Perception and Comprehension of Prosody in children with ASD

By

Pumpki Lei Su

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY
in
Hearing and Speech Sciences
May 8, 2020
Nashville, Tennessee

Approved:
James W. Bodfish, Ph.D.
Stephen Camarata, Ph.D.
Duane Watson, Ph.D.
Melissa Duff, Ph.D.

To my parents,

who encouraged me to follow my curiosity even if it leads to a disastrous mess

and who also instilled in me persistence by insisting that I clean up my own disastrous mess

To my grandmother,

who is one of few grown-ups who remembers exactly what it is like to be a child

and to my little brother, Mingming,

who shall always remain in the present tense in my heart

and who has taught me that the deep and genuine human connections that we cherish

transcend death and separation

ACKNOWLEDGEMENTS

I still remember when I was interviewing for doctoral programs five years ago, one potential mentor at a different institution only asked me one question during my interview, “Are you sure you want to do this???” I remember thinking, “Of course I am sure. Why else would I be here?” Now, five years later, while my answer to this question remains positive, I now understand what she meant by that question. Only through learning and experiencing what a scientific lifestyle is like can one truly discover whether the pursuit of knowledge and the exhilaration of discovery outweighs the constant uncertainty and obscurity, frequent frustration, and a seemingly unrelenting stream of rejection. Looking back, I am greatly indebted to many people for the support, encouragement, and love I have received throughout this journey.

First and foremost, I would like to thank my two advisors, Drs. Jim Bodfish and Stephen Camarata, for their exceptional mentorship and commitment. I feel extremely fortunate to have not just one, but two most inspiring and supportive mentors anyone could ever ask for. In Shanghainese dialect, we would describe such luck as “having a forehead that can reach the ceiling” (额角头碰到天花板). The full subtlety of the meaning may be lost in translation but my gratitude shall show through between these words. Thank you both for numerous scientific and life discussions throughout the past five years that not just taught me how to complete a research project but also fostered my intellectual and personal growth as a young scientist. And Steve, your constant dad jokes did make grad school just plain good fun sometimes.

I’d also like to thank other faculty, committee members, and mentors who have been immensely supportive. Dr. Paul Yoder, thank you for seeing something in me that I had yet to see in myself and investing so much into my career development. Thank you for providing a perfectly balanced dosage of incisive criticism and encouragement that pushed and inspired me to think deeply and write clearly. Dr. Duane Watson, thank you for your considerable contribution to my dissertation work and your sense of humor. The award of “most likely to find the most accurate meme to post on twitter or syllabus” goes to you without a single doubt. Dr. Melissa Duff, thank you for being a remarkable female scientist role model and for saying loudly that research is messy. Dr. Antje Mefferd, thank you for many thought-provoking discussions on research, teaching, and many shared background and interests. You will also always be my favorite international food buddy! Dr. Linda Watson, you are one of my favorite language researchers and human beings. Thank you for generously offering helpful input at various stages of my graduate career and always lifting my morale with kindest words of encouragement. Drs. Casey Lew-Williams and Megan Roberts, thank you both for welcoming me into your labs during my undergraduate education and getting me hooked on science. Finally, I would like to express my deep gratitude to Mary Camarata, who has been much more than a clinical mentor to me. Thank you for empowering me in the clinic and for being a trusted friend in life.

This work would not have been possible without my funding: Vanderbilt International Scholar Program, VICTR VR52148, and Graduate Research Grant from Organization for Autism Research. I am also deeply thankful for all children and families who participated in this research and who I had the opportunity to work with in the clinic.

Thanks to my wonderful lab mates from both Developmental Disabilities Lab and Bodfish Lab, past and present, for stimulating discussions and lovely company: Tonia Davis, who automatically took on the big sister role and has always been a phone call away, George Castle, Hope Lancaster, Allison Whitten, Robin Shafer, and Mary Alice Keller. Thanks to my marvelous officemates, Natalie Covington and Carlos Benitez-Barrera, for always being willing to talk through anything. You know it’s true friendship when you spent over 4,000 hours (which is probably a conservative estimate) with two other people in a confined space over 4 years and still enjoy their company. Thanks to my friends in Hearing and Speech, Laurel Teller and Sylvia Shih-Yuan Liang in particular, for being there for me through my ups and downs.

Next, I want to extend an extra special thanks to all my friends. Thanks to Yue, Betty, Yan, Yongtai, Chao, Susu, Kathy, Adam, Simeng, Bobo, Weiran, Long for being my family 8,000 miles away from home. Thanks to Rita for giggles, fabulous pun jokes, and cat memes. Thanks to Ariel Wu, OC, Annie Kong, Sherry Xu, and Chun Chen for your support and love from afar. Gargantuan thanks going out to Adela Zhang for introducing me to *Friends*, being there for me for 16 years and counting, reading numerous paper and application drafts, and even helping me with word choices for this Acknowledgement section.

Finally, I owe everything to my family. Thanks to my parents and my grandmother for their unconditional love and daily reminders to be kind to myself and to the world. Thanks to my Danish parents, Christa and Palle, for loving me as their own. I think of you both every time I see a white daisy. Thanks to my cats, Sleepy and Softie, for making a teeny tiny but indispensable contribution to human science by supporting an aspiring human scientist emotionally and by sitting on her lap day after day to supervise the progress and the quality of this work. Lastly, to my partner Ling, thank you for being my biggest cheerleader, favorite dance partner, personal programming instructor and debugger, fearless travel buddy, best-Chinese-food-chef-in-Nashville-and-probably-Tennessee, peer mentor, proof reader, research participant, friend, and many many more. I couldn’t have asked for a better partner and can’t wait for the day when we get to live in the same city again. All my love to you.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Social impairment and restricted and repetitive behaviors are the defining characteristics of Autism Spectrum Disorder (ASD; American Psychiatric Association, 2013); however, language deficits are also common in this population (Magiati, Tay, & Howlin, 2014) and are likely related to the core features of the disorder. Pragmatic language, defined as the social use of language, stands at the intersection of the social and language domains and could provide important insights into both social and language deficits in ASD (Prutting, 1982; Tager-Flusberg, 1999). In particular, prosody, one crucial aspect of pragmatic skills, is essential to social participation and interaction and, accordingly, is a reasonable candidate for mechanistic studies designed to advance current understanding of social and language impairments in ASD. Though structural language skills vary in this population, individuals with ASD demonstrate significant difficulties with at least one aspect of prosody (McCann & Peppé, 2003; McCann et al., 2007). Furthermore, prosodic deficits observed in this population tend to persist and affect long-term social and communicative competence even when other aspects of language improve (Paul, Shriberg, et al., 2005). These findings suggest that accurate identification and effective interventions that target prosodic deficits are necessary to improve long-term social and communication outcomes for individuals with ASD.

