fine, Bartolucci, Ginsberg, and Szatmari (1991) described the speech as pedantic, voluble, tangential, and lacking in inflection. It is interesting that these adjectives lack similarity; moreover, several of the adjectives appear to describe opposing characteristics. Although some of these words appear quite opposite, they are simultaneously used to describe "autistic speech" quite consistently. Adjectives like odd, exaggerated, halting, rapid, and jerky are used as part of diagnostic measures including the ADOS (Lord, Rutter, DiLavore, & Risi, 1999) and ADI-R (Le Cauteur, Lord, & Rutter, 2003).

The current literature regarding prosodic characteristics of people with ASD is scarce. The majority of researchers have relied on their own subjective judgments; although few researchers have used objective acoustic analysis. Most have recruited older children or adolescents, and some have investigated prosodic differences between autistic disorder and Asperger syndrome. Clearly, acoustic analysis would be optimal to confirm the presence and describe the nature of prosodic differences in people with ASD. Acoustic analyses would provide an objective way to define the speech of people with ASD rather than relying on the wide variety of undefined adjectives. Furthermore, it is essential to study younger children with ASD to increase the understanding of the development of expressive prosody in this population.

Several studies have investigated receptive prosody in children with ASD. Diehl, Benneto, Watson, Gunlogson, and McDonough (2008) conducted a study to determine whether 21 adolescents with HFA used prosody to understand syntax. The researchers provided three sentence conditions. First, sentences where only prosody could distinguish the meaning of the sentences were given (e.g., [put the dog in the basket][on the star]). Second, sentences where syntax was not ambiguous, but no prosodic cues were provided (e.g., [put the dog that's in the basket on the star]). Third, sentences where prosodic cues were provided with the syntax (e.g., [put the dog that's in the basket][on the star]). Adolescents with HFA experienced greater difficulty than their typical peers understanding the first group of sentences. These results suggested that people with HFA may be relying solely on sentence content and ignoring prosodic information. Furthermore, deficits in prosody may be grammatical in nature, as well as both pragmatic and affective, suggesting a deficit in the prosodic system as a whole. This study also suggests there is a deficit in the integration of different elements of communication (e.g., syntax and prosody). Diehl and colleagues (2008) found that the deficits in receptive prosody were correlated to receptive language scores, supporting the hypothesis that prosody and language are closely related.

Peppe and colleagues (2007) conducted a study to examine the receptive and expressive characteristics of prosody in children with HFA. The participants included 31 children with HFA and excluded persons with Aspergers syndrome. Profiling Elements of Prosodic Systems in Children (PEPS-C), which measures receptive and expressive prosody, was used in this study. In this assessment, test items were presented at two levels: *form* tasks, which examined auditory discrimination and imitation, and *function* tasks, which examined pragmatic, grammatical and affective prosody. The PEPS-C lists prosodic function tasks in four categories which include: *turnend* (rising intonation or falling intonation), *chunking* (prosodic boundaries in phrases), *focus* (stress or emphasis), and *affect* (understanding and using intonation which signals liking or disliking). The authors displayed photos on a computer screen that allowed for either the expression of a target utterance or the receptive identification of the utterance produced by the computer.

Peppe and colleagues (2007) found that participants with HFA were more likely to incorrectly identify two items that were the same as different. When children with HFA were asked to repeat a phrase from a speaker who sounded as though they liked something, the children with HFA often sounded disliking. Furthermore when the children with HFA were asked to repeat a sequence of items, they often failed to make prosodic breaks (i.e., pauses). For example, if a child was asked to say "fruit, salad, and milk", it often sounded as though he or she was saying, "fruit-salad and milk". In the expressive function tasks, children with HFA were significantly more likely to place stress on the incorrect word within a sentence (e.g., focus task). For example, when presented with a picture of a red sheep kicking a soccer ball and the verbal stimulus "the red cow's got the ball", children with HFA often responded with "No, the RED sheep's got it." Receptively, children with HFA were also significantly more likely to misinterpret the placement of stress in a sentence. During turnend tasks, children with HFA were perceived as questioning when a statement was required; in other words, their pitch would increase at the end of a phrase rather than decrease. During affect tasks, children with HFA tended to produce liking responses as disliking responses and disliking responses as liking responses. This was similar to the results of the auditory discrimination tasks described earlier. In the receptive function tasks, the children with HFA were similar in their confusion of liking and disliking. The children's difficulties in expressing stress within a sentence have been supported by previous studies (Shriberg et al. 2001). One limitation of this study was that subjective analysis was implemented rather than acoustic analysis.

Strong correlations between receptive and expressive prosody were found within the HFA group. This suggests that implementing therapy to target receptive prosody may also improve expressive prosody. The authors found that receptive prosodic abilities were correlated with verbal mental age. Yet, they found that expressive prosodic abilities were not correlated with verbal age or chronological age. These results suggest that receptive prosody in children with HFA is delayed, and expressive prosody in children with HFA is deviant (Peppe et al. 2007). When compared to previous literature, it is surprising that children with ASD were unable to correctly imitate the adults. Kanner (1943) observed that children with autism who were verbal repeated words with the exact intonation of another speaker, and it is often reported that children with ASD are able to repeat video games or television commercials verbatim, maintaining the same prosodic characteristics. These discrepancies suggest the need for further research in intonation of children with

ASD. Furthermore, although these authors suggested that expressive prosody is deviant, it is unclear as to how expressive prosody deviates.

McCann and colleagues (2007) emphasized the interplay between language and prosody in a follow-up study with the same 31 children with HFA who participated in the Peppe and colleagues (2007) study. They assessed speech, language, non-verbal, and pragmatic abilities with a variety of standardized tests and compared these measures to prosody measures to determine if there was a correlation. Surprisingly, prosodic ability appeared to be a more significant deficit than receptive language in the children with HFA; these authors also found that prosodic skills were highly correlated with expressive language ability and receptive language ability.

Both of these results tend to support the prosodic bootstrapping hypothesis (Fernald, 1989). Whether disordered prosody is a result of language delay, or language delay is a result of disordered prosody remains a perpetual question. Another question that persists is whether Theory of Mind deficits cause deficits in prosody or prosodic deficits cause deficits in Theory of Mind. Regardless of the direction and nature of these relationships, McCann and colleagues (2007) reiterate the need for acoustic analysis to objectively confirm the results of the PEPS-C.

Shriberg and colleagues (2001) examined prosody in an older population of speakers with HFA and with AS. In their study, 15 males with AS were compared to 15 males with HFA; both of these groups were compared to 53 typically developing male speakers ranging from 10-50 years old. The authors chose to include participants with Aspergers syndrome because they present with similar deficits in social interaction, communication, and play as persons with HFA. The Prosody-Voice Screening Profile

(PVSP; Shriberg, Kwiatkowski & Rasmussen1990) was selected to evaluate these speakers. The PVSP examines prosody and voice characteristics of speech in a conversational sample. It provides information regarding the speaker's phrasing, pitch, stress, rate, loudness, and quality. The examiner assumes that a prosodic deficit exists in one of these areas when the speaker has 80% of his or her utterances coded as inappropriate in that specific area. Few differences were found between persons with HFA and persons with AS; however, significant differences were found between both of these groups and the control group in the areas of: stress, phrasing, and resonance. The speakers with HFA and AS had 10% lower appropriate phrasing scores than the control group. Both groups were coded as having sound/syllable word repetitions. It is interesting that these phrasing abnormalities were unrelated to the speakers' increased rate of speech; in fact, phrasing errors were marginally associated with reduced speech rate. Misplaced word stress was most common in persons with HFA, although prolongations were most common for AS speakers. Both groups were coded as having less appropriate nasal resonance compared to the control group.

Shriberg and colleagues (2001) suggested that speakers with ASD are likely to experience normal conversation as a stressful task and they may produce the same dysfluencies as typical speakers would produce during a stressful task. This implies that it may be more beneficial to assess prosody for people with ASD through a continuous speech sample rather than to assess prosody through imitation and elicitation. Most of the inappropriate stress coded in this study was due to misplaced stress of a word within a sentence or phrase. This form of stress reflects a pragmatic decision instead of a grammatical decision (e.g., placing stress on a specific syllable of a word) suggesting that these speakers have a greater deficit in pragmatic stress than grammatical stress. These results differ from Diehl and colleagues (2008) who found deficits in grammatical prosody for adolescents with ASD. The results from Shriberg and colleagues' (2001) study may also be explained by deficits in Theory of Mind because a speaker must understand what information a listener would know, if he or she intends to emphasize the new information. The hypernasality noted in this study may be central to their listeners' inabilities to perceive the speakers' emotional states. These authors not only suggested that instrumental studies be conducted, but that they are conducted on younger children with ASD to understand prosody and voice characteristics through development.

Paul, Shriberg, and colleagues (2005) provided further information on the participants in the Shriberg and colleagues (2001) study. The authors administered the ADOS and the VABS Survey Form to assess communicative and socialization abilities. The authors compared these scores to phrasing, stress and resonance abilities from the previous study. Their goal was to ascertain which measures (phrasing, stress or resonance) most influenced listeners' perceptions of communicative competence. Stress and resonance were found to significantly influence listeners' perceptions of social and communicative competence in the individuals with ASD. The authors observed that abnormal stress is not necessarily predictive of abnormal resonance; rather, the two are independent of one another. These results suggested that clinical assessments must determine which areas of prosody are disordered before beginning treatment. According to this study, prosodic deficits in one area do not predict prosodic deficits in another.

Grossman, Bemis, Skwerer, and Tager-Flusberg (2010) recently conducted a study similar to the previously mentioned studies, but with one important difference - the authors used an acoustic analysis technique. Sixteen children and adolescents with HFA were compared to 15 TD children on measures including: perception of affective prosody, perception of lexical stress, and lexical stress production. For the production task, children listened to a series of sentences and were instructed to say the missing words which were illustrated and written on a notebook in front of them. Grossman and colleagues analyzed the mean pitch, intensity and duration of the words. Children with HFA disambiguated words (e.g., DOORmat, SMALL fish) through stress as measured by pitch and intensity. However, all of their productions were significantly longer than their typical peers. The authors noted that children with HFA often exaggerated the pauses between syllables. This study demonstrated that the productions of children and adolescents with HFA are acoustically different, even when performing a narrowly specified task as was done in this study. These objective differences provide further insight into the suprasegmental differences found in previous studies (McCann et al., 2007; Peppe et al., 2007; Shriberg et al., 2001).

Intonation and ASD

Paul, Augustyn, and colleagues (2005) examined perception and production of stress, intonation, and phrasing in 27 participants with ASD ranging in age from 14-21 years. The authors examined grammatical and pragmatic/affective prosody. The participants were provided with a verbal stimulus and asked to mark a response, point to a picture, or provide a verbal response depending on the nature of the item. These production tasks were subjectively scored by an examiner. This study found that stress is a deficit in people with ASD (receptively and expressively) which supported Shriberg and colleagues (2001). The authors found no differences between the TD group and the group with ASD on pragmatic/affective production of intonation. The intonation abilities were examined by instructing both groups to produce child-directed speech and adult-directed speech. The authors speculated that the TD children were embarrassed by the task, while the group with ASD was not bothered. They suggested this as a reason for the lack of disparity between the groups. Contrary to the suggestions made by Paul, Shriberg, and colleagues (2005) to treat specific deficits in one area of prosody, Paul, Augustyn, and colleagues (2005) suggested a metalinguistic approach to treatment that focuses on a range of prosodic elements rather than just one. There is apparent confusion regarding whether multiple prosodic characteristics are related to one another or if they are independent of one another.

Although Paul, Augustyn, and colleagues (2005) found that children with ASD did not differ from TD children, other studies have found unusual intonation in children with ASD (Diehl, Watson, Bennetto, McDonough, & Gunlogson, 2009; Paul et al., 2009; Fine et al., 1991; Hubbard & Trauner, 2007). Fine et al. (1991) studied three groups of adolescents ranging from 7- to 32-years of age to determine the differences in intonation of adolescents with HFA, adolescents with AS, and adolescents with a psychiatric out-patient diagnosis. A 10-minute interview regarding common topics such as, school, family, and vacations, was conducted with each participant. Each intonation boundary was denoted and determined as an appropriate or inappropriate location. The concepts of marked and unmarked boundaries was employed in this study; marked boundaries occur when the pattern of an utterance is dependent on the context of the communication exchange (e.g., she *gave* the book to Mary), whereas, unmarked boundaries can be used in a variety of communicative situations (e.g., she gave the book to *Mary*). Marked stress typically contrasts another thought or statement. Sentences that employ this stress pattern are only appropriate in specific contexts. The authors found that all groups were similar in their use of unmarked boundaries; however, the HFA group used inappropriate intonation when certain marked boundaries were expected. These results suggested that people with ASD have more difficulty using intonation to convey information in specific, socially-accepted ways.