To date, the vast majority of existing research on prosody has primarily focused on expressive abilities (Peppé & McCann, 2003). Receptive prosody remains poorly understood and no objective measure is currently available to assess receptive prosody in this population. This knowledge gap has hindered practitioners from identifying specific receptive prosodic deficits, leaving them untreated. Untreated deficits in receptive prosody can have a direct functional impact in everyday life. To illustrate, the functional importance of receptive prosody can be evident when an individual with ASD misses a job opportunity because he or she misinterpreted a question with raising intonation as a statement and did not answer, or when a child with ASD fails to make friends at school because he or she constantly misunderstands other children's meaning due to difficulties differentiating sarcastic tones from genuine tones.

In this introduction, section 1.1 provides an overview of prosody by establishing its definition, components, acoustic correlates, and functions. Section 1.2 discusses the significant role of prosody, specifically receptive prosody, for language learners and language users. Section 1.3 concludes by introducing the aims of the present work and providing an overview of the rest of the chapters in this dissertation.

1.1 Prosody: Definition, Components, Acoustic Correlates, and Functions

In their summary of the research literature on prosody, Cutler, Dahan, & van Donselaar (1997) noted the extent to which definitions of prosody have varied across studies. In some works, an abstract and theoretical definition of prosody is used that does not emphasize the realization or production of prosody (e.g. "the organizational structure of speech"; Beckman, 1996). In other studies prosody is defined as the acoustic properties of an utterance ("pitch, tempo, loudness, pause"). In this work, I adopt Cutler et al.'s definition which falls between these two extremes and defines prosody as "the linguistic structure which determines the suprasegmental properties of utterances" (Cutler et al., 1997). Such a definition considers both the abstract linguistic structure and the acoustic realization of prosody in spoken communication.

Prosodic components include emphasis (also termed contrastive pitch accent or word prominence), lexical stress, intonation, intonational breaks (also termed boundary, boundary tones, or pause), and rhythm (or more generally, timing)(Wagner & Watson, 2010). Acoustically, these components are conveyed by variations in fundamental frequency (F0), intensity, and duration (Lehiste, 1970). Their respective perceptual correlates are pitch, loudness, and length. It is important to note that there is no one-to-one correspondence between prosodic phenomena (i.e. intonation, stress, rhythm) and acoustic parameters (i.e. F0, intensity, duration). Each prosodic phenomenon is comprised of multiple acoustic correlates. For instance, contrastive pitch accent is acoustically associated with longer duration, increased intensity, and greater F0 movement on the stressed segment (Watson, 2010). A boundary tone that marks the end of a linguistic unit is comprised of longer duration of the final syllable, the presence and length of pauses, and F0 excursions at prosodic boundaries (Wagner & Watson, 2010).

Regarding linguistic functions, prosody serves grammatical, semantic, and pragmatic functions (Crystal, 1986; Kehoe, 2013). Grammatical prosody refers to the use of prosody to signal syntactical information within linguistic units (Fry, 1955; Selkirk, 1995). For example, intonation at the end of an utterance indicates whether an utterance is a statement (falling pitch) or a question (rising pitch). Lexical stress within a word signals whether a lexicon is used as a noun (e.g. REcord) or a verb (e.g. reCORD). Placement of pauses can disambiguate phrases such "cinnamon buns and chocolate" to refer to two items (i.e. "cinnamon buns / and chocolate") or three items (i.e. "cinnamon / buns / and chocolate"). Semantic prosody enhances or modulates information beyond the literal meaning of an utterance (Bolinger, 1961; Watson, 2010). For example, contrastive pitch accent is often used to direct listener's attention to a piece of salient information in a discourse context or to contrast a piece of information with possible alternatives (Bolinger, 1961, 1972; Pierrehumbert, 1980). The interpretation of the same sentence, "Sophie has four hedgehogs", changes as the speaker varies the location of the contrastive pitch accent. One would say, "Sophie has four HEDGEHOGS," to emphasize the type of pets Sophie owns, or "Sophie has FOUR hedgehogs," to emphasize the quantity. Lastly, pragmatic functions of prosody convey a speaker's intentions or emotional states (Bolinger, 1989) and help listeners interpret the message in a discourse context (Cutler et al., 1997). For example, the following utterance, "I am having so much fun," could be interpreted either as genuine or sarcastic depending on the prosodic characteristics of the utterance and the context of the discourse.

Given that prosody can be conceptualized having both form (auditory-perceptual features) and function (linguistic functions that prosody serves), throughout this work, regarding receptive prosody I use *perceive/perception* or *discriminate/discrimination* to refer to an individual's ability to process prosody on the form level and *understand/comprehend* to refer to the ability to process prosody on a higher cognitive-linguistic level.

1.2 The Role of Prosody for Language Learners and Users

The ability to perceive and comprehend prosody contributes significantly to early language acquisition for infants and children as language learners and successful communication for adults as language users (Cutler & Swinney, 1987; Frazier, Carlson, & Clifton, 2006; Mehler et al., 1988; Paul et al., 2005). Research in this area has focused on the language learner's sensitivity and preference to certain prosodic cues and their use of prosody to learn other aspects of language (i.e. word learning, syntax parsing, and phonological awareness and emergent literacy), and on the language user's use of prosody to allow more organized and efficient processing of linguistic information.

1.2.1 Prosody contributes to language learning

A large body of infant speech perception research has suggested that infants are sensitive to prosodic patterns from very early on in life (See Gervain, 2018 for a review). Many studies have used the high-amplitude sucking paradigm and the head-turn preference procedure and have found that infants within 6 months are capable of discriminating varying pitch and rhythm patterns in synthetic and natural speech (Karzon & Nicholas, 1989; Morse, 1972), at both word (Nazzi, Floccia, & Bertoncini, 1998) and sentence levels (Jusczyk et al., 1992), in their native language and across unfamiliar languages with different rhythmic classes (Nazzi, Bertoncini, & Mehler, 1998; Nazzi, Floccia, et al., 1998; Nazzi, Jusczyk, & Johnson, 2000). Studies that used low-pass filtered speech stimuli or stimuli produced by different speakers have further confirmed that infants are discriminating patterns based on prosodic features instead of phonetic- or speaker-related acoustic information (Jusczyk et al., 1992; Mehler et al., 1988; Nazzi, Bertoncini, et al., 1998). Regarding prosodic elements, infants discriminate words with final vowels differing in F0, amplitude, and duration (Bull, Eilers, & Oller, 1984, 1985; Eilers, Bull, Kimbrough, & Lewis, 1984). Not only are infants sensitive to prosodic cues, but they also exploit prosodic information to learn language. The facilitative role of prosody in language acquisition has been investigated in three areas of research: prosodic features in child-directed speech (CDS), early word segmentation and word acquisition, and prosodic bootstrapping.

1.2.1.1 Functional significance of prosodic features in CDS

Child-directed speech is a distinctive type of speech that caregivers use when interacting with young children and displays unique syntactic, lexical, and prosodic characteristics, such as shorter utterance, simpler vocabulary, larger vowel space, exaggerated pitch contour, and slower rate compared to adult-directed speech (Cristia, 2013; Ferguson, 1964; Fernald & Simon, 1984; Soderstrom, 2007). Typically developing (TD) children show a strong preference for CDS over ADS from the very beginning of life (Cooper, Abraham, Berman, & Staska, 1997; Cooper & Aslin, 1990). A recent meta-analysis analyzed findings from 34 studies and reported that infants' preference for CDS had an average effect size of Cohen's *d* of .72 (Dunst, Gorman, Hamby, & Hamby, 2012). Additionally, abundant literature has provided convergent evidence for the beneficial role that CDS plays in early childhood development (Begelsonm et al., 2017). Important functions of CSD identified include attracting, engaging, and maintaining infant's attention (Butler, O'Sullivan, Shah, & Berthier, 2014; Cohen et al., 2013; R. P. Cooper & Aslin, 1990), facilitating early language learning (Thiessen, Hill, & Saffran, 2005), and boosting social recognition (Schachner & Hannon, 2011).

Though infant's preference for CDS and the positive impact of CDS on language development have been replicated using various stimuli and paradigms in the past two decades in typically developing population, most studies focused on global properties of CDS and until recently it remained unclear the extent to which prosodic aspects of CDS contribute to language learning. A recently published meta-analysis reviewed 15 studies that tested relations between prosodic features of CDS and infant outcomes and reported an overall significant positive relation between prosodic features typical of CDS and a variety of infant outcomes, including attentional, pre-linguistic, and linguistic outcomes (Spinelli, Fasolo, & Mesman, 2017). In particular, the association has been shown to be stronger for infants younger than 9 months compared to older infants and for prelinguistic (e.g. vocal responses and imitations) outcomes compared to linguistic outcomes (e.g. vocabulary comprehension and production). Findings from this meta-analysis confirm early theories by Fernald that prosodic forms of CDS have attentional and affective functions for young children (Fernald, 1989, 1992).

1.2.1.2 Prosody and early word segmentation and acquisition

A second line of research that supports the facilitative role that prosody serves in language development has involved work on word segmentation and acquisition. Many theories have been put forth to explain how infants segment the continuous speech stream into words, one of which proposes that infants rely on prosodic cues in their native language to detect word boundaries (Jusczyk, Houston, & Newsome, 1999). For example, in English, the predominant stress pattern of words is trochaic (i.e. a strong syllable followed by a weak syllable; Cutler & Carter, 1987). Accordingly, Cutler and colleagues hypothesized that that English-speaking infants apply a Metrical Segmentation Strategy and use stressed syllables as indicators of word onsets in the speech stream (Cutler, 1990).

Support for this hypothesis was reported in a study by Jusczyk and Thompson (1978) that used the high-amplitude sucking paradigm and found that 2-month-old English-speaking infants could detect stress pattern changes in speech stimuli (e.g. DAga vs. daGA). In addition, a series of head-turn preference experiments by Jusczyk, Aslin, Houston, & Newsome further showed that 7.5-month-old English-speaking infants were also highly biased to interpret stressed syllables as word onsets (Jusczyk & Aslin, 1995; Jusczyk et al., 1999). In particular, in their 1999 study, 7.5-month-olds succeeded at segmenting bisyllabic words with strong-weak stress patterns (e.g. doctor, kingdom) from sentences but failed to detect bisyllabic words with weak-strong patterns (e.g. guitar, device) in sentences. A follow-up experiment also showed that 7.5 month old infants mis-segmented weak-strong words followed by the same monosyllabic words (e.g. “guitar is”) as if they detected words with a strong-weak pattern (e.g. “taris”), which further lends support for the Metrical Segmentation Strategy Hypothesis. These findings collectively suggest that prosody serves as a powerful yet not completely error-free cue for word segmentation for young infants. In contrast, ten-month-olds were successful at segmenting weak-strong words and did not display the same biased tendency to interpret stressed syllables as word onsets in weak-strong words, which suggests that by that age, infants has learned to use other information such as syllable distributional or coarticulation cues in addition to prosodic cues to segment words. Interestingly, two studies (Johnson & Jusczyk, 2001; Johnson & Seidl, 2009) subsequently tested the extent to which infants weigh prosodic information when integrating multiple cues during word segmentation. Both studies found that 8-month-olds and 11-month-olds rely more heavily on stress pattern cues when they conflict with statistical distributional cues. These findings suggest that prosodic cues remain highly salient for infants at 11 months and that the 10-month-olds in Jusczyk et al. (1999)’s study likely exploited cues other than distributional cues[1] to segment weak-strong words.

In addition to parsing the speech stream into words, young children also take advantage of prosodic cues to learn word-object pairings. Curtin (2009) used an adapted switch design and tested whether 12-month-old infants could learn two novel word-object pairings that differ only in stress patterns. In this design, infants were habituated to two novel objects, each paired with a novel word (in this study, the two novel words were “BEdoka” and “beDOka”). Once habituated, they were exposed to two types of test trials. In the control trial, a familiar pairing was presented. In the test trial, the pairing was switched to create a new word-object combination. Results showed that infants looked significantly longer to the switched pairing compared to the same

[1] The authors suggested that infants in Jusczyk et al (1999) likely used allophonic or coarticulation cues instead of distributional cues to learn weak-strong stress patterned words.

familiarized pairing. Using the same design, Curtin (2010) replicated this finding in 14-month-olds and extended prior work by showing that infants encode both the stressed syllable (e.g. BE vs Do in the example of “BEdoka” vs. “beDOka”) and the position of the stressed syllable (e.g. “BEdoka” vs. “doBEka”) in their lexical representations of newly learned words.

Two theoretical frameworks have emerged from the early word learning research that have included a focus on prosody: PRIMIR (Processing Rich Information from Multidimensional Interactive Representations; Werker & Curtin, 2005), and WRAPSA (Word Recognition and Phonetic Structure Acquisition; Jusczyk, 1992, 1997). Both models maintain that infants have access to a wide variety of rich information when processing lexicons, including prosodic information. These two models differ in the extent to which each model places emphases on prosody. WRAPSA assumes that infants access prosodic information first and then syllabic and phonetic cues across developmental levels, whereas PRIMIR proposes that infants have access to all information (i.e. prosodic, syllabic, phonetic, indexical characteristics) simultaneously and that priority was given to a specific information based on both a child’s developmental level and task demands. Despite their different assumptions, both acknowledge that infants access prosodic information as a strategy to learn words.