The socially appropriate use of intonation in ASD has been investigated in other studies as well. Paul and colleagues (2009) used the Pragmatic Rating Scale (PRS; Landa, 1992) in 29 individuals with ASD from 12-18 years of age. The participants were divided into a HFA/PDD-NOS group and an AS group. It is important to note that this study and the studies by Shriberg and colleagues (2001), Paul, Shriberg, and colleagues (2005) and Paul, Augustyn, and colleagues (2005) were conducted at the Yale Child Study Center; thus, some of the participants may span across the studies. The PRS was used in this investigation, although it was originally developed to assess the parents of persons with ASD as a means of determining whether family members also presented with pragmatic difficulties. This scale is divided into three major behavior groups: pragmatic behaviors, speech and prosodic behaviors, and paralinguistic behaviors. The first 30 min of the interview taken during the ADOS were analyzed using the PRS. The ratings (normal, moderately inappropriate, and absent or highly inappropriate) were given at the end of 3 min segments of the interview. These were averaged and compared to the other major behavior groups. The participants with HFA/PDD-NOS had ratings that were significantly different from the typically developing control group on the following ratings: unusual intonation, inappropriate use of gaze, and conversationally “out of sync”.

Overall, Paul and colleagues (2009) reported that the largest differences for persons with HFA/PDD-NOS were in areas of intonation and gaze, and that those with AS also showed similar deficits in intonation. The AS group performed better on their use of gaze and worse on their use of formal language when compared to participants in the HFA/PDD-NOS group. This not only suggests that intonation abnormalities are present and noticeable in persons with an ASD, but that there may be gradations of prosodic difference on the autism spectrum.

The research by Diehl and colleagues (2009) supports this idea as their study found that pitch range correlated with ADOS scores. Paul and colleagues (2009) predicted that problems in volume, rate, or timing of speech may also be present in younger, lower functioning children with ASD. The authors concluded that the differences in prosody are currently unidentified and these authors explained that they are presently using neuroimaging to explore differences in intonation. They also discussed limitations in research regarding conversational skills in persons with ASD due to the lack of appropriate and valid measurement tools. The fact that the PRS, which was created as a rating scale for parents of children with ASD, was used to measure the use of intonation in this population emphasizes the lack of available, reliable instruments to measure prosody.

Evaluating prosody in ASD. The PVSP and PEPS-C are two of the most prominent prosody assessments that are currently used in the ASD population. Rating scales have been used to study various areas of speech and language including prosody; for example, Crawford, Edelson, Skwerer and Tager-Flusberg (2008) used a rating scale to determine the subjective impression of the prosody of children with William's

Syndrome from a rating of 0 (monotonous) to a rating of 3 (more expressive or dramatic). Similar ratings scales have not yet been used to understand the prosody in children with ASD; although, their use could provide insight into degrees of variability within this population.

Acoustic analysis in ASD. Acoustic analyses are lacking in the research of prosody in ASD. Studies that include these measures are important for a number of reasons. First, unusual intonation is a diagnostic marker of ASD (APA, 2000) and is assessed through clinical judgment on gold standard assessments of ASD including the ADOS and ADI-R (Diehl et al., 2009). More research at the quantitative level will allow the development of sensitive prosodic measures for this population. If these studies are conducted in a young population, they will provide an earlier, more accurate diagnostic tool. Second, if subjective descriptions can be matched to acoustic measures, then a universally accepted method to describe the speech of persons with ASD can be used with an objective basis. Third, acoustic analysis software can be used not only as a diagnostic tool, but also used to assess the progress of treatment. Various treatment programs could be developed to allow individuals with prosodic disorders a method of visualizing the changes in their prosody; this would be particularly useful in the ASD population as other studies have reported success when using visual cues or video modeling to teach other communicative behaviors (e.g., social skills) to children with ASD (Litras, Moore, & Anderson, 2010).

Several acoustic studies have revealed differences in the vocal behavior of young children with ASD. Schoen, Paul, and Chawarska (2010) recently studied vocal productions in toddlers with ASD through acoustic analysis. This study was divided into two sections. For both sections, the CSBS DP was administered to the toddlers in the study. First, the authors analyzed the duration and pitch range of non-linguistic vocalizations in 18- to 36-month-old toddlers and compared their measurements to the measurements of TD children. Using PRAAT software, the researchers concluded there was no significant difference between the groups for low pitch or pitch range, but a chi square test revealed a greater number of high pitch vocalizations in toddlers with ASD when compared to TD toddlers. The ASD group most often used a complex and rise-fall contour for their vocalizations. These complex contours consisted of irregular rises and falls within a breath group. The TD group most often used both flat and rise-fall contours for their vocalizations. The toddlers with ASD more often produced longer than 0.5 sec vocalizations when compared to the TD group. This study did not analyze words or word approximations because the authors were interested only in non-linguistic vocalizations.

The second part of the study was a perceptual analysis of speech and nonspeech vocalizations. A different group of toddlers with ASD and TD toddlers were included in this study. Vocalizations were separated into speechlike vocalizations and nonspeech vocalizations. Nonspeech vocalizations were segmented into utterances based on breath group (all productions within one breath) or when a pause of greater than one second occurred between utterances. Speechlike vocalizations were coded based on the number of vowels and consonants present in each utterance (i.e., level 1 included vowels and continuant single consonants, level 2 included single consonant vowel combinations which could be reduplicated, level 3 included syllables containing two or more different consonants). Nonspeech vocalizations were categorized as distress, hum, delight, atypical vocalization, or other. Toddlers who produced 10 or more meaningful words or word approximations were classified as being in the meaningful speech (MS) stage, and those who produced fewer than 10 meaningful words were classified as being in the premeaningful speech (PS) stage. The authors found that the TD group produced significantly greater number of consonants than either the ASD PS and ASD MS groups. The ASD group produced a significantly greater number of nonspeech vocalizations than the TD group.

This study supports the hypothesis that children with ASD use prosody differently than TD children. Schoen et al. (2010) found greater fluctuation of pitch in the prelinguistic vocalizations of children with ASD. These authors used perceptual analysis and acoustic analysis, but they did not perform these analyses on the *speech* of children with ASD. Furthermore, the perceptual analysis and the acoustic analysis were done on different groups of children; thus, the two analyses could not be compared to one another. Their finding that toddlers with ASD produce more complex utterances consisting of greater pitch fluctuation suggests that children with ASD may use exaggerated prosody as Asperger described. Furthermore, the descriptions of longer vocalizations in ASD suggest an atypical rhythmic quality. These authors suggest that this analysis be performed during the language learning process of young children to understand how their prosody develops alongside language.

However, the majority of acoustic analyses have been conducted on older children with ASD. Shaw and Nadig (2010) also used the PRAAT software to analyze contrastive stress in the speech of children with HFA. These authors calculated change in pitch, amplitude, and duration of the primary syllable of the adjective relative to the other syllables in a sentence (e.g., pick up the BIG cup). Shaw and Nadig found that both typical children and children with HFA used increased pitch, amplitude and duration to mark these adjectives; however, children with HFA used amplitude less often to mark these syllables when compared to their typical peers.

Diehl and colleagues (2009) analyzed prosody production in adolescents with HFA through naturalistic speech samples. First, they examined the within-subject variation in fundamental frequency in adolescents with HFA compared to TD matched peers. The participants watched a cartoon and retold the story to someone they were told had not seen the cartoon before. The narratives were digitized and analyzed using a speech analysis and synthesis system. The authors collected a data point every 250 ms and analyzed pitch range by calculating the standard deviation. The adolescents with HFA had significantly larger standard deviations (pitch ranges) than their typically developing peers; however, there was not a significant difference in the average pitch of the participants. The authors also found a significant correlation between the ADOS score of participants with HFA and their pitch range, suggesting that greater variation in intonation is associated with greater perceived communication impairment. This implies a correlation between acoustic measures and diagnostic measures.

Diehl and colleagues (2009) repeated this study with a younger group of children with ASD (ages 6-14 years). The participants listened to a story while looking at a picture book, rather than watching a cartoon, and were then asked to tell the story without looking at the pictures. This study also revealed a greater within-subject variation in intonation for the children with ASD, however it was not correlated to ADOS scores as was found for the adolescent group. This repetition of the study did not support a correlation between acoustic and diagnostic measures. Furthermore, although Diehl and colleagues (2009) found significant differences between fundamental frequency variation in children with ASD and typical children, they also found a significant overlap between the two groups. This suggests that either children with ASD produce several utterances with atypical prosody and other utterances with typical prosody, or that some children with ASD are more typical in their productions than others. The latter of these may explain the variation in descriptions used to categorize the speech of people with ASD (Hubbard & Trauner, 2007; Peppe et al. 2007).

Furthermore, Diehl and colleagues (2009) suggested that calculating standard deviation of fundamental frequency as the sole measure of prosody variance will not adequately represent the characteristics of prosody in ASD. Although these authors found inconsistent results regarding the relationship between clinician judgments of communication ability and acoustic analysis, Hubbard and Trauner (2007) found marked trends between subjective interpretations and acoustic measures, however the relationship did not reach statistical significance. The authors compared acoustic features (i.e., amplitude and fundamental frequency) to subjective ratings of affect (i.e., happy, sad, and angry) in repetition and free-response tasks for 28 children ranging in age from 6-21. Their study compared children with ASD, children with Aspergers and children who were TD. Hubbard and Trauner predicted that children with ASD would have a more flat pitch contour (i.e., monotone), but their results indicated that children with ASD have a much greater pitch range than TD children. This monotone expectation comes from one of the many adjectives used to describe the speech of children with ASD (see Peppe et al., 2007). However, their results suggested a sing-song prosody which is another adjective used to describe this speech.

In repetition tasks, children with ASD were found to have the greatest pitch range compared to children with Aspergers and TD children. However, all participants tended to make emotional distinctions of their pitch in the same way; they all used the highest pitch to represent a happy emotion, a mid-pitch to represent an angry emotion, and the lowest pitch to represent a sad emotion. Although children with Aspergers and TD children used amplitude to distinguish anger from other emotions, children with ASD did not. Children with ASD were also the only children who tended not to use duration as an emotional cue (e.g., did not use slower speech when expressing a sad emotion).

Hubbard and Trauner (2007) obtained subjective ratings through a “free-response” measurement, which was elicited by telling the children a story that was intended to evoke a specific emotion (i.e., happy, sad, angry) and asking them to complete the story in one sentence. However, this form of elicitation is not guaranteed to produce the same emotions in all participants and its validity may be questionable. Furthermore, there were difficulties obtaining free-response data from children with ASD due to their perseverating on the content of the story, failing to make first-person statements, and responding with single-words rather than sentences. Four of the children with ASD did not complete this task because they did not understand the instructions. However, the results of the study are intriguing. Those who completed this section showed a less than significant correlation between emotion and pitch range. Children with ASD appeared to have atypical locations of maximum pitch within a phrase. This suggests an atypical rhythmic quality, perhaps described as jerky speech.

Primarily, Hubbard and Trauner suggested that children with ASD have a much greater pitch range than typically developing children and may “overshoot” their intonation. The children with ASD tended to have an exaggerated pitch range more often during repetition tasks when compared to free-response tasks; the authors proposed that children with ASD are engaging in sound mimicry rather than formulated, emotional expression. This study also demonstrated that children with ASD produce flatter amplitude with less variant duration in speech. Subjective ratings had marked trends with acoustic features, but were not significantly correlated with acoustic measurements. This discrepancy between subjective measurements and objective measurements of prosody suggests the need for future research to examine the use of individual, acoustic features in speech and compare these features to subjective measurements. Lastly, this study allows for the development of hypotheses to explain a few of the adjectives used to describe the speech of people with ASD. Perhaps speech in ASD is described as monotone due to flat amplitude and flat duration, and yet still described as sing-song due to their great pitch range and overshoot of intonation.

Subjective descriptions. Both, speech that is monotone and speech that is sing-song may present linguistic and pragmatic challenges for the communication partner. Peppe and colleagues (2006) explain that people with ASD who use monotonous speech often give the impression that the speaker is depressed; moreover, the emphasis of the utterance is sometimes lost and the listener may find it difficult to recognize when the utterance is finished. However, prosody that is exaggerated may seem patronizing and less socially acceptable. Pitch movement exceeding the typical range may make the speaker appear inattentive to the pragmatic context (Diehl et al., 2009). These different qualities will create a speaker who sounds odd and a listener who is confused and frustrated.

Asperger (1952/1991) described four children who he believed were prototypical of children with ASD. He described the first boy, Fritz, as delayed in his motor milestones, but as learning to talk quite early. Fritz began using sentences very early, and was described as sounding "like an adult" (p. 39). His voice sounded high, thin, and far away. Asperger wrote that natural speech melody was missing in Fritz's voice. He spoke slowly and each word had a long duration. In other words, his prosody was characterized by atypical rhythm and halting speech. Fritz used a high degree of modulation in his voice; his speech often sounded "sing song" (p. 42) and he frequently spoke this way when demands were made on him. Although Asperger described much of Fritz's speech as sing-song in quality, Asperger also described Fritz's speech as missing melody and using a long duration which appear characteristic of monotone. If Fritz was both monotone and sing-song in his prosody, perhaps these two characteristics are not mutually exclusive. Interestingly, Fritz tended to encompass a variety of other prosodic attributes suggesting that an array of adjective are necessary to adequately describe the speech of an individual child with ASD.

The next child Asperger described was Harro who displayed quite different vocal characteristics. His voice was very deep and "appeared to come from very far down, in his abdomen" (p. 52). His speech appeared to be quite monotone as he spoke slowly and without modulation. Ernst, the third child, was described as having speech characteristic of a caricature; possibly very sing-song in nature as well. Ernst talked continually and would go on-and-on even when a short answer would have suited. The last child, Hellmuth, was described similarly to Fritz. He learned to speak at a young age, and was quickly talking "like a grown up" (p. 65). He spoke slowly and sounded clever and dignified. He seemed to be speaking in verse or in sing-song. Asperger (1952/1991) drew several conclusions regarding the nature of speech in children with ASD,

> Sometimes the voice is soft and far away, sometimes it sounds refined and nasal, but sometimes it is too shrill and ear splitting. In yet other cases, the voice drones on in a sing-song and does not even go down at the end of a sentence. Sometimes speech is over-modulated and sounds like exaggerated verse-peaking. However many possibilities there are, they all have one thing in common: the language feels unnatural, often like a caricature, which provokes ridicule in the naïve listener. (p. 70)

Research Objective

The purpose of the current study is to determine which of 12 chosen prosodic descriptors used in research literature and diagnostic tests (see chapter 3) are best at characterizing the speech of children with ASD. This will be accomplished by analyzing two listeners' auditory perceptions of the speech of the children with ASD.

First, the listeners' perceptions of the prosodic descriptors within each child will be investigated to determine the degree to which the prosodic descriptors are present within the children with ASD. Second, the relationships between the 12 descriptors will be evaluated. Specifically, the relationship between two descriptors – sing-song and monotone – will be investigated. Third, the first 11 descriptors will be evaluated to determine their unique contribution to an odd quality within individual speech utterances, within individual children, and within the entire group of children. This will be assessed separately (individual vs. group) as Diehl and colleagues (2009) suggested that narratives of children with HFA have more within-subject variability in their prosody than typical peers. Fourth, a pitch analysis will be performed on two of these descriptors – sing-song and monotone – to objectively quantify them. This study might provide greater clarity regarding the melodic extremes that are so often perceived in the speech of children with ASD.

Hypotheses

In regard to the research objectives stated above, the following hypotheses are proposed:

1. It is thought that each of the prosodic descriptors used in the ratings will be identified in the children with ASD at an atypical level (>1 rating on 1 – 7 scale). Kanner (1943) and Asperger (1952/1991) described the verbal children with ASD with multiple adjectives (e.g., mechanical, odd, sing-song, etc.) However, it is also thought that children with ASD will receive typical ratings on a large percentage of their speech; Diehl and colleagues (2009) found a greater Fo variation in children with ASD, but also found significant overlap between several children with ASD and typical children.

2. It is hypothesized that individual children with ASD will produce individual thought groups characterized by atypical levels of multiple descriptors at one time. In fact, the same thought group may appear to encompass qualities of seemingly opposing characteristics. Asperger (1952/1991) described Fritz as using sing-song speech as well as speech that was lacking typical melody.

3. It is likely that sing-song and monotone will be negatively correlated; although, they will likely be perceived to co-occur within certain thought groups. Furthermore, it is probable that sing-song and monotone will be found to have a relationship with intonation and rhythm; Paul and colleagues (2009) hypothesized that both pitch

abnormalities and timing differences would be present in younger children with ASD. Several studies have described children with ASD whom produced questions as though they were statements and sounded "staccato" as also being sing-song or monotone (Asperger, 1952/1991; Kanner, 1943; Peppe et al. 2007).

4. Furthermore, it is hypothesized that thought groups rated as highly odd (prosodic descriptor 12) will also receive atypical ratings on the majority of the 11 remaining descriptors. In other words, high ratings of the first 11 descriptors will predict a high rating of the 12^{th} descriptor. Specifically, it is suspected that high ratings of monotone and sing-song will be most associated with oddness in children with ASD. It is likely that there will be less predictability within the group of children, as compared to the individual child, due to the likelihood that different characteristics predict oddness in different children.

5. Lastly, it is hypothesized that monotone and sing-song will result in their own explicit differentially patterned pitch traces. For example, a high degree of fundamental frequency fluctuation within the thought group may be a predictor of sing-song quality while a low degree of fundamental frequency fluctuation within the thought group may be a predictor of monotone quality. However, it is also possible that this fundamental frequency fluctuation is present within both monotone and sing-song thought groups and the overall differentiating acoustic characteristic is the slope of fundamental frequency change throughout the entire thought group. In the latter scenario, a shallow slope might be perceived as more monotone while a steeper slope is perceived as more sing-song.

Chapter 3

Method

Participants

Children were selected from an ongoing research study being conducted at The University of Memphis. The purpose of the research study is to identify early markers of ASD in young children. To be included in the ongoing research study, the children had to meet the following criteria: (a) diagnosis of ASD within the previous 6 months, (b) no known hearing or visual impairments or co-morbid diagnoses, (c) monolingual, English speakers, (d) no prematurity, (e) no low birthweight, and (f) self-identified as Caucasian or African American. To be selected for the current study the children also were required to be verbal communicators (i.e., producing at least 2-word combinations).

For the current study, 7 children were selected from the database (see Table 1 for demographic information). The participants included 5 males and 2 females between 39 and 63 months of age (M = 50.79, SD = 8.84). However, one of the children was excluded from the analyses due to inefficient audio playback. All participants were white and recruited from communities in the Mid South region of the United States. A diagnosis of ASD was confirmed through administration of the ADI-R (Le Cauteur et al., 2003) and the ADOS (Lord et al., 1999), as well as by adhering to the diagnostic criteria established by the DSM-IV-TR (American Psychological Association, 2000).

Table 1

Summary of Participant Demographics

Sample Size	7	
Child's Age at Initial Evaluation in Months (*M*, *SD*)	50.79	8.84
Gender		
Female	2	
Male	5	

Table 2

Summary of Parent Demographics

	Parents' Education in Years Completed		Parents' Age at Child's Birth	
	M	SD	M	SD
Mother	16	2.03	30	2.73
Father	16	2.41	36	8.41

Procedures

All evaluations were completed in a small clinical room and were audio and video recorded for later data analyses. Video recordings were conducted in a sound-treated room outfitted with four cameras. The multiple cameras were used to give the child freedom to move around the room without losing a frontal view of the child. To obtain the audio recordings, the children wore vests that were equipped with a wireless microphone (Countryman Isomax EMW Lavalier) and wireless transmitter (Samson Airline UHF AL1) that sent a signal to a receiver (Samson UHF AM1). TF32 software operating a DT322 acquisition card (Data Translation, Inc., Marlboro, MA) was used to digitize the speech signals at 48 kHZ after low-pass filtering at 20 kHz using a Data Translation AAF-3 antialiasing board. This equipment assured high quality audio recordings for later analyses.

The children and their caregivers participated in an evaluation including the Mullen Scales of Early Learning (MSEL; Mullen, 1995), Communication and Symbolic Behavior Scales Developmental Profile (CSBS DP; Wetherby & Prizant, 2002), and a structured play sample. The MSEL is used to evaluate the cognitive development of children through visual reception, fine motor, and receptive and expressive language scales. Participants were more than 1 SD below the mean on all measures of visual reception, fine motor, receptive language, and expressive language with the exception of two children who demonstrated visual reception skills within 1 SD above the mean (for additional information see Plexico, Cleary, McAlpine, & Plumb, 2010). The CSBS DP is used for children with a functional communication age between 6 and 24 months and may be used for older children whose functional communication age falls within this range. The CSBS DP is designed to elicit spontaneous communication and play behavior. For the current study, administration of the CSBS DP allowed for a structured context for sampling natural communication, regardless of the child's language level. The 20 minute play sample between the caregiver and the child included a standard set of toys and books. Caregivers were encouraged to actively participate. The parents were instructed to respond naturally and to avoid directing the child's behavior.

In a previous analysis, the archived speech samples including portions of the CSBS DP and the play sample were transcribed. Diehl and colleagues (2009) suggested that future acoustic analyses in the ASD population include both a spontaneous speech sample (e.g. play) and a standardized speech sample (e.g. CSBS DP) to avoid the magnitude of variance in content, length, and manner of utterances that is often present in a purely narrative sample. A total utterance, word, and syllable count for each sample was calculated. The mean length of utterance in morphemes (MLU) was calculated using 75 utterances from the transcribed sample (see Plexico et al., 2010). The mean MLU was 2.82 (SD = 0.92). These results indicate the participants in the current study were producing multi-word utterances; however, 6 of the 7 participants were producing utterances that contained fewer morphemes than expected for their chronological age.

Data Analyses

The recording for each child was digitized and converted into video and audio files used for coding in Action Analysis Coding and Training (AACT) system (Delgado, 1996) and TF32 software (Milenkovic, 2001). This software displays videos, waveforms, and spectrograms. A 600 syllable sample was taken from the archived speech samples of the CSBS DP and structured play (see Plexico, et al. 2010). This sample was divided into utterances where each utterance included an uninterrupted thought referred to as *thought groups*. Thought groups included one idea or notion in the child's natural play or conversational turn. Gottman (1983) defined these segments as "one expressed idea or unit" and used these to assess conversation between children. He suggested that this unit of analysis serves as a data reduction technique to allow a set of words or short phrases with the same meaning to be coded as one. If the child was interrupted by the parent or the examiner, the thought was broken into several thought groups so the coding groups did not contain adult vocalizations. Thought groups were chosen to maintain the most natural listening situation for the raters. Furthermore, the majority of these segments also provide verbalizations of adequate length to complete all 12 ratings on each thought group. The exception is that thought groups consisting of one-word utterances are inadequate to rate temporal qualities (e.g., rhythm).

Excluded thought groups. Thought groups were excluded from analysis if the mother or clinician was speaking at the same time as the child thus making it difficult to develop clear impressions of the speech. Thought groups that contained vocalizations rather than word approximations (e.g. laughing, squealing) or animal noises (e.g. baa) were excluded from the analysis.

***Subjective Analysis and Ratings*.** The current study implemented a rating scale to measure the prosody of children with ASD as judged by naturalistic listeners. Two graduate students, representing naturalistic listeners, rated each thought group on a 7-point scale for 12 descriptors divided into the following five groups: group 1 (soft/loud and slow/fast), group 2 (sing-song and monotone), group 3 (intonation and modulation), group 4 (rhythmic, jerky and halting), and group 5 (fluency, intelligibility, and odd) (see Table 3). These ratings were chosen from a larger list of the most salient descriptors found in a literature review (see chapter 2) as well as characteristics mentioned in diagnostic measures such as the ADOS (Lord et al., 2000) and ADI (Le Cauteur et al., 2003). The raters rated group 1 (soft/loud and slow/fast) on Likert scales. When rating loudness, a rating of 1 represented extremely soft speech and a rating of 7 represented extremely loud speech. When rating the rate of speech, a rating of 1 represented extremely slow speech and a rating of 7 represented extremely fast speech. Group 1 was different from groups 2-5 because on each of the remaining scales, a rating of 1 represented a typical level of the descriptor and a rating of 7 represented an extreme level of the descriptor. After piloting the rating scale with two different graduate students (not the raters), it was decided that it was often difficult to rate the descriptors in group 4 when the thought groups consisted of only one word. Thus, a "not applicable" (N/A) category was added to this group's ratings for one-word thought groups because these did not present enough information to be adequately described as atypical or typical rhythm, jerkiness, or halting quality. The raters were permitted to play the thought group once before each group of ratings (5 times total), but no more than 5 times in order to simulate the most naturalistic listening environment. An instruction sheet providing specific descriptions of the extremes for each rating scale (ratings of 1 and 7) was given to both graduate students as a reference to be used while they were rating the thought groups (see Appendix A). The raters listened to thought groups of the 600 syllable sample from each of the seven selected children and used Microsoft Access 2007 to rate each selected thought group. Prior to beginning the ratings, the graduate student raters were played five examples that the primary researcher judged to be unambiguous regarding the various characteristics to be rated. The raters were permitted to discuss their reasoning for particular ratings during the introduction session only and were then required to make judgments independently throughout the rating process. The raters were not trained to reliability in an effort to keep the judgments as naturalistic as possible.

Table 3

Groups of Descriptors

Group	Descriptors
1	Soft/Loud
	Slow/Fast
2	Sing-Song
	Monotone
3	Intonation
	Modulation
4	Rhythmic
	Jerky
	Halting
5	Intelligibility
	Odd

Acoustic Analysis. TF32 uses a pitch determination algorithm with variable parameters that allows the computer to track changes in fundamental frequency. The tracking of these changes results in a pitch trace. There are six parameters that must be adjusted to provide the most accurate pitch trace. Each parameter has a default setting which can be adjusted globally or locally. In this study, the pitch analysis consisted of three phases: (a) global parameters, (b) local parameters, and (c) hand-editing. The global parameters phase allows the researcher to adjust any or all of the 6 parameters to numbers that appear to provide the best version of the pitch trace for the entire sample. These global changes do not affect each section of the pitch trace equally; thus, further changes must be made. Next, the parameters are adjusted at specific locations where the global changes are insufficient for a particular utterance. Lastly, hand-editing is the most detailed stage and involves marking individual glottal pulses, interpolating gaps, and zeroing pitch information that was not produced by the participant (e.g., if the mother's voice was loud enough to be recorded from the child's microphone).