The studies reviewed above collectively highlight the importance of prosody in communicating information about word boundaries and word structures to beginner language learners. Two recent studies have shifted attention to the extent to which prosody signals the semantic meanings of words. To study whether prosodic elements communicate about the content of spoken language, Nygaard, Herold, & Namy (2009) asked adult speakers to produce novel words (e.g. blicket, seebow) in sentences where the novel words were used to convey one of two meanings from pairs of dimensional adjectives (e.g. happy/sad, big/small, yummy/yucky).

They found that prosodic elements associated with particular semantic meanings. For instance, higher F0 and greater F0 variation were associated with positive valence (e.g. happy, yummy). Greater amplitude and longer duration were associated with words that indicate big size (e.g. tall, big). Additionally, they found that both adult listeners and 5-year-olds were significantly more accurate at inferring word meanings when the word has matched prosody than mismatched prosody (Herold, Nygaard, Chicos, & Namy, 2011; Nygaard et al., 2009). These two findings extend the current understanding of the ways prosody facilitates word learning.

1.2.1.3 Prosody and syntax

A widely discussed hypothesis is termed prosodic bootstrapping hypothesis, which proposes that children use prosodic information to segment sentences into linguistic units such as phrases or clauses (Gleitman & Wanner, 1982; Jusczyk et al., 1992; Morgan, 1986; Morgan & Demuth, 1996). This hypothesis has been supported by two types of evidence. The first involves descriptive studies that found that certain prosodic cues and their underlying acoustic characteristics are associated with syntactic boundaries. In particular, pauses are more likely to be present or be longer at syntactic boundaries than within phrases (Goldman-Eisler, 1972; Morgan, 1986; Scott, 1982). Syllables at the end of linguistic units appear to be lengthened compared to syllables in other positions (Crystal & House, 1988; Klatt, 1975; Lehiste, 1973). F0 declines at the end of linguistic units and rises at the beginning of the following unit (Cooper & Sorensen, 1977; Pierrehumbert, 1980).

The second type of evidence has suggested that infants are sensitive to prosody-syntax associations and use them to segment the speech stream into phrases. Early work has used the pause insertion preferential listening paradigm to test infants' perception of prosody-syntax associations. In this paradigm, infants were exposed to sentences with pauses inserted either at syntactic boundaries or non-boundary positions and infants' listening time was analyzed. The first study that adopted this paradigm found that 7 to 10 months old infants listened longer to speech samples with pauses inserted at clause boundaries compared to samples with within-clause pauses (Hirsh-Pasek et al., 1987). Two subsequent studies (Gerken, Jusczyk, & Mandel, 1994; Jusczyk et al., 1992) compared pauses inserted at phrase boundaries versus within-phrase positions and found that 9 months old English-speaking infants listened significantly longer to sentences where pauses were inserted at phrase boundaries (i.e. between noun phrase and verb phrase; e.g. "That / looks great", "Would it help if I / move the chair?") compared to sentences with pauses inserted at within-phrase positions (i.e. after verb; e.g. "That looks / great", "Would it help if I move / the chair?"). Though these early studies have been widely cited and interpreted as evidence that supports the prosodic bootstrapping hypothesis, it is important to point out that these findings mostly suggest that infants are sensitive to prosodic cues that associate with syntactic boundaries but did not provide direct evidence that infants use these cues to segment speech stream (Jusczyk et al., 1992).

The prosodic bootstrapping hypothesis has recently been revisited by Hawthorne and colleagues (Hawthorne & Gerken, 2014; Hawthorne, Rudat, & Gerken, 2016). As an attempt to provide evidence that infants can use prosody-syntax associations to segment the speech stream into constituent-like units, Hawthorne and Gerken familiarized 19-month-olds with sentences either with 1 clause (ABCDEF, where each letter represents a nonsense word, such as /bʌp/, /dɪv/) or 2 clauses (ABC, DEF). Infants were then tested on sentences that either represent a grammatical "movement" of the 2 clauses (i.e. DEF, ABC) or an ungrammatical one (i.e. EFA, BCD). A significant interaction between group and test condition was detected: the effect of condition was significant for the 2-clause group but not for the 1-clause group, which suggests that infants familiarized with the 2-clause sentences treated prosodically-grouped phrases as cohesive linguistic units that can be reordered. Hawthorne and colleagues (2016) provided further evidence that 20-months-olds used the prosodic hierarchy to learn syntactic constituents at multiple levels. After being familiarized to nonsense sentences with adverbial modifier + clause prosody (i.e. A, BCDE, such as "Thankfully, the girl likes cheese"), 20-month-olds succeeded at discriminating between constituents and non-constituents at both the phonological phrase level (BC is a constituent vs CD is not) and intonational phrase level (BCDE is a constituent vs. ABCD is not). Findings from these two studies extend prior work on the prosodic bootstrapping hypothesis and indicate that 19- and 20-month-olds use prosody to extract phrases and treat them as cohesive linguistic units that can be reordered. Even though one could argue that infants at this age do not yet have the concept of syntactic constituents, infants' ability to extract smaller linguistic units from larger sentences using prosody allows them to "divide and conquer" and suggests that prosodic cues are helpful for syntax learning.

Despite the popularity of this hypothesis and the supportive evidence discussed above, other researchers have pointed out that this hypothesis may have been over-stated and over-interpreted in the literature. Morgan & Demuth (1996) argued that the "prosodic" part of the prosodic bootstrapping hypothesis is "confining" and "misleading" (p.2) because prosody is not unique in contributing relevant information for syntax acquisition. The authors suggested that it would be more accurate to term this hypothesis "phonological bootstrapping" instead to reflect the idea that children use a wide variety of information in spoken language input to learn syntax such as phonetic, phonotactic, distributional information. Another criticism is that the association between prosody and syntax may not be reliable or robust enough to support direct access to syntactical structures (Fernald & McRoberts, 1986; Seidl, 2007). For instance, prosodic boundaries and syntactic structures can diverge such that pauses might be present at a relatively minor syntactic boundary over a major boundary. Fisher & Tokura (1996) showed that prosodic cues would incorrectly cue the syntactic boundary in sentences with unstressed pronoun subjects (e.g. "You like the doggy") because the pause would naturally fall between the verb and the article (i.e. "You like / the doggy") and would not correspond to its major syntactic boundary (i.e. NP / VP). Watson & Gibson (2004) also introduced the notion that the presence and the duration of prosodic boundaries are also impacted by prosodic and phonological constraints for the speaker. In particular, a prosodic boundary is more likely to occur before and after a long syntactic constituent to allow the speaker to plan, process, and recover (Wagner & Watson, 2010; Watson & Gibson, 2004). Because of the divergence between prosody and syntax, Fernald and McRoberts (1996) argued that though previous studies that showed that prosodic boundaries tend to occur at syntactic boundaries (i.e. *p*(prosodic cue | syntactic structure)), the opposite (i.e. *p*(syntactic structure | prosodic cue)) does not hold true. These critiques do not undermine the proposed importance of prosody in language development but rather suggest that prosody *on its own* may not be a reliable or sufficient cue for syntax acquisition and that the prosodic bootstrapping hypothesis may have over-simplified the role of prosody in syntax acquisition. In line with this point, Fernald & McRoberts suggested that prosody may help infants discover distributional regularities in the spoken language input and prosody in combination with other types of cues facilitates syntax acquisition.

1.2.1.4 Summary

The results of multiple research studies provide support for the theory that infants are sensitive to various prosodic elements from very early on in life and use them to learn other aspects of language, such as semantics and syntax. It is important to stress that though this review supports the argument that children use prosodic signals to learn language, it does not suggest that prosody is unique in providing supportive cues for early language development (Morgan & Demuth, 1996). Additionally, a common pitfall for studies on the usefulness of prosody is that most theories appear to have taken for granted infant's ability to learn prosody in the first place. For instance, many theorists (Gleitman & Wanner, 1982; Pinker, 1984) have assumed that infants were born with perceptual filters and have argued that infants are innately biased to attend to prosodic cues. Others have carefully pointed out that such a position may not represent the role of prosody in early language acquisition and does not further our understanding of the specific means infants employ to acquire prosody and take advantage of prosody to serve language development (Fernald & McRoberts, 1986; Fisher & Tokura, 1996).

1.2.2 Prosody contributes to spoken language comprehension

In additional to facilitating language learning in young children, receptive prosody is also integral to successful spoken language comprehension and social communication (Watson, 2010). Research on the role of prosody in spoken language comprehension for adults who are proficient language users has focused on three main areas: the use of prosody in resolving syntactic ambiguities, the role of prosody in conveying and modulating semantic interpretations, and the pragmatic effect of prosody in reference resolution and conveying meanings in discourse context. Due to the scope of this chapter, I have limited this review to two prosodic elements, prosodic boundary and pitch accent, and will discuss how these two cues aid listeners' processing of syntactic, semantic, and pragmatic information.

1.2.2.1 Resolving syntactic ambiguities

Prosodic boundaries. Results from multiple studies on prosodic boundaries have established that listeners take advantage of prosodic boundaries and groupings to resolve both local and global syntactic ambiguities (Kjelgaard & Speer, 1999; Lehiste, 1973; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Schafer, 1997; Snedeker & Trueswell, 2003; see Cutler et al., 1997 and Wagner & Watson, 2010 for two reviews). These studies have collectively shown that listeners use prosodic boundaries to disambiguate the scope of modifiers, the distance of the prepositional or adverbial phrase attachment, and the location of attachment for a phrase in the middle of a sentence (Lehiste, 1973; Price et al., 1991). Prosodic boundary can also be used to differentiate between parentheticals vs. subordinate clauses, appositions vs. attached noun phrase or prepositional phrase, and early vs. late closure in sentences with local ambiguities (Clifton, Carlson, & Frazier, 2002; Hirschberg, 2017; Schafer, 1997; Speer, Kjelgaard, & Dobroth, 1996; Wagner & Watson, 2010). To illustrate, Table 1 summarizes the types of global and local syntactic ambiguities that can be resolved by prosodic boundary cues and provides examples for each type.

Table 1. Summary of the types of syntactic ambiguities that can be resolved by prosodic boundary cues

Type	Example	Reference
Near vs. far attachment of final phrase	The hostess greeted / the girl with a smile. (PP as an attachment to the near NP "the girl") The hostess greeted the girl / with a smile. (PP as an attachment to the far NP "the hostess")	Lehiste, 1971
	I've read a review of nasality / in German. (near attachment) I've read / a review of nasality in German. (far attachment)	Price et al., 1991
	Mary maintained / that the CEO lied when the investigation started. Mary maintained that the CEO lied / when the investigation started.	Carlson, Clifton, Charles, & Frazier, 2001
Ambiguous scope of modifier	The old men / and women stayed home. The old men and women / stayed home.	Lehiste, 1973
Ambiguous conjunctions	Steve / or Sam and Bob Steve or Sam / and Bob	Lehiste, 1973
	Fruit / salad / and milk (3 items) Fruit salad / and milk (2 items)	Peppé & McCann, 2003
Parenthetical clause vs. non-parenthetical subordinate clauses	Mary knows many languages / you know (parentheticals) Mary knows many languages you know ("you know" as a subordinate clause)	Price et al., 1991 Schafer, 1997
Appositions vs. attached NP or PP	The neighbors who usually read, the Daleys, were amused. (appositions) The neighbors who usually read the Dailies were amused. (attached NP)	Price et al., 1991;
	The animal that usually fights, the lion, is missing. (Apposition) The animal that usually fights the lion is missing. (attached NP)	Hirschberg, 2017

Preposition vs. particle	John laughed / at the party. (preposition) John laughed at / the party. (particle)	Price et al., 1991; Hirschberg, 2017
Left vs right attachment of middle phrase	They left early / in May. (left attachment) They left / early in May. (right attachment)	Price et al., 1991; Schafer, 1997
	When you learn gradually / you worry more. (left) When you learn / gradually you worry more. (right)	Shattuck-Hufnagel & Turk, 1996
Object NP + vocatives vs. complex NP	I will take the eggs, Benedict. (object NP + vocatives) I will take the Eggs Benedict. (complex NP)	Shattuck-Hufnagel & Turk, 1996
Ambiguous PP attachment: VP vs. NP attachment	Tap the frog / with the flower (VP attachment) Tap / the frog with the flower (NP attachment)	Snedeker & Trueswell, 2003
Restrictive vs non-restrictive relative clause	My brother who is a writer / needs a new job. (restrictive relative: I have multiple brothers and the one who is a writer needs a new job) My brother / who is a writer / needs a new job. (unrestrictive relative: I may or may not have other brothers)	Hirschberg, 2017
Early vs late closure	Whenever the guard checks the door / it's locked. (late closure) Whenever the guard checks / the door is locked. (early closure)	Speer, Kjelgaard, & Dobroth, 1996

Note: NP: noun phrase; VP : verb phrase ; PP : prepositional phrase.