These three phases were completed for the same 600-syllable sections coded in the subjective analysis. The researcher who conducted the pitch analysis was also responsible for separating the speech sample into individual thought groups for the subjective analysis. The pitch analysis was conducted prior to the subjective analysis so that the experimenter had no access to the subjective ratings during the pitch analysis. Although this three-phase pitch analysis was conducted on the entire 600-syllable sample for each child, individual thought groups were selected for further analysis based on the sum value (rating of sing-song/monotone of rater 1 + rating of sing-song/monotone of rater 2) of both sets of ratings for group 2 (sing-song and monotone). Specifically, the following four categories were selected based on sing-song and monotone qualities: 1) both = the highest numbers obtained when the sums of sing-song and monotone (sum of both raters) ratings were multiplied, 2) sing-song = high ratings on sing-song (sum >/= 4) with low ratings on monotone (sum = 2), 3) monotone = high ratings on monotone (sum >/=4) with low ratings on sing-song (sum = 2), and 4) neither = low ratings on both sing-song and monotone with a sum value of 2 for both descriptors. The low ratings were required to equal 2 because a rating of 2 could only occur when both raters judged the child to be typical (rating of 1). Multiplication was used for the both category because the product was more sensitive to high ratings by both coders than a sum. These delineations were chosen to isolate thought groups either where both monotone and sing-song components were present in the segment, where only one characteristic was present in the absence of the other, or where neither characteristics were present. The goal was to analyze three exemplars in each of the categories; however, fewer exemplars were analyzed for certain children when three thought groups did not meet the stated criteria.

For each of the selected thought groups, the Fo value for each thought group (sampling rate = 100 Hz) was extracted from AACT into an excel file. The mean Fo, standard deviation (Hz), length of thought group (ms), number of syllables, and length of individual modulations in the pitch contour (ms) were recorded directly from the AACT program. Within the extracted excel file, the data were used to measure the depth of fundamental frequency modulation across a thought group, the slope of the fundamental frequency change across the thought group, and the duration of the thought groups. The measurements were made by creating a trend line with fundamental frequency (on the ordinate axis) over time in milliseconds (on the abscissa) which provided the slope of the fundamental frequency change across time. This measurement represented the overall declination of the thought group. The trend line was subsequently subtracted from the fundamental frequency values to derive modulation in the absence of the overall declination of the thought group. This provided an opportunity to measure the maximum and minimum pitch within the thought group to assess pitch variability independent of across-utterance pitch changes (de-trended Fo SD) and compare this to the pitch variability which included the across-utterance pitch changes. Therefore, the resulting measurements for each thought group included: Fo SD, duration, number of syllables, Fo mean, slope of trend line, and de-trended Fo SD. If the perception of sing-song quality was most evident within a subsection of the entire thought group, the calculations were made based on this subsection. However, for the majority of segments where a sing-song quality was audibly present, this quality was present throughout the entire thought group.

Cycles of pitch modulation were identified in several thought groups. These cycles were defined as rhythmic increases and decreases in fundamental frequency that repeated a minimum of two times. They were not present in all thought groups; however, the thought groups that did have cycles were further analyzed to investigate the rhythmic qualities of speech in these children. Specifically, the duration of the cycle from peak-to-peak was measured for each cycle within a thought group. This was done for the thought groups in the sing-song group, monotone group, and both group. The neither group was excluded and these groups were chosen to investigate the acoustic qualities that Sing-Song and Monotone individually add to a speech segment in terms of rhythm.

Chapter 4

Results

Descriptive Analyses

The mean and standard deviation (SD) were calculated for each of the 12 prosodic descriptors within the group of children. The first 2 descriptors (soft/loud and slow/fast) were rated on a scale where 1 and 7 represented two extremes of one feature; atypical levels would be evident if the means were close to a rating of 1 (soft or slow) or close to a rating of 7 (loud or fast). Both soft/loud ($M = 3.77$) and slow/fast ($M = 3.88$) were judged to be fairly typical. Table 4 includes the remaining 10 descriptors (sing-song, monotone, intonation, modulation, rhythmic, jerky, halting, fluent, intelligibility and odd). The means and SD of these descriptors revealed that the children's speech was judged to be atypical relative to the rating scale provided on each of the 10 descriptors. Oddness at the level of the thought group was the most atypical of all the descriptors ($M = 3.63$, $SD = 1.17$). Modulation ($M = 1.37$, $SD = .74$) and fluent ($M = 1.12$, $SD = .53$) were the least atypical of the descriptors (see appendix for means and SD separated by rater). The means and SDs illustrate that different children's speech was determined to be atypical in different areas. For example, Child G was judged to have atypical rhythmic and halting speech when compared with the other children. Child F was judged to be monotone and highly unintelligible. Child B and Child C were perceived as very sing-song with atypical intonation patterns, whereas, Child E was judged to be monotone with atypical rhythmic qualities.

Table 4

Means and Standard Deviations of Descriptors

	N	Sing-song	Monotone	Intonation	Modulation	Rhythmic	Jerky	Halting	Fluent	Intelligibility	Odd	Overall odd
Child												
A	96	1.76 (.70)	1.95 (.79)	2.38 (.89)	1.31 (.59)	2.19 (.71)	1.98 (.87)	1.65 (.86)	1.05 (.29)	2.47 (1.20)	3.31 (.91)	3
B	166	2.79 (.96)	1.86 (.77)	3.03 (.90)	1.16 (.42)	2.60 (.76)	2.61 (.86)	1.98 (.92)	1.03 (.22)	3.57 (1.55)	3.54 (.91)	3.5
C	117	3.37 (1.08)	1.38 (.56)	3.49 (1.00)	1.47 (.81)	2.60 (.96)	3.36 (.89)	2.18 (1.23)	1.62 (1.27)	2.92 (1.90)	3.72 (1.11)	4
D	147	1.76 (.82)	1.86 (.77)	2.30 (.92)	1.03 (.13)	2.79 (.74)	2.73 (.69)	1.94 (.93)	1.10 (.43)	2.73 (1.47)	3.34 (1.07)	4.5
E	130	1.54 (.74)	2.13 (1.07)	1.99 (.97)	1.32 (.70)	3.24 (1.02)	2.89 (1.04)	1.61 (.84)	1.05 (.24)	2.28 (1.63)	3.52 (1.14)	4.5
F	150	1.80 (.97)	2.71 (.95)	2.52 (.88)	1.36 (.62)	1.75 (.74)	2.99 (.88)	2.24 (.97)	1.05 (.18)	3.74 (1.52)	3.90 (1.04)	5
G	309	2.03 (1.06)	2.54 (1.03)	2.50 (1.13)	1.66 (.97)	3.39 (.81	3.10 (.96)	2.48 (1.09)	1.09 (.34)	2.15 (1.51)	3.81 (1.41)	6

Table 4 (continued)

Means and Standard Deviations of descriptors

	N	Sing-song	Monotone	Intonation	Modulation	Rhythmic	Jerky	Halting	Fluent	Intelligibility	Odd	Overall odd
Rater												
1		2.19 (1.33)	2.37 (1.41)	2.44 (1.46)	1.41 (.94)	3.08 (1.17)	2.84 (1.32)	2.18 ((1.21)	1.12 (.55)	2.93 (1.74)	3.73 (1.27)	
2		2.08 (1.17)	1.94 (1.02)	2.74 (1.18)	1.34 (.76)	2.58 (1.14)	2.83 (1.12)	1.92 (1.22)	1.13 (.56)	2.62 (1.94)	3.53 (1.36)	
Total		2.14 (1.07)	2.15 (.99)	2.59 (1.07)	1.37 (.74)	2.83 (.90)[a]	2.83 (.97)[a]	2.05 (1.03)[a]	1.12 (.53)	2.78 (1.66)	3.63 (1.17)	4.36

[a] N = 842 (excluded one-word utterances)

Table 4 presents the children in order from the least overall odd to the most overall odd. Child A was determined to be the most typical sounding child overall. This child also had the lowest rating on jerky and the lowest rating of odd at the level of the thought group. Child G was determined to be the most atypical sounding child overall and had the highest rating on rhythmic and modulation; unlike child A (the lowest overall odd), he had high ratings on jerky and was the only child with equally high mean monotone ($M = 2.54$) and sing-song ($M = 2.14$) ratings.

Correlations of descriptors

Pearson product-moment correlation coefficients were computed among the 10 descriptors for the entire group of children. The results of the correlation analyses presented in Table 4 show a large number of statistically significant correlations among the descriptors. These correlations were analyzed to determine which characteristics, excluding odd, were correlated with one another and were then analyzed to determine which characteristics were most correlated with a rating of extreme oddness at the level of the thought group. Table 5 presents these data which were calculated from the means of rater 1 and rater 2. The raters' individual correlation matrices are available in Appendix B. The descriptors representing temporal aspects of the speech signal (rhythmic, jerky, and halting) were all significantly correlated to one another with the strongest relationship between jerky and halting ($r = .527, p < .001$). Sing-song and intonation were the most largely correlated descriptors amongst the group of children.

Table 5

Correlations Among Raters' Mean Ratings of Descriptors for Total Group of Children (N = 1115)

	Monotone	Intonation	Modulation	Rhythmic[a]	Jerky[a]	Halting[a]	Fluent	Intelligibility	Odd
Sing-Song	-.315***	.672***	.051	.034	.225***	.085*	.139***	0.206***	.339***
Monotone		-.307**	.178***	.366***	.136***	.217***	-.057	.057	.234***
Intonation			.143***	.130***	.328***	.219***	.094**	.237**	.483***
Modulation				.303***	.303***	.222***	.101**	.003	.393***
Rhythmic					.446***	.296***	.105**	.113**	.563***
Jerky						.527***	.245***	.106**	.575***
Halting							.269***	.037	.406***
Fluent								-.001	.089***
Intelligibility									.521***

*** p < .001.

** p < .01.

* p < .05.

[a] N = 842 (excluded one-word utterances)

Pearson product-moment correlation coefficients were calculated among the 10 descriptors for each individual child. These correlations were calculated from mean values of rater 1 and rater 2. Among all descriptors, sing-song and intonation were found to be the most correlated for Child D ($r = .787, p < .001$), Child E ($r = .592, p < .001$), Child C ($r = .815, p < .001$) and Child G ($r = .584, p < .001$). Although Child A, B, and F presented with correlations between sing-song and intonation, their largest correlations were among other descriptors. Jerky and halting were most correlated for Child F ($r = .589, p < .001$) and Child A ($r = .841, p < .001$); although Child B was found to have the largest correlation between jerky and monotone ($r = .530, p < .001$). These correlation tables are available in Appendix C.

Sing-song and monotone. The descriptors of greatest interest – sing-song and monotone – were negatively correlated with one another, $r = -.315(1113), p < .001$. It was hypothesized that sing-song and monotone would be related to intonation and rhythm; however no hypothesis was made regarding the nature of that relationship. Sing-song was largely correlated with intonation ($r = .621, p < .001$) and monotone was correlated with rhythmic ($r = .366, p < .001$). Monotone was found to have a significantly negative correlation with intonation, $r = -.307(1113), p < .001$; whereas, sing-song was found to have virtually no relationship with rhythmic.

Each of the individual children also had significantly negative correlations between monotone and sing-song. Monotone was positively correlated with rhythmic in all of the children excluding Child B and Child G where correlations did not reach significance. Monotone was negatively correlated with intonation in all of the children excluding Child C where the correlation did not reach significance (see Appendix C).

***Oddness*.** The 9 remaining prosodic descriptors were largely correlated with the perception of oddness of the thought groups at $p < .001$. Oddness was most correlated with jerky, $r = .575(840), p < .001$, rhythmic, $r = .563(840), p < .001$, and intelligibility, $r = .521(1113), p < .001$ (see Table 5). Child B, D, and E were judged to have odd correlate most highly with rhythmic, jerky, and intelligibility. Child A, F, and G were judged to have odd correlate most highly with intonation, jerky, and intelligibility. Child C had the greatest number and greatest strength of correlations between odd and the remaining descriptors; jerky ($r = .691, p < .001$), intonation ($r = .643, p < .001$), intelligibility ($r = .618, p < .001$), and sing-song ($r = .607, p < .001$). Child A, F, and G did have correlations between rhythmic and odd and Child B, D, and E also had high correlations between intonation and odd. These correlations were not as large as the relationships described above (see Appendix C).

Predicting the Quality of Oddness

A hierarchical multiple regression analysis was conducted to predict odd quality from the remaining 9 descriptors (see Table 6). The regressions were calculated using the mean values from the two raters. Sing-song and monotone were chosen as the first set of predictors because they were the two descriptors of most interest. The results of this analysis indicated that sing-song and monotone account for a significant proportion of the odd descriptor, $R^2 = .17, F(2, 839) = 85.23, p < .001$ indicating that children with higher ratings of sing-song and monotone tended to have higher ratings of odd. A second analysis was conducted to evaluate whether the remaining descriptors (intonation, modulation, rhythmic, jerky, halting, fluent, intelligibility) predicted the odd perception above and beyond sing-song and monotone. The remaining descriptors also accounted for a significant proportion of odd, *R^2 change = .49, $F(7, 832) = 174.44$, $p < .001$*. When the remaining descriptors were evaluated, sing-song was no longer a significant predictor of oddness. Intelligibility was a strong predictor of oddness; although, it was determined that 'intelligibility' is not a true area of prosody. Therefore, a separate regression was conducted which excluded intelligibility from the analysis (see Table 7).