The effect of prosodic boundaries on listener's processing is most convincingly shown in Speer et al. (1996). While early studies mainly presented listeners with syntactically ambiguous sentences with or without informative prosodic boundary cues and asked listeners to judge in a forced-choice task, Speer and colleagues presented listeners with sentences with local ambiguities (early closure vs. late closure) in three conditions: felicitous prosody (i.e. informative prosodic boundary cues), conflicting prosody (i.e. misleading prosodic boundary cues), and ambiguous prosody (i.e. control condition). Some examples of sentences in these three conditions are provided below:

Felicitous prosody:

(1a) Whenever the guard checks the door / it's locked. (late closure with late boundary)

(1b) Whenever the guard checks / the door is locked. (early closure with early boundary)

Conflicting prosody:

(2a) Whenever the guard checks / the door it's locked. (early closure with late boundary)

(2b) Whenever the guard checks the door / is locked. (late closure with early boundary)

Baseline condition with neutralized prosody boundary cues:

(3a) Whenever the guard checks the door is locked.

Listeners were instructed to press a lever as soon as each sentence ends. Participants' reaction time data revealed that both early and late closure sentences with felicitous prosody showed a significant processing advantage compared to baseline sentences. In contrast, participants processed both types of sentences with conflicting prosody significantly slower than baseline sentences. Another finding from this study is that an interaction was detected between the type of sentence (early closure vs. late closure) and condition: participants' processing of early closure sentence was significantly slower than late closure sentence in both the baseline and the conflicting prosody condition but was comparable in the felicitous prosody condition. These findings suggest that participants are biased to interpret the noun as a direct object in the verb phrase instead of as the subject of the main clause. The finding that this bias was reduced in the felicitous prosody condition for early closure sentences again shows that informative prosodic cues can help with challenging syntactic processing.

Though the studies reviewed above have shown that prosodic boundaries can inform and facilitate the syntactic parsing process, the findings of other studies have also identified some limitations to this model. One limitation is that prosodic boundaries can only resolve ambiguities when surface syntactic structure differs across interpretations (Lehiste, 1973; Lieberman, 1967; Shattuck-Hufnagel & Turk, 1996; Wagner & Watson, 2010). Wagner and Watson (2010) provided an example from Liberman (1967): the sentence “flying airplanes can be dangerous” is globally ambiguous because “flying” can either be interpreted as a gerund or an adjective modifier. Given that both interpretations share the same syntactic configuration and that the major syntactic breaks are at the same position, prosodic boundaries can’t provide additional information to disambiguate the sentence.

Another limitation is that studies on prosodic boundary and syntactic disambiguation have mainly examined the extent to which listeners use prosodic cues without discussing whether speakers typically provide these cues in natural conversations (Allbritton, McKoon, & Ratcliff, 1996; Cutler et al., 1997). Snedeker and Trueswell (2003) used a referential communication task to study if speakers consistently produce informative prosodic boundary cues to disambiguate meanings of ambiguous sentences. Participants were randomly assigned a speaker or a listener role. The speaker was instructed to communicate some commands that the listener will complete (e.g. “Tap the frog with the water”). Results show that the availability of such cues depends on the speaker's knowledge of the context. Speakers only reliably produced prosodic boundary cues when they were aware of alternative interpretations and deemed that prosodic cues were necessary for clear communication. Listeners, on the other hand, were able to rapidly integrate prosodic cues in their processing when these cues were present. These findings suggest that even though prosodic boundary cues depend on speakers' knowledge of the referential context and may not be frequently used, they do guide listeners' syntactic parsing when they are available.

Pitch Accent. Though the majority of research on prosody and syntax in adult language processing has focused on prosodic boundaries and their underlying prosodic elements (i.e. pause, boundary tone, pre-boundary syllable lengthening), several studies have identified that pitch accents can also impact syntactic parsing (Carlson & Tyler, 2018; Grosz & Hirschberg, 1992; Kutik, Cooper, & Boyce, 1983). Specifically, Carlson and Tyler presented participants with sentences with ambiguous attachments with manipulated pitch accents (see below for two examples):

Ambiguous final adverbial phrase attachment:

(4a) Sally DISCOVERED that Pam had returned on Sunday.

(4b) Sally discovered that Pam had RETURNED on Sunday.

Ambiguous final prepositional phrase attachment:

(5a) Alison ENTERTAINED a toddler with many toys.

(5b) Alison entertained a TODDLER with many toys.

Results show that in sentences with ambiguous final adverbial phrases, pitch accents increased the attachment of the adverbial phrase as a modifier of the accented verb in listeners' interpretations. Likewise, for sentences with ambiguous final prepositional phrases, pitch accents on verbs or object nouns increased the attachment of the prepositional phrase to the accented head. In one experiment (Experiment 3) in this study, the authors also presented pitch accents without prosodic boundaries and found that the same effect, suggesting that the effect of pitch accents does not depend on prosodic boundaries. Other studies have found that the use of pitch accents can distinguish parenthetical phrases from subordinate clauses in sentences such as “WE only SUSPECTED THEY all KNEW that a burglary had been committed”(Hirschberg, 2017; Marcus & Hindle, 1990). Additionally, Liberman and Sproat (1992) suggested that pitch accents are useful for disambiguating phrases when the surface syntactic structure does not differ across interpretations. For example, the phrase “German teachers” can be interpreted as teachers of German language when a pitch accent is placed on “German” or as teachers who are German when a pitch accent is placed on “teachers.” These observations will need to be empirically tested to confirm that listeners are sensitive to such use of pitch accent in disambiguating complex noun phrases.

1.2.2.2 Conveying semantic and pragmatic meaning in discourse contexts

Though prosodic boundary or pitch accent can modulate or change the semantic or pragmatic interpretations of individual sentences, the role of prosody in signaling information of utterances is much more evident in the broad context of discourse. Findings from previous studies have indicated that prosodic cues can signal discourse structure (Grosz & Hirschberg, 1992; Herman, 2000; Yule, 1980), convey or modulate discourse meaning (Arnold, 2008; Bolinger, 1972; Dahan, Tanenhaus, & Chambersc, 2002; Ito & Speer, 2008; Terken & Nooteboom, 1987; Watson, Arnold, & Tanenhaus, 2008), and enhance listeners’ memory by strengthening representations of referenced items (Fraundorf, Watson, & Benjamin, 2010a).

Previous studies have shown that prosody encodes discourse structure in conversational interactions or monologues using corpus-based methods (Grosz & Hirschberg, 1992; Herman, 2000; Hirschberg & Nakatani, 1996; Nakatani, Hirschberg, & Grosz, 1995). Hirschberg and Nakatani (1996) analyzed acoustic and prosodic features of a corpus of elicited monologues produced by four untrained speakers who were instructed to give directions (Boston Directions Corpus; Nakatani, Hirschberg, & Grosz, 1995) and found that for both reading and conversational speech, sentences at various positions of a discourse context (initial, medial, and final) have distinctive prosodic patterns. Discourse-initial sentences differ from discourse-medial or discourse-final sentences in both pitch accent and rhythm, whereas discourse-medial utterances differ from discourse-final utterances in rhythm. Though this specific study did not test whether listeners were able to use such prosodic cues exclusively to understand or predict discourse structure, other studies have shown that listeners were able to use prosodic cues (prosodic boundary and pitch accent) to reliably and accurately identify the discourse position of isolated utterances when these cues are available (Herman, 2000).

Further evidence that supports the semantic and pragmatic role of prosody in spoken language comprehension comes from a large body of literature that examines contrastive pitch accent and the type of information that it signals to listeners. Previous studies have proposed that speakers use various pitch accents to signal information that is new, unpredictable, or less accessible in the discourse context (Arnold, 2008; Aylett & Turk, 2004; Birch & Clifton, 1995; Bolinger, 1972; Selkirk, 1995; Watson, Arnold, & Tanenhaus, 2006; Watson et al., 2008; see Wagner & Watson, 2010 for an excellent review on this topic). Recent advances in technology have also allowed researchers to examine the effect of contrastive pitch accent on listeners' language processing using eye-tracking methods. One widely-used paradigm is called the visual-world paradigm (VWP;Tanenhaus, Spivey, Eberhard, & Sedivy, 1995). In this paradigm, participants sit in front of a monitor or a display and are instructed to look at visual scenes and listen to sentences while their eye movements are recorded. Research designs using this approach have typically manipulated the linguistic stimuli and examined how the experimental manipulation affects participants' eye movements. The underlying rationale of this paradigm is that listeners make saccadic eye movements to target items as they hear relevant words in the linguistic stimuli and thus analyzing their eye movements could provide insights into their spoken language comprehension process (Tanenhaus et al., 1995).

A series of studies have used the VWP to understand whether listeners process pitch accent in real-time language processing. Dahan et al. (2002) is the first VWP study that provided evidence that listeners can rapidly use pitch accent information during reference resolution. In this study, participants were instructed to move objects to different locations on a computer and listened to sentences such as "Put the candle/candy below the triangle. Now put the CANDLE/candle above the square". Results of this study indicated that participants were biased to look at the new referent when the accent was present in the second sentence and look at the previously mentioned referent when the accent was absent. Additionally, participant's eye movement patterns between conditions began to differ around 300 ms after the onset of the second sentence, which suggests that participants were able to rapidly integrate pitch accent cues to guide their processing. Using the same experimental design as Dahan et al. (2012), Arnold (2008) replicated and extended Dahan et al.'s finding and found that adults and 4- to 5-year-olds were biased toward the previously-mentioned referent when hearing an unaccent noun. Ito and Speer (2008) tested a specific type of pitch accent (L+H*) and found that listeners were able to use L+H* contrastive pitch accent to anticipate an incoming referent. Listeners were found to fixate on a potential contrastive object faster when they heard a context-appropriate contrastive pitch accent on a prenominal adjective and slower when they heard a context-inappropriate accent. A follow-up study by the same group (Ito, Bibyk, Wagner, & Speer, 2014) found the same effect of L+H* contrastive pitch accent in children as young as 6-year-olds. These findings collectively show that listeners are sensitive to pitch accent cues in discourse contexts and use them to efficiently process spoken language.

Finally, several studies have suggested that prosodic cues in communication also have mnemonic benefits (Fraundorf et al., 2010a; Shintel, Anderson, & Fenn, 2014). Fraundorf and colleagues tested how two types of pitch accents (non-contrastive H* vs. contrastive L+H*) affected listeners' recognition memory. Participants listened to short discourses, each of which began with a short context passage that establishes two contrastive sets followed by a continuation passage that mentions one word from each contrast set. An example is provided below in (6a) and 6(b)

(6a) Context passage: Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys.

(6b) Continuation passage: Finally, the *British/French* spotted one of the monkeys in *Malaysia/Indonesia* and planted a radio tag on it.

Pitch accents of the italic target words were manipulated so that either the first target word was assigned an L+H* and the second word was assigned a H* or vice versa. After listening to 48 discourses, participants completed a recognition memory test. Results show that words presented with an L+H* accent were more accurately retrieved and these benefits persisted as long as a day. Shintel et al. (2014) reported similar findings that prosodic cues enhanced participants' memory in a word learning task. These findings suggest that prosodic cues not only facilitate the initial processing of a specific referent in a discourse context but also enhance memory by strengthening the representation of the referent and improving the encoding of relevant information.

1.2.2 Section summary

A review of current literature suggests that prosody facilitates young children's language development by attracting their attention to linguistic input in the environment, providing informative cues for word and phrase segmentation, and supporting the learning of syntactic structures. For proficient language users, previous studies have identified its facilitative role in the syntactic parsing process, referent resolution in discourse contexts, and even recognition memory.

1.3 The Present Work

The results of existing research on the relationship of prosody to language learning and language use summarized above can be used to guide the development of studies to examine the nature of prosody deficits in the context of clinical conditions associated with atypical language acquisition and use. Among the variety of conditions associated with language deficits, Autism Spectrum Disorder (ASD) has been reported to be associated with a high prevalence of a variety of prosodic deficits (Fusaroli, Lambrechts, Bang, Bowler, & Gaigg, 2017). Atypical prosody was among one of the original features that Leo Kanner included in his seminal paper (American Psychiatric Association, 2013; Kanner, 1943; McCann & Peppé, 2003) on autism. Clinical descriptions of prosodic deficits in ASD include flat, exaggerated, or "sing-song" intonation, and wide variation in intensity (Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2007). These prosodic deficits may pose tremendous barriers to language learning and everyday social participation for children with ASD. Deficits in expressive or receptive prosody can lead to communication breakdowns, negatively impact one's social competence, and are considered as one of the most significant challenges to social participation and integration (McCann, Peppé, Gibbon, O'Hare, & Rutherford, 2007; Paul et al., 2005). Despite the apparent significance of prosody in the context of autism, this has been a markedly under-studied aspect of the condition.

The goal of this dissertation was to use the established findings on prosody in the context of typical language development to design a novel line of research on prosody in children with ASD to help fill the gaps in the small literature of prosody in ASD. In the next chapter, I will review the existing research literature on prosody in ASD and identify specific gap areas in the literature. Chapter 3 presents the development and application of a visual-world eye-tracking task to test comprehension of specific prosodic cues in TD children and children with ASD. Finally, Chapter 4 concludes by summarizing findings from preceding chapters and discussing future directions and clinical implications inspired by this work.