Predicting Sing-Song

Two multiple regression analyses were conducted to determine predictors of sing-song and monotone. The first analysis included intonation and rhythm as the predictors for sing-song (see Table 8). The regression equation was significant, $R^2 = .408$, $F(2, 839) = 288.92$, $p < .001$. The individual regression coefficients revealed that intonation is a significant predictor of sing-song, $t(839) = 24.00$, $p < .001$; although rhythmic does not significantly predict sing-song, $t(839) = -1.823$, $p > .05$.

Predicting Monotone

The second analysis included intonation and rhythm as the predictors for monotone (see Table 9). The regression equation was also significant, $R^2 = .308$, $F(2, 839) = 186.415$. The individual regression coefficients revealed that rhythmic is a significant predictor of monotone, $t(839) = 14.512$, $p < .001$; although, the absence of intonation predicts monotone, $t(839) = -14.510$, $p < .001$.

Table 6

Prediction of Oddness from 2 Sets of Descriptors

Model	Descriptors	*B*	p-value	R	R^2	R^2 change	F
Set 1				.411	.169	.169	85.233***
	Sing-Song	.384	.000				
	Monotone	.354	.000				
Set 2				.814	.663	.494	174.44***
	Sing-Song	.042	.122				
	Monotone	.089	.001				
	Intonation	.193	.000				
	Modulation	.188	.000				
	Rhythmic	.281	.000				
	Jerky	.239	.000				
	Halting	.100	.000				
	Fluent	-.084	.000				
	Intelligibility	.348	.000				

Table 7

Prediction of Oddness excluding Intelligibility (N = 842)

Descriptors	β	T	p-value	F
Sing-Song	.082	2.631	.009	126.792***
Monotone	.121	4.063	.000	
Intonation	.234	7.073	.000	
Modulation	.160	6.347	.000	
Rhythmic	.308	10.857	.000	
Jerky	.257	8.306	.000	
Halting	.087	3.037	.002	
Fluent	-.102	-4.089	.000	

Table 8

Prediction of Sing-Song

Descriptors	β	p-value	R	R^2	R^2 change	F
			.639	.408	.408	288.921***
Intonation	.643	.000				
Rhythm	-.049	.000				

Table 9

Prediction of Monotone

Descriptors	β	p-value	R	R^2	R^2 change	F
			.555	.308	.308	186.415***
Intonation	-.420	.000				
Rhythm	.420	.000				

Pitch Analysis

Analyses of subjective ratings and acoustics were conducted as soon as the raters had completed their portion and the three phase pitch analysis had been completed. In these initial calculations, the thought groups were divided into groups based on the sums of the ratings from both raters (see Chapter 3). These group divisions were conducted for the individual child's thought groups to include a representative sample from every child. After the subjective ratings were obtained, reliability was calculated for the total group of children and each rater presented with what appeared to be an independent scheme for rating specific descriptors; therefore, many of the analyses conducted after this point analyzed each rater independently. Rater 2 was chosen as the primary rater because this rater presented with less of a correlation between monotone and sing-song suggesting that she was using a more independent scheme for determining sing-song and monotone. Furthermore, if there was a difference between the two prosodic descriptors, rater B's data would be more likely to find the true difference as well as exploring what makes thought groups encompass both sing-song and monotone.

The variables Fo SD, Fo mean, duration, slope of trend, and de-trended Fo SD (DTSD) were analyzed for the groups: sing-song, monotone, and both. The variable of most interest was the DTSD. The de-trended value removes irrelevant sources of variability that often are contaminants of SD measurements; specifically, the measurements account for differences in Fo mean and the slope of the trend line. The DTSD values, the Fo SD values and the duration values were converted to a logarithmic scale which provided a more normal distribution that is more valid for conducting statistical analyses.

An ANOVA of the log of DTSD found that all three groups were indistinguishable, $F(2, 64) = 2.949, p = .060$ (see Figure 1). A Tukey's HSD post hoc test revealed a statistically significant difference between sing-song (M = 3.585) and monotone $(M = 3.084), p < .05$. The results suggest that there is significantly greater Fo variation in sing-song quality regardless of overall declination of the speech segment or the Fo mean of the speaker. The results indicate that the monotone and sing-song groups are more significantly different from one another than they are different from speech where both qualities are present.

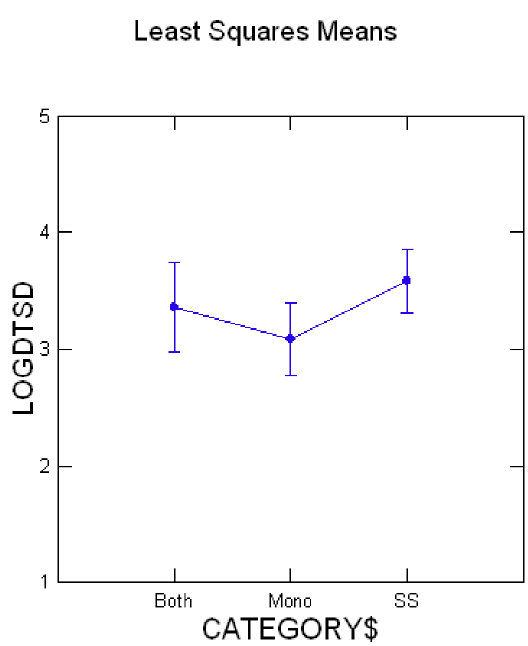

Figure 1. Logarithmic DTSD for All Groups *(N = 67)*

ANOVA revealed significant differences in Fo mean *F(2, 64) = 4.392, p < .05* and Fo SD *F(2, 64) = 3.574, p < .05* among the three groups. A Tukey's HSD post hoc test of Fo mean revealed a significant difference between sing song and monotone, but not between both and either sing-song or monotone, p < .05 (see Figure 2). A Tukey's HSD post hoc test of Fo SD similarly revealed a significant difference between sing-song and monotone only, p < .05 (see Figure 3). These post hoc analyses determined that sing-song quality is significantly higher in Fo with greater Fo variation across the thought group. ANOVA showed an insignificant difference between the three groups in terms of duration *F(2,64) = .634, p > .05* and slope *F(2,53) = 1.333, p > .05*.

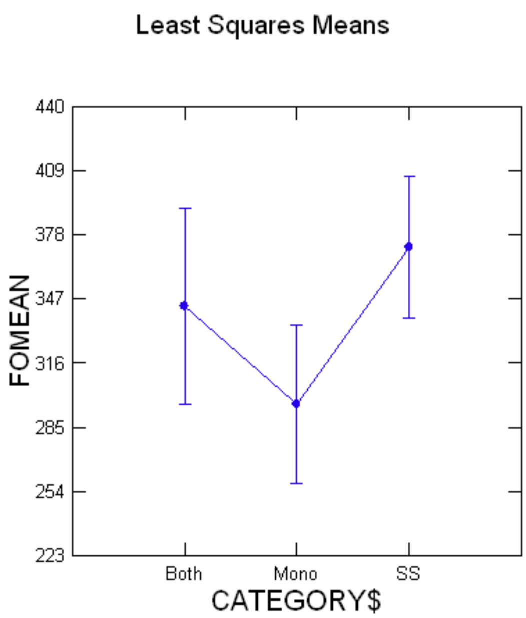

Figure 2. Significant Difference in Fo Mean Among the Three Groups *(N = 67)*

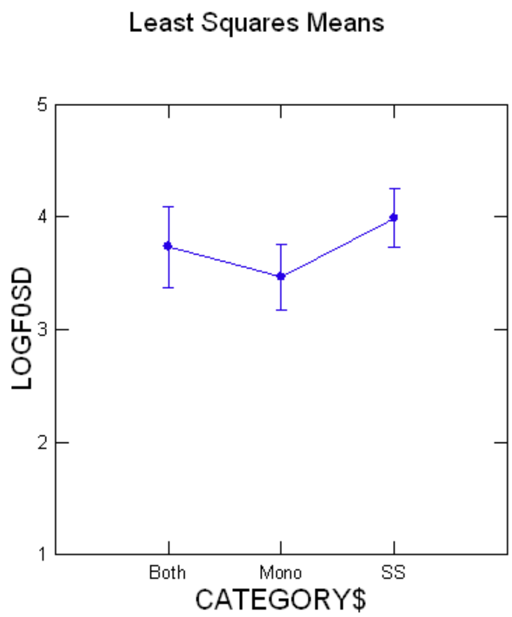

Figure 3. Significant Difference in FoSD Among the Three Groups *(N = 67)*

Cycle Durations

In the subjective analysis, atypical rhythm was found to be highly correlated with monotone but found to have no correlation with sing-song (see Table 10). A post-hoc analysis was conducted to analyze the rhythmic characteristics of the groups: both, sing-song and monotone. Cycles of pitch modulation were measured based on the presence of rhythmic fluctuation within the thought group as demonstrated by repetitive pitch contour peaks and valleys. Three sample thought groups within each of the three groups of interest were found to have two or more modulation cycles. Table 9 includes the four measurements obtained from each group: mean cycle duration of each thought group, SD of cycle duration within the thought group, mean cycle duration within each group, and mean SD within each group.

Table 10

Characteristics of Cycle Modulation

Groups	Thought Group	Mean Cycle Duration	SD Cycle Duration	Mean	
				Duration	SD
Both					
	1	537	46.67		
	2	592	105.88		
	3	764	275.06	631	142.54
Sing-Song					
	1	599	98.75		
	2	855	186.68		
	3	955	414.69	803	233.37
Monotone					
	1	343	24.04		
	2	1208	55.34		
	3	391	161.22	647	80.20

Duration values in milliseconds (ms)

The finding of greatest interest is the mean of the SD within each group. The monotone group presents with the smallest SD value, the both group presents with the middle SD value, and the sing-song group presents with the largest SD value. It appears that the quality of monotone may be characterized by speech that is constrained to strict cycle duration with little room for variation; whereas, the quality of sing-song may be characterized by greater fluctuation in cycle duration. As expected, when sing-song and monotone co-occur there is greater cycle duration variability than monotone alone but less cycle duration variability that sing-song alone.

Chapter 5

Discussion

This study analyzed prosodic descriptors that have characterized the speech of ASD in both research literature and diagnostic tests. Next, the descriptors were analyzed to determine which were correlated with one another. Then, this study investigated which descriptors were most predictive of an overall "odd" speech quality. Lastly, a pitch analysis was conducted to objectify the two most commonly used descriptors – sing-song and monotone – and to determine whether these two descriptors are similar or dissimilar in their acoustic parameters.

This study expands on previous research in several ways. First, a younger group of children were analyzed in this study which provides more information regarding the development of prosody within the ASD population. Previous research in young children with ASD has tended to focus on quantity of vocalizations, function of vocalizations, as well as vocal and phonemic quality of vocalizations rather than suprasegmental quality. Second, the prosody analysis was not based solely on subjective impressions as has been the case in many previous studies. Rather, both acoustic analysis techniques and subjective impressions were used to evaluate the verbal communication of children with ASD. The subjective ratings considered 12 prosodic descriptors that have not been assessed collectively in previous studies. Not only has previous research neglected to investigate these descriptors, but there are no studies comparing naturalistic subjective impressions to acoustic analyses of children with ASD. Furthermore, the primary researcher could have developed a coding system that would train listeners how to rate certain features. This type of coding system would be expected to achieve good reliability and would be less naturalistic in nature. However, the goal of this particular study was to determine naturalistic, everyday listener's perceptions rather than trained researchers' judgments. This study contributes information to discussion of whether prosodic differences are interrelated (Paul, Augustyn, et al. 2005) or independent (Paul, Shriberg, et al. 2005). Third, a natural speech sample was used rather than imitation as various authors have suggested the importance of eliciting natural speech when assessing prosody (Diehl et al., 2009; Shriberg, et al., 2001; Paul, Shriberg, et al. 2005). This rating scale could have been used for speech that was segmented differently (breath groups, single words, entire speech sample) which may have been more standardized and less naturalistic.

It was hypothesized that the group of children would be perceived as atypical to some degree on each descriptor. The results support this hypothesis. The group of children had an average rating above 1 on each of the 10 descriptors that were analyzed. The children received the most extreme ratings in the odd quality of their thought groups and least extreme in regards to atypical modulation and atypical fluency of their speech (although atypical judgments were still present in these two descriptors). This result suggests that ASD is characterized by unnatural sounding speech, as proposed by Asperger (1952/1991), due to the contribution of a variety of atypical qualities that are realized within a single child. Furthermore, although all of the children in the current study were judged to speak in varying degrees of "oddness", none of the children were rated as typical or completely lacking oddness. Each child was characterized by differing descriptors which supports other studies that have suggested there is not an individual aspect of prosody that is disordered in every child with ASD (Diehl et al., 2009; McCann et al., 2007; Peppe et al., 2007; Shriberg, et al., 2001). It seems that the autism spectrum is characterized by a spectrum of prosody deficits. Furthermore, Child G was rated as the oddest overall and also had the greatest number of high correlations between the descriptors; these ratings suggest that the greater number of mutually occurring prosodic deficits within the child will likely result in the perception of highly unnatural speech.

Relationships Among Prosodic Descriptors

It was also hypothesized that the children with ASD would produce thought groups that were characterized by more than one descriptor. In support of this hypothesis, a great number of large correlations were found among the descriptors within individual thought groups. However, it is possible that the descriptors could exist independently of one another. As expected, sing-song and monotone were negatively correlated. However, sing-song and monotone were also found to co-occur as evidenced by the thought groups selected for the "both" category in the pitch analysis. Thought groups that were described as sing-song also were described as having atypical intonation. This supports the acoustic findings of Schoen et al., (2010) who found that children with ASD produced irregular rise-fall contours (i.e., atypical intonation) and had greater pitch fluctuation (i.e., sing-song). However, thought groups that were monotone were described as missing intonation. These results are fairly intuitive; the greater the perceived pitch change at the end of a thought group, the more sing-song the child will appear. The less the perceived pitch change at the end of a thought group, the more monotone the child will appear. Interesting and less often reported, monotone speech is also atypical in terms of rhythm.

These results could be applied to recent studies that have reported similar results. For example, the children in the Peppe and colleagues (2007) study who sounded “questioning” when a statement was required (i.e., atypical intonation) would sound more sing-song than monotone. Similarly, Child A, E, and F were determined to have intonation correlate most with odd. The children in the Peppe and colleagues (2007) study who failed to make appropriate prosodic breaks (i.e., atypical rhythmic) would sound more monotone than sing-song. Similarly, Child B, D, and E were determined to have rhythm correlate most with odd. Both groups of children in the Peppe and colleagues (2007) study would likely both sound odd, but for different reasons. Although sing-song and monotone can simultaneously occur within a thought group, they rarely do. Each child presented with an average of 3 thought groups that were rated as highly monotone and highly sing-song. The study conducted by Peppe and colleagues (2007) did not describe the characteristics of the individual children so it is unclear whether the characteristics they described occurred simultaneously or independently.

Oddness

It was hypothesized that thought groups rated as highly odd would also be characterized as atypical on the remaining descriptors. The results showed that within the group of children, oddness was significantly correlated with all of the descriptors. This suggests that the unnatural sounding speech in children with ASD is not only related to the presence of dysfluencies, unintelligibility, abnormal rhythm, abnormal pausing, presence of fluctuation in the voice, rising or falling of pitch, etc; instead, the unnatural quality of speech in ASD is likely related to a combination of all of these features. Moreover, oddness was consistently largely correlated with high jerky and high unintelligibility ratings. These results support previous studies that have suggested the timing and stress of speech in ASD are large contributors to the odd quality (Paul, Augustyn, et al. 2005; Shriberg et al. 2001). Additionally, highly unintelligible speech in children from 3 – 5 years of age would sound quite unnatural.

The regression analysis initially suggested that sing-song and monotone were strong predictors of oddness; however, once the remaining descriptors were added to the regression, sing-song was no longer a strong predictor. It appears that sing-song and intonation may be so largely correlated that intonation is more truly the predictor of oddness rather than sing-song.

Sing-Song: Intonation, Pitch, and Cycle Durations

It makes sense that the pitch analysis would find a greater fluctuation in pitch to be most representative of the thought groups that were rated as sing-song and not monotone. Even when overall declination (i.e., slope) was controlled, sing-song quality presented with greater pitch fluctuation than monotone. In other words, atypical intonation (rises or falls) was not solely responsible for the sing-song quality of speech in the children with ASD because when its contribution was removed, there was still a significant difference between sing-song and monotone in terms of pitch fluctuation. However, when the overall intonation contour was present for the pitch analysis, the sing-song quality was easier to differentiate from the monotone quality. This confirms the raters' judgments that atypical intonation is the characteristic that uniquely contributes to sing-song.

Initially, it seemed possible that the overall slope of the thought group could contribute more to sing-song quality than Fo variation (see Chapter 2), however the results suggest that greater Fo variation most clearly separates sing-song from monotone. The high Fo variation within the ASD population is supported by a number of previous studies (Diehl et al., 2009; Schoen et al., 2010; Hubbard & Trauner, 2007). However, the current study is the first that associates this high pitch fluctuation with a subjective description of speech in children with ASD.

The increased SD of cycle duration in sing-song quality is a new finding that has not previously been reported. Hubbard and Trauner (2007) described atypical locations of maximum pitch *within* a phrase which is similar to the definition of "cycles" in the current study. Hubbard and Trauner's descriptions may lend support for this rhythmic or jerky quality. In conclusion, although the depth of erratic pitch fluctuation contributes significantly to sing-song quality, it appears that the width of this pitch fluctuation may also be a contributor.

Monotone: Rhythm, Pitch, and Cycle Durations

The strong correlation between monotone and rhythm led to an investigation of the differences in rhythm between sing-song and monotone. Previous research has used adjectives that appear to encompass monotone and rhythmic quality. For example, Peppe and colleagues (2007) used dull, robotic, and wooden to describe the speech of persons with ASD. Kanner (1943) described children who produced short, staccato, forceful sounds and were mechanical-sounding in their speech. These adjectives may be attempts at describing speech with very specific, strict cycle durations as was found to characterize monotone in the current acoustic analyses. The smaller SD of cycle duration in monotone corroborates the idea that rhythm is a significant predictor of monotone. The results of the Grossman and colleagues (2010) study support the notion that rhythmic qualities of speech may be the main contributing prosodic deviance in the speech of children with ASD. The restricted range of pitch fluctuation may also contribute to this "robotic" quality so often associated with monotone. The pitch results support the subjective analyses where the absence of intonation was found to be a predictor of monotone.

The results suggest that a child might use sing-song speech when his or her intonation parameters are fluctuating with less rigid boundaries (higher pitch SD, higher cycle duration SD) whereas this child might use monotone speech when his or her intonation parameters are more rigid and strict (smaller pitch SD, smaller cycle duration SD). It is interesting that Child G was rated as the oddest overall and was also the only child to have average ratings above 2 for both sing-song and monotone. Although, sing-song and monotone are not often both highly characteristic of the same child, their dual presence may increase the perception of oddness.

Interdependence or Independence of Prosodic Components

The results appear to add to the confusion regarding the independence or interdependence of deviant prosodic components (Paul, Augustyn, et al., 2005; Paul, Shriberg, et al. 2005). Although all of the characteristics were judged to be atypical in some degree within the group of children, the greatest prosodic deficit was not consistently the same from child to child. Moreover, several descriptors were consistently atypical on the same thought groups such as intonation and sing-song. This suggests that deficits in various domains of prosody tend to co-occur. However, it is unclear whether intonation and sing-song should be considered different prosodic components as they may describe the same prosodic deficit. In addition, certain prosodic characteristics tend to predict the degree of unnaturalness in conversation. For example, a child that is perceived as sing-song or monotone is likely to also be described as odd (see Chapter 4,

Table 6). However, this is not always the case as child B had one of the highest mean values of sing-song and one of the lowest ratings of overall odd (see Chapter 4, Table 5).

Clinical Implications

Understanding unusual prosody in children with ASD has important implications for assessment, treatment, and ultimately social acceptance in persons on the autism spectrum. Words such as “sing-song”, “monotone”, or “jerky” may be used clinically with the assumption that therapists intuitively understand them.

Assessment. Unusual prosody is a diagnostic characteristic of ASD according to the DSM-IV (APA, 2000); furthermore, the ADOS (Lord et al., 2000) and ADI-R (Le Cauteur et al., 2003) use descriptions of atypical prosody to diagnose ASD. Unfortunately, these diagnostic measures are not yet sensitive to the specific prosodic deficits in the ASD population. Furthermore, the proposed revisions of the DSM-V do not include a criterion related to verbal output of children when diagnosing ASD. This study exemplifies explicit areas of atypical prosody that may be present in children with ASD who do not have significant language deficits. Awareness of these prosodic characteristics in ASD is essential because children with ASD and typical language development may still present with social difficulties secondary to unnatural sounding speech.

It would be useful to provide a rating scale similar to the scale developed for this study to a parent, grandparent, or teacher to rate his or her perceptions of the child’s speech. This type of rating scale could supplement diagnostic tests that are currently used. Other areas within the field of communication sciences and disorders are using similar rating scales such as the Consensus Auditory-Perceptual Evaluation of Voice

(CAPE-V; Kempster, Gerratt, Abbott, Barkmeier-Kraemer, & Hillman, 2009), which is used to subjectively evaluate voice quality. The resulting information would provide helpful information in determining which areas of prosody must be targeted to increase natural sounding speech and social acceptance.

The acoustic measurements echo previous studies that have recommended screening children with ASD for prosody involvement (Shriberg et al., 2001). The relationship between the perception of monotone and the acoustic measurements of rhythm as well as the relationship between the perception of sing-song and the acoustic measurements of pitch fluctuation suggest that our perceptions of these qualities are fairly reliable at predicting acoustic involvement. This is important in assessment as it suggests that the diagnostician's or parent's perceptions of monotone or sing-song will likely occur when there truly are atypical acoustic qualities.

Treatment. The overall goal of speech and language therapy for children with ASD who are verbal should be to improve the naturalness of their communication. Children with ASD are typically receiving intensive language therapy which often includes "social interaction" training. Perhaps the focus of intervention should be on increasing natural interaction via explicit instruction in accepted prosody manipulation in addition to standard language therapy. There are several considerations the therapist should make before implementing a treatment plan to address prosodic differences in children with ASD.

First, the clinician should be aware of which area of prosody is the most atypical in the child with ASD. For example, if the child is judged to use extremely atypical prosody within a number of prosodic domains, then a metalinguistic approach should be employed to create an awareness of prosody and to provide auditory bombardment of accepted prosodic conventions (e.g. questions, statements, appropriate stress, appropriate pausing, etc.). Paul, Augustyn, and colleagues (2005) proposed a similar approach that could be used with school-age children and adolescents which includes explicit instruction to teach the purpose of prosody.

A modification to the Cycles approach for children with phonological disorders (Hodson & Paden, 1991) could be used for children with multiple deficits in prosody. This approach might consist of auditory bombardment, production practice, and developing semantic awareness of prosodic contrasts.

However, if the child's assessment indicated a greater deficit in one area of prosody then that specific domain should be trained to improve the child's naturalness in speech. For example, Child E in this study was found to have slightly atypical prosodic differences in a variety of areas assessed on the subjective analysis; however, rhythmic and jerky descriptors were judged to be the most atypical for this child. These descriptors were both defined using words like bumpy, robotic and mechanical. Therefore, therapy that focused on improving pitch control would be less helpful for this child than therapy that focused on improving the smooth, melody of speech. It is noteworthy that one of the descriptors most consistently correlated with odd was intelligibility. Although this is not a true prosodic quality, it was a large contributor to the odd characteristic of the children's speech within this study. Traditional articulation therapy is not often reported as an approach used with children with ASD. However, if the noted unintelligibility is related to articulation deficits, traditional articulation therapy would be a reasonable first approach for a child with highly unnatural speech. It is evident that a thorough and accurate assessment of speech and prosody is essential before efficacious therapy can be implemented.

Second, the clinician should be aware of his or her personal use of prosody within the therapy environment. Gerken and McGregor (1998) explain that clinicians often use a "therapeutic prosody" (p. 45) that has a higher pitch with increased pitch variability and a slower rate of speech during their interactions with children with language disorders. Gerken and McGregor propose this therapeutic prosody is helpful in accenting the targets of therapy and maintaining the child's attention much like CDS functions to promote prosodic bootstrapping. However, the current findings suggest that pitch variability in ASD is a large contributor to their unnatural sounding speech. It seems that clinicians should avoid this strategy when accentuating therapeutic targets only if their clients are found to have already developed prosodic deviations. Moreover, CDS has been shown to be very effective in promoting language development in infants and toddlers and this study does not suggest otherwise.

Third, clinicians working with young children with language disorders tend to use a great deal of telegraphic speech to facilitate quick language learning of content words. The use of grammatically incorrect language structure to facilitate language development in children with language disorders seems contradictory in all circumstances (Bedore & Leonard, 1995; Gerken & McGregor, 1998); moreover, the presence of abnormal prosody in children with ASD suggests that telegraphic speech should be used sparingly, if not at all, in this population. When function words are eliminated in speech, the natural strong-weak pattern of speech is also eliminated. It seems logical that this could result in a different understanding of rhythm in speech production. Perhaps “prosodic recasts” or expansions should be used to facilitate language and prosody in children with ASD.

Social acceptance. Pragmatics and social interaction are often described as the most noticeable deficits in persons with ASD. Shriberg and colleagues (2001) suggested that even inconsistent voice and prosody differences can affect the listener’s perceptions of the speaker’s attractiveness and affect. Paul, Shriberg, and colleagues (2005) suggested an association between deficits of prosody and perceptions of social/communicative competence. The communication partners of people with ASD, who present with intonation deficits, may find it difficult to perceive the communicative intent (questions vs. statements) of these speakers in conversation. It may also be difficult to perceive the affect of a statement when the person uses inconsistent or atypical rhythm. These conversational difficulties are similar to the descriptions of pragmatic and affective prosody deficits in people with ASD (Grossman, et al., 2010; Paul et al., 2009; Peppe et al., 2007; Shriberg et al., 2001). Impairment in social skills is the hallmark of ASD and the affect the prosody has on social interaction should be forefront in the minds’ of clinicians when making decisions regarding assessment and intervention.

Limitations

The limitations of this study should be considered when interpreting these findings. This study did not collect data for age-matched, typically developing peers or persons with developmental delay not ASD. Therefore, this study can merely describe the characteristics that are judged to be present in children with ASD. Fortunately, the two raters had been exposed to typical language development and were made aware that the rating of 1 should be provided when the thought group sounded similarly to a typically developing child.

Furthermore, this study did not differentiate degrees of the autism spectrum. Paul and colleagues (2009) suggested that prosody may vary from HFA to AS and Diehl and colleagues (2009) suggested that prosody differences correlated with ADOS scores; this should be considered in future studies. However, Diehl and colleagues (2009) found this correlation to exist only in adolescents and adults with ASD and not in school-age children with ASD. Therefore, the severity of autism may be less significant in studies which focus on achieving an earlier diagnosis of ASD.

This study also included a small sample size of children which limits the power of the results. For example, the sample of thought groups which included more than two measureable cycles of Fo variation was limited to nine in this study; for this specific analysis, there was not enough data to draw significant conclusions.

The rating scale used in this study defined a rating of 1 as typical but it is likely that typically developing children would also receive ratings above 1 if they were rated on the same scale. Therefore, the interpretation of atypical qualities in these children can only be appreciated relatively to one another, rather than in comparison to typically developing children.

Lastly, although the thought group unit of analysis worked well for establishing a naturalistic listening environment for the raters, the thought groups are likely less efficient for measuring the trend line across a phrase. This could account for the insignificant relationship between monotone, sing-song and both groups as more than one phrase was often included within a thought group.

Future Research

A rating scale similar to the scale used in this study should be implemented in future studies to compare children with ASD to typically developing children and children with DD. Future studies should implement research designs to acoustically investigate rhythm and duration. This study has demonstrated that rhythmic variation may be particularly salient and meaningful in the ASD population; however, a very small number of samples were considered for analysis and the decisions of how to mark these "rhythmic cycles" were more intuitive to the primary researcher rather than objectively defined. Furthermore, duration should be analyzed in words or word approximations as Schoen et al. (2010) found significantly longer word durations in the ASD population.

Future research should consider the importance of carefully defining terms that are used to characterize children with ASD. Studies should focus on developing and standardizing descriptions of terms that can be used consistently and translated from researcher to therapist. Furthermore, the prosody of typically developing children and developmentally delayed children should be compared to the prosody of children with ASD to ascertain whether these atypical productions are unique to ASD.

Although research investigating the nature of prosody differences in children with ASD is certainly warranted, research focusing on methods of treating these deficits is equally necessary. There is a lack of knowledge regarding the impact and efficacy of metalinguistic activities focused on prosody for children with ASD. Future research should implement intervention to address specific prosodic deficits

Conclusions

Pragmatic language intervention is typically the focus of therapy for children with ASD. Unusual prosody is listed as a diagnostic marker of ASD but has barely been investigated until recently. This study contributes to the knowledge regarding our perceptions of the prosody in children with ASD and the acoustics of speech in children with ASD. Although prosody may seem less essential as a therapy target than other areas in the ASD population, it has been shown to significantly affect the naturalness of their speech and the perception of social competence in these individuals. Specifically, this study suggests that atypical rhythmic and intonation qualities tend to result in monotone and sing-song perceptions, respectively. These perceptions lead to an appreciation of odd sounding speech within the ASD population. Therapists are trained to teach children ways of communicating that are more natural and socially acceptable. This study provides specific areas for which deficits may be recognized and from which therapy tasks may be developed.

References

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

American Psychiatric Association [APA]. (2010). *DSM-5 Development*. Arlington, VA: Author. Retrieved from http://www.dsm5.org/ProposedRevisions/Pages/proposedrevision.aspx?rid=94

Asperger, H. (1952/1991). "Autistic psychopathology" in childhood. In U. Frith (Ed. And Trans.), *Autism and Asperger syndrome* (pp. 37–92). Cambridge, UK: Cambridge University Press.

Bedore, L. M., & Leonard, L. B. (1995). Prosodic and syntactic bootstrapping and their applications: A tutorial. *American Journal of Speech-Language Pathology, 4(1),* 66-72.

Buder, E.H., Chorna, L.B., Oller, D.K., & Robinson, R.B. (2008). Vibratory regime classification of infant phonation. *Journal of Voice, 22(5)*, 553-564.

Center for Disease Control and Prevention. (2010). Autism Spectrum Disorders. Retrieved from http://www.cdc.gov/ncbddd/autism/index.html

Chiang, C., Soong, W., Lin, T., & Rogers, S. (2008). Nonverbal communication skills in young children with autism. *Journal of Autism and Developmental Disorders, 38,* 1898-1906. doi: 10.1007/s10803-008-0586-2

Crawford, N., Edelson, L., Skwerer, D. P., & Tager-Flusberg, H. (2008). Expressive language style among adolescents and adults with Williams syndrome. *Applied Psycholinguistics, 29*, 585-602.

Delgado, R. (1996). Action Analysis coding and Training [Computer software]. Miami, FL: Intelligent Hearing Systems.

Diehl, J., Benneto, L., Watson, D., Gunlogson, C., & McDonough, J. (2008). Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and Language, 106*, 144-152.

Diehl, J., Watson, D., Bennetto, L., McDonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics*, *30*, 385-404.

Esposito, G., & Venuti, P. (2009). Comparative analysis of crying in children with autism, developmental delays, and typical development. *Focus on Autism and Other Developmental Disabilities*, *24(4),* 240-247.

Fernald, A. (1989). Intonation and communicative intent in mother's speech to infants: Is the melody the message? *Child Development*, *60,* 1497-1510.

Fine, J., Bartolucci, G., Ginsberg, G., & Szatmari, P. (1991). The use of intonation to communicate in pervasive developmental disorders. *Journal of Child Psychology, 32(5)*, 771-781

Furrow, D. (1984). Young children's use of prosody. *Journal of Child Language, 11,* 203-213.

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

Gottman, J. (1983). How children become friends. *Monographs of the Society for Research in Child Development, 48 (3),* 1-86.

Grossman, R., Bemis, R., Skwerer, D., & Tager-Flusberg, H. (2010). Lexical and affective prosody in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research, 53*, 778-793.

Halle, P. A., de Boysson-Bardies, B., & Vihman, M. M., (1991). Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech, 34(4),* 299-318.

Hodson, B.W., & Paden, E.P. (1991). Targeting Intelligible Speech: A Phonological Approach to remediation (2nd Ed). Austin, TX: Pro-Ed.

Hubbard, K., & Trauner, D. A. (2007). Intonation and emotion in Autistic Spectrum Disorders. *Journal of Psycholinguistic Research, 36*, 159-173.

Jusczyk, P. W. (2002). Language development: From speech perception to first words. In A. Slater & M. Lewis (Eds.). *Introduction to infant development* (pp. 147-164). New York: NY: Oxford University Press.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2*, 217-250.

Kempster, G., Gerratt, B., Abbott, K., Barkmeier-Kraemer, J., & Hillman. R. (2009). Consensus Auditory-Perceptual Evaluation of Voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*, 124-132.

Koegel, L., Koegel, R., Green-Hopkins, I., & Barnes, C. (2010). Brief report: Question asking and collateral language acquisition in children with autism. *Journal of Autism and Developmental Disorders*, *40*, 509-515.

Le Cauteur, A., Lord, C., & Rutter, M. (2003). Autism Diagnostic Interview – Revised (ADI-R). Los Angelos, CA: Western Psychological Services.

Litras, S., Moore, D. W., & Anderson, A. (2010). Using video self-modelled social stories to teach social skills to a young child with autism. *Autism Research and Treatment*, 2010, 1-9. doi: 10.1155/2010/834979

Lord, C., Risi, S., Lambrecht, L., Cook, E. Leventhal, B., DiLavore, P., et al. (2000). The Autism Diagnostic Observation Schedule- Generic: A standard measure of social and communication deficits associated with spectrum disorders. *Journal of Autism and Childhood Schizophrenia*, *30(3)*, 205-223. doi:10.1023/A:1005592401947

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). Autism diagnostic observation schedule-generic. Los Angelos, CA: Western Psychological Services.

Loveland, K. A., Landry, S. H., Hughes, S. O., Hall, S. K., & McEvoy, R. E. (1988). Speech acts and the pragmatic deficits of autism. *Journal of Speech and Hearing Research, 31,* 593-604.

McCann, J., Peppe, S., Gibbon, F., 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(6),* 682-702.

Milenkovic, P. (2001). TF32 [Computer software]. Madison, WI: University of Wisconsin-Madison.

Mullen, E. (1995). *The Mullen Scales of Early Learning.* Circle Pines, MN. American Guidance.

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J. (2003). *Early Social Communication Scales.* University of Miami. Coral Gables, FL.

Paul, R. (2005). Assessing communication in autism spectrum disorders, In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of Autism and Pervasive Developmental Disorders: Assessment, interventions and policy* (3rd ed., Vol. 2, pp. 799-816). Hoboken, NJ: John Wiley & Sons.

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. (2005). Perception and production of prosody by speakers of autism spectrum disorders. *Journal of Autism and Developmental Disorders, 35(2),* 205-220.

Paul, R., Shriberg, L. D., McSweeny, J., Cicchetti, D., Klin, A., & Volkmar, F. (2005). Brief report: Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 35(6*), 861-869.

Paul, R., Orlovski, S. M., Marcinko, H. C., & Volkmar, F. (2009). Conversational behaviors in youth with high-functioning ASD and Asperger syndrome. *Journal of Autism and Developmental Disorders, 39*, 115-125.

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

Peppe, 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,* 1776-1791.

Peppe, 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.

Plexico, L., Cleary, J., McAlpine, A., & Plumb, A. (2010). Disfluency characteristics observed in young children with autism spectrum disorders: A preliminary report. *Perspectives on Fluency and Fluency Disorders, 20,* 42-50.

Rutter, M., Bailey, A., Berument, S., Lord, C., & Pickles, A. (2001). Social communication questionnaire research edition. Los Angeles, CA: Western Psychological Services.

Schoen, E., Paul, R., Berkovits, L. & Volkmar, F. (2010, May). Prosody in school-age children with autism spectrum disorder. *9th Annual International Meeting for Autism Research.* Philadelphia, PA.

Schoen, E., Paul, R. & Chawarska, K. (2010). In R. Paul & P. Flipson, Jr. (Eds.), *Speech Sound Disorders in Children* (pp. 181-204). San Diego, CA: Plural Publishing.

Shaw, H., & Nadig, A. (2010, May). Do children with high-functioning autism appropriately mark contrastive stress in speech to a partner? *9th Annual International Meeting for Autism Research.* Philadelphia, PA.

Sheinkopf, S., Mundy, F., Oller, D., & Steffens, M. (2000). Vocal atypicalities of preverbal autistic children. *Journal of Autism and Developmental Disorders, 30(4),* 345-354.

Shriberg, L. D., & Kent, R. (2003). *Clinical Phonetics (3rd ed.).* Boston: Pearson Education, Inc.

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., Klin, A.,Cohen, D., & Volkmar, F. (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.

Shumway, S., & Wetherby, A. M. (2009). Communicative acts of children with autism spectrum disorders in the second year of life. *Journal of Speech, Language, and Hearing Research, 52*, 1139-1156.

Snow, D. (2006, March/April). Regression and reorganization of intonation between 6 and 23 months. *Child Development, 77(2),* 281-296.

Snow, D. (2007). Polysyllabic units in the vocalizations of children from 0:6 and 1:11: Intonation-groups, tones and rhythms. *Journal of Child Language, 34,* 765-797.

Snow, D., & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua, 112*, 1025-1058.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). Vineland Adaptive Behavior Scales. Circle Pines, MN: American Guidance.

Stone, W., Ousley, O. Y., Yoder, P. J., Hogan, K. L., & Hepburn, S. L. (1997). Nonverbal communication in two- and three-year old children with autism. *Journal of Autism and Developmental Disorders, 27,* 677-696.

Tager-Flusberg, H. (1999). A psychological approach to understanding the social and language impairments in autism. *International Review of Psychiatry, 11*, 325-334.

Wetherby, A. M., Cain, D. H., Yonclas, D. G. & Walker, V. G. (1988). Analysis of intentional communication of normal children from the prelinguistic to the multiword stage. *Journal of Speech and Hearing Research, 31,* 240-252.

Wetherby, A., & Prizant, B. (2002). Communication and symbolic behavior scales developmental profile—First Normed edition. Baltimore, MD: Paul H. Brookes.

Wetherby, A. M. & Prutting, C. A. (1984). Profiles of communicative and cognitive-social abilities in autistic children. *Journal of Speech and Hearing Research, 27*,

Wetherby, A. M., Watt, N., Morgan, L. & Shumway, S. (2007). Social communication profiles of children with autism spectrum disorders late in the second year of life. *Journal of Autism and Developmental Disorders, 37,* 960-975. doi: 10.1007/s10803-006-0237-4 364-377.

Wetherby, A. M., Woods, J., Allen, J., Cleary, J., Dickinson, H. & Lord, C. (2004). Early indicators of autism spectrum disorders in the second year of life. *Journal of Autism and Developmental Disorders, 34(5),* 473-493.

Wetherby, A., Yonclas, D., & Bryan, A., (1989). Communication profiles of preschool children with handicaps: Implications for early identification. *Journal of Speech and Hearing Disorders, 54*, 148-158.

Appendix A:

Rater Instructions

You will be listening to utterances of variable length and rating each utterance on several characteristics. These ratings are divided into 5 groups. You are permitted to play the utterance before beginning the initial ratings and replay it as you transition from one group of ratings to another. The utterances should not be played more than 5 times each. Each characteristic will be rated on a 7-point scale. Below, the characteristics are included in order of how they should be rated:

GROUP 1

1. Slow/Fast

   This 7-point scale will represent a continuum where 1 represents very slow, 4 represents typical rate of speech for a child from ages 3-5, and 7 represents very fast.

2. Soft/loud

   This 7-point scale will represent a continuum where 1 represents very soft, 4 represents typical loudness for a child from ages 3-5, and 7 represents very loud.

   Below is a description of the remaining rating scales. Each of these is rated on a continuum from typical speech (rating of 1) to extreme variance (rating of 7) as it relates to the respective characteristic. All intermittent rating levels should be used to describe increasing degrees of the specific characteristic.

GROUP 2

3. Sing-song

   This is a measurement of the melodic, patterned, fluctuating nature of speech. A rating of 1 represents a typical amount of fluctuation in the child's voice, and a rating of 7 represents when an extreme level of sing-song quality is present in the child's voice.

4. Monotone

   A measurement where a rating of 1 represents a typical amount of fluctuation in the child's voice, and a rating of 7 represent excessively dull, droning, or flat speech.

GROUP 3

5. Intonation

   A measurement of rising or falling pitch, where 1 represents typical movement/direction of pitch and 7 represents atypical rising and/or falling of pitch.

6. Modulation

   A measurement where a rating of 1 represents a typical, stable voice and a rating of 7 represents an extremely atypical shaky voice characterized as unstable, fluttery or trembling.

GROUP 4

- This group includes a “not applicable” (N/A) category for one-word utterances that cannot be described as either typical or atypical.

7. Rhythmic

   A measurement where a rating of 1 represents typical smooth and flowing speech and a rating of 7 represents extremely atypical robotic speech, characterized as mechanic, computerized speech.

8. Jerky

   A measurement of rhythmic and stress variability. A rating of 1 represents typical stress placement and a rating of 7 represents extreme variability in stress emphasis also characterized by bumpy speech and erratic fluctuation.

9. Halting

   A measurement of the number of spaces present in speech. A rating of 1 represents consistent speech with an even pace and a rating of 7 is used if there are an excessive number of gaps, pauses, or silent intervals.

GROUP 5

10. Fluent

A measurement of the stuttering-like dysfluencies present in speech. A rating of 1 is appropriate if the speech appears easy, flowing, and smooth; whereas, a rating of 7 is appropriate if the speech consists of repetitions, prolongations and/or blocks and is characteristic of severe to profound stuttering. Ratings between 1 and 7 should be used to describe utterances that are representative of mild to moderate stuttering.

11. Intelligibility

A measurement of how well the child is understood. A rating of 1 is representative of speech that is highly intelligible and a rating of 7 is appropriate if the child's articulation is extremely unintelligible (jumbled, indecipherable, or incoherent).

12. Odd

This is the final rating and should represent the rater's overall impression of the child's speech. A rating of 1 is appropriate if the overall quality of the speech sounds typical and a rating of 7 is appropriate if the overall quality of the speech sounds extremely atypical.

EXAMPLE:

PLAY UTTERANCE

Slow			Typical			Fast
1	2	3	4	5	6	7

Soft			Typical			Loud
1	2	3	4	5	6	7

PLAY UTTERANCE

SING-SONG						
Typical						Extreme
1	2	3	4	5	6	7

MONOTONE						
Typical						Extreme
1	2	3	4	5	6	7

PLAY UTTERANCE

INTONATION						
Typical						Extreme
1	2	3	4	5	6	7

MODULATION						
Typical						Extreme
1	2	3	4	5	6	7

PLAY UTTERANCE

RHYTHM

Typical						Extremely Atypical
1	2	3	4	5	6	7
N/A						

JERKY						
Typical						Extremely Jerky
1	2	3	4	5	6	7
N/A						

HALTING						
Typical						Extremely halting
1	2	3	4	5	6	7
N/A						

PLAY UTTERANCE

FLUENCY						
Typical						Extremely dysfluent
1	2	3	4	5	6	7

INTELLIGIBILITY						
Typical						Extremely Unintelligible
1	2	3	4	5	6	7

ODD						
Typical						Extremely Odd
1	2	3	4	5	6	7

Appendix B:

Rater 1 Correlations

Table B1

Rater 1: Correlations Among Descriptors for Total Group of Children (N = 1115)

	Monotone	Intonation	Modulation	Rhythmic[a]	Jerky[a]	Halting[a]	Fluent	Intelligibility	Odd
Sing-Song	-.403***	.648***	-.069*	-.047	-.042	.007	.073*	.154***	.234***
Monotone		-.489***	.151***	.276***	.249***	.224***	-.069*	.024	.201***
Intonation			.073*	.051	.046	.109**	.076*	.097**	.308***
Modulation				.268***	.226***	.271***	.083**	.027	.340***
Rhythmic					.328***	.353***	.083*	.067	.501***
Jerky						.549***	.221***	.036	.421***
Halting							.300**	.096**	.402***
Fluent								.033	.053
Intelligibility									.458***

*** p < .001 ** p < .01 * p < .05 [a] N = 842 (excluded one-word utterances)

Appendix C

Rater 2 Correlations

Table C1

Rater 2: Correlations Among Descriptors for Total Group of Children (N = 1115)

	Monotone	Intonation	Modulation	Rhythmic[a]	Jerky[a]	Halting[a]	Fluent	Intelligibility	Odd
Sing-Song	-.163***	.451***	.139***	.092**	.417***	.169***	.145***	.150***	.345***
Monotone		.004	.222***	.464***	-.008	.168***	-.013	.136***	.273***
Intonation			.070*	.235***	.421***	.235***	.041	.239***	.489***
Modulation				.232***	.245***	.094***	.096***	.029	.331***
Rhythmic					.374***	.146***	.062	.163***	.511***
Jerky						.331***	.156***	.172***	.613***
Halting							.152***	-.039	.330***
Fluent								-.012	.103**
Intelligibility									.501***

*** p < .001 ** p < .01 * p < .05 [a] N = 842 (excluded one-word utterances)

Appendix D:

Child A Correlations

Table D1

Correlations Among Descriptors for Child A (N = 96)

	Monotone	Intonation	Modulation	Rhythmic	Jerky	Halting	Fluent	Intelligibility	Odd
Sing-Song	-.320**	.692***	-.063	.147	.293**	.191	.024	.124	.458***
Monotone		-.253*	.342**	.486***	.185	.248*	.080	-.034	.119
Intonation			-.045	.161	.330**	.256*	.004	.078	.539***
Modulation				.197	.323**	.259*	.139	-.022	.283**
Rhythmic					.399***	.301**	.170	.220*	.484***
Jerky						.841**	.342**	.142	.552***
Halting							.301**	.049	.440***
Fluent								.096	.099
Intelligibility									.519***

*** p < .001
** p < .01
* p < .05

Appendix E

Child B Correlations

Table E1

Correlations Among Descriptors for Child B (N = 166)

	Monotone	Intonation	Modulation	Rhythmic[a]	Jerky[a]	Halting[a]	Fluent	Intelligibility	Odd
Sing-Song	-.350***	.412***	-.238**	.114	-.215**	-.098	.081	.071	.168*
Monotone		-.294***	.293***	.061	.530***	.222**	.107	.031	.213**
Intonation			-.123	.468***	-.010	.101	.081	.254**	.454***
Modulation				.053	.255**	.137	.031	-.029	.199*
Rhythmic					.287***	.158	-.067	-.019	.504***
Jerky						.487***	.249**	.179*	.512***
Halting							.254**	.187*	.394***
Fluent								.038	.123
Intelligibility									.569***

***p < .001
**p < .01
*p < .05

Appendix F

Child C Correlations

Table F1

Correlations Among Descriptors for Child C (N = 117)

	Monotone	Intonation	Modulation	Rhythmic[a]	Jerky[a]	Halting[a]	Fluent	Intelligibility	Odd
Sing-Song	-.147	.815***	.348***	.284**	.519***	.103	-.079	.319***	.607***
Monotone		-.025	.157	.279**	.059	,270**	.084	.033	.137
Intonation			.443***	.427***	.574***	.220*	-.038	.303**	.643***
Modulation				.710***	.495***	.598***	.232*	.253**	.567***
Rhythmic					.556***	.530***	.341***	.258**	.576***
Jerky						.432***	.365***	.318**	.691**
Halting							.474***	.031	.449***
Fluent								-.020	.106
Intelligibility									.618***

*** p < .001

** p < .01

* p < .05

[a] N = 106 (excluded one-word utterances)

Appendix G

Child D Correlations

Table G1

Correlations Among Descriptors for Child D (N=147)

	Monotone	Intonation	Modulation	Rhythmic[a]	Jerky[a]	Halting[a]	Fluent	Intelligibility	Odd
Sing-Song	-.346***	.787***	.181*	-.055	.339**	-.078	.305***	.230**	.419**
Monotone		-.451***	.101	.506***	.076	.352***	.014	.173*	.333***
Intonation			.133	-.126	.413***	.023	.081	.267**	.449***
Modulation				.161	-.148	-.010	.131	.087	.156
Rhythmic					.465***	.339**	.084	.340**	.571***
Jerky						.510***	.092	.181	.648***
Halting							.199	-.044	.447***
Fluent								.039	.141
Intelligibility									.655***

*** $p < .001$

** $p < .01$

* $p < .05$

[a] N = 96 (excluded one-word utterances)

Appendix H

Child E Correlations

Table H1

Correlations Among Descriptors for Child E (N = 130)

	Monotone	Intonation	Modulation	Rhythmic[a]	Jerky[a]	Halting[a]	Fluent	Intelligibility	Odd
Sing-Song	-0.288**	.592***	-.045	.124	.288**	.135	-.078	.045	.347***
Monotone		-.385	.403***	.536***	.018	.102	.120	.257**	.304***
Intonation			-.122	-.060	.297**	.247*	-.126	.026	.321***
Modulation				.260**	-.049	-.024	-.066	.159	.193*
Rhythmic					.468***	.366***	.117	.253**	.661***
Jerky						.362***	-.073	-.025	.574***
Halting							.146	-.091	.387***
Fluent								-.090	.012
Intelligibility									.464***

*** p < .001

** p < .01

* p < .05

[a] N = 106 (excluded one-word utterance)

Appendix I

Child F Correlations

Table I1

Correlations Among Descriptors for Child F (N = 150)

	Monotone	Intonation	Modulation	Rhythmic[a]	Jerky[a]	Halting[a]	Fluent	Intelligibility	Odd
Sing-Song	-.299***	.470***	-.024	-.002	.192*	.203*	.044	.124	.295***
Monotone		-.357***	- .071	.242**	- .016	.143	-.030	- .080	-.139
Intonation			.002	.147	.346***	.241**	.080	.166*	.465***
Modulation				.250**	.404***	.108	-.062	.081	.325***
Rhythmic					.456***	.290**	.105	.172	.375***
Jerky						.589***	.103	.189*	.561***
Halting							.181*	.066	.326***
Fluent								- .016	.044
Intelligibility									.658***

*** p < .001

** p < .01

* p < .05

[a] N = 126 (excluded one-word utterances)

Appendix J

Child G Correlations

Table J1

Correlations Among Descriptors for Child G (N = 309)

	Monotone	Intonation	Modulation	Rhythmic[a]	Jerky[a]	Halting[a]	Fluent	Intelligibility	Odd
Sing-Song	-.146**	.584***	.067	.096	.128	-.065	.054	.217***	.410***
Monotone		-.177**	.068	.154	.088	.106	.001	.132*	.292***
Intonation			.279***	.246**	.321**	.155	.028	.222***	.585***
Modulation				.215**	.345***	.096	.061	.004	.470***
Rhythmic					.354***	.066	.017	.174*	.496***
Jerky						.507***	.240**	-.050	.503***
Halting							.182*	-.091	.179*
Fluent								.012	.136*
Intelligibility									.540***

*** p < .001

** p < .01

* p < .05

[a] N = 159 (excluded one-word utterances)

THE UNIVERSITY OF MEMPHIS

Institutional Review Board

To:	Rachel Fields Audiology and Speech-Language Pathology
From:	Chair, Institutional Review Board For the Protection of Human Subjects irb@memphis.edu
Subject:	Pitch Analysis and Subjective Ratings of the Speec (012711-233)
Approval Date:	February 25, 2011

This is to notify you that the Institutional Review Board has designated the above referenced protocol as exempt from the full federal regulations. This project was reviewed in accordance with all applicable statuses and regulations as well as ethical principles.

When the project is finished or terminated, please complete the attached Notice of Completion form and send it to the Board via e-mail at irb@memphis.edu.

Approval for this protocol does not expire. However, any change to the protocol must be reviewed and approved by the board prior to implementing the change.

____

Chair, Institutional Review Board
The University of Memphis

Cc: Dr. Julie E. Cleary