CHAPTER 2

A Systematic Review of Receptive Prosody in Individuals with ASD

2.1 Introduction

Atypical prosody has been frequently identified in children and adults with Autism Spectrum Disorder (ASD; American Psychiatric Association, 2013; McCann & Peppé, 2003). Though receptive prosody deficits have also been reported in previous studies (Paul, Augustyn, Klin, & Volkmar, 2005; Wang, Dapretto, Hariri, Sigman, & Bookheimer, 2001), the majority of existing research on prosodic deficits in the ASD population have focused on expressive prosody (McCann & Peppé, 2003), leaving receptive prosody in ASD poorly understood. McCann and Peppé published a narrative review summarizing the current evidence base for prosodic deficits in ASD (McCann & Peppé, 2003), yet only three studies among 16 included studies examined receptive prosody.

Given that prosody generally does not have orthographic representation other than punctuations, the terminology of prosody and the scope of the topic tend to vary across studies (Peppé, 2009; Xu, 2015). One frequently used theoretical framework conceptualizes prosody as having both prosodic forms and communicative functions (Wagner & Watson, 2010; Xu, 2015). Early references to the form vs. function framework can be dated back to Pierrehumbert's 1980 thesis where she described the distinction between the formal and the functional approach of analyzing prosody. Specifically, the formal approach attempts to define components of prosody based on observed prosodic patterns, such as F0 contours, whereas the functional approach focuses on identifying prosodic patterns that convey similar or different meanings.

More recent studies have adopted the form vs. function theoretical framework to investigate prosodic deficits in ASD (Järvinen-Pasley, Peppé, King-Smith, & Heaton, 2008; Peppé et al., 2007). Peppé and colleagues developed a semi-standardized prosody assessment procedure, the Profiling Elements of Prosodic Systems in Children (PEPS-C; Peppé & McCann, 2003), to assess prosodic deficits at both form and function levels. Based on their definitions, the form level involves the ability to perceive and produce prosodic forms (i.e. auditory discrimination and vocal imitation). The function level requires the ability to cognitively and linguistically process communicative functions that prosody serves in spoken communication, including grammatical, semantic/pragmatic, and affective/emotional functions (Crystal, 1986; McCann & Peppé, 2003; Paul et al., 2005). Specifically, grammatical prosody refers to the use of prosody to signal syntactical information within linguistic units (Fry, 1955; Selkirk, 1995). For example, intonation at the end of an utterance indicates whether an utterance is a statement (falling pitch) or a question (rising pitch). Lexical stress within a word signals whether a lexicon is used as a noun (e.g. REcord) or a verb (e.g. reCORD). Semantic or pragmatic prosody can be used to enhance or modulate information beyond the literal meaning of an utterance (Bolinger, 1961; D. G. Watson, 2010). For example, contrastive pitch accent is often used to direct listener's attention to a salient information in a discourse context or to contrast a piece of information with possible alternatives (Bolinger, 1961, 1972; Pierrehumbert, 1980). The interpretation of the same sentence, "Sophie has four hedgehogs", changes as the speaker varies the location of the contrastive pitch accent. One would say, "Sophie has four HEDGEHOGS", to emphasize the type of pets Sophie owns, or "Sophie has FOUR hedgehogs", to emphasize the quantity. Lastly, affective/emotional prosody conveys a speaker's intentions or emotional states (Bolinger, 1989). For example, the utterance, "I am having so much fun", could be interpreted either as genuine or sarcastic depending on the prosodic characteristics of the utterance.

Previous studies have considered the ability to perceive prosody acoustically at the form level as a prerequisite of the ability to process prosody cognitively and linguistically at the function level (Diehl & Paul, 2012; McCann & Peppé, 2003; Peppé et al., 2007). Children with ASD have been found to demonstrate prosodic deficits at both form and function levels (Järvinen-Pasley et al., 2008; Peppé et al., 2007). Children with ASD were significantly worse at discriminating prosodic patterns in both words and utterances (Peppé et al., 2007), detectingspeakers' emotions (Peppé et al., 2007), and identifying syntactic boundary (Järvinen-Pasley et al., 2008). Intriguingly, these findings, particularly reported form-level deficits, seem to be inconsistent with other findings that reported intact performance on function-level tasks in children with ASD (Grossman et al., 2010; Paul, Augustyn, et al., 2005). These findings suggest a potential dissociation between form-level and function-level ability to process prosody in children with ASD and warrants further investigations into patterns of prosodic deficits in this population (Järvinen-Pasley et al., 2008).

2.1.1 Existing reviews of prosodic deficits in ASD

There are currently two published reviews on prosodic deficits in individuals with ASD (Fusaroli et al., 2017; McCann & Peppé, 2003). McCann and Peppé published a narrative review and summarized prosodic deficits in individuals with ASD (McCann & Peppé, 2003). The authors identified 16 studies between 1980 and 2002 and summarized the findings using a functional approach (i.e. whether participants with ASD demonstrated the ability to use certain functions of prosody effectively). It was concluded that findings were inconsistent in all areas of prosodic functions identified in the review. Additionally, it is difficult to draw conclusions on the receptive prosody in this population because only 3 out of 16 studies examined receptive prosodic abilities with each article focusing on different functions of prosody (Frankel et al., 1987; Paul, Augustyn, Volkmar, & Cohen, 2000; Rutherford, Baron-Cohen, & Wheelwright, 2002).

In addition, a recently published article reported a systematic review and meta-analysis on expressive prosody in individuals with ASD (Fusaroli et al., 2017). The authors concluded significant differences in mean pitch and pitch range between individuals with ASD and typically developing individuals. However, the scope of this review was limited to the production of prosody.

2.1.2 This current systematic review

In the past two decades, there have been more attempts to investigate receptive prosody in individuals with ASD. Given the current gap in scientific knowledge in receptive prosodic abilities in ASD, this systematic review aims to evaluate the current evidence base on prosodic deficits in individuals with ASD on both form and function levels. Specifically, this systematic review was guided by three primary research questions:

1. Do individuals with ASD demonstrate prosodic deficits on the form level (i.e. discriminate prosodic forms acoustically)?
2. Do individuals with ASD demonstrate prosodic deficits on the function level (i.e. cognitively and linguistically process communicative functions that prosody serves, including grammatical, semantic/pragmatic, and affective/emotional)?
3. On an exploratory note, is the ability to process prosody on the form- or function-level associated with broader linguistic or social functioning in individuals with ASD?

2.2 Methods

Inclusion criteria were developed *a priori* based on participant characteristics, comparison group, study design, outcome measure, and language of the article (see Table 2). Considering the paucity of research on receptive prosody in children with ASD, studies that examined children or adults with ASD or Asperger Syndrome (AS) were included to ensure sufficient articles. To be included, a study must have a typically developing comparison group, include at least one behavioral measure of receptive prosody, and be published after 1980. Case studies or single-case designs were excluded. Studies that were not written in English or assessed non-English-speaking population were excluded given that prosodic structures and patterns have been found to vary across languages (M. Beckman, 1992).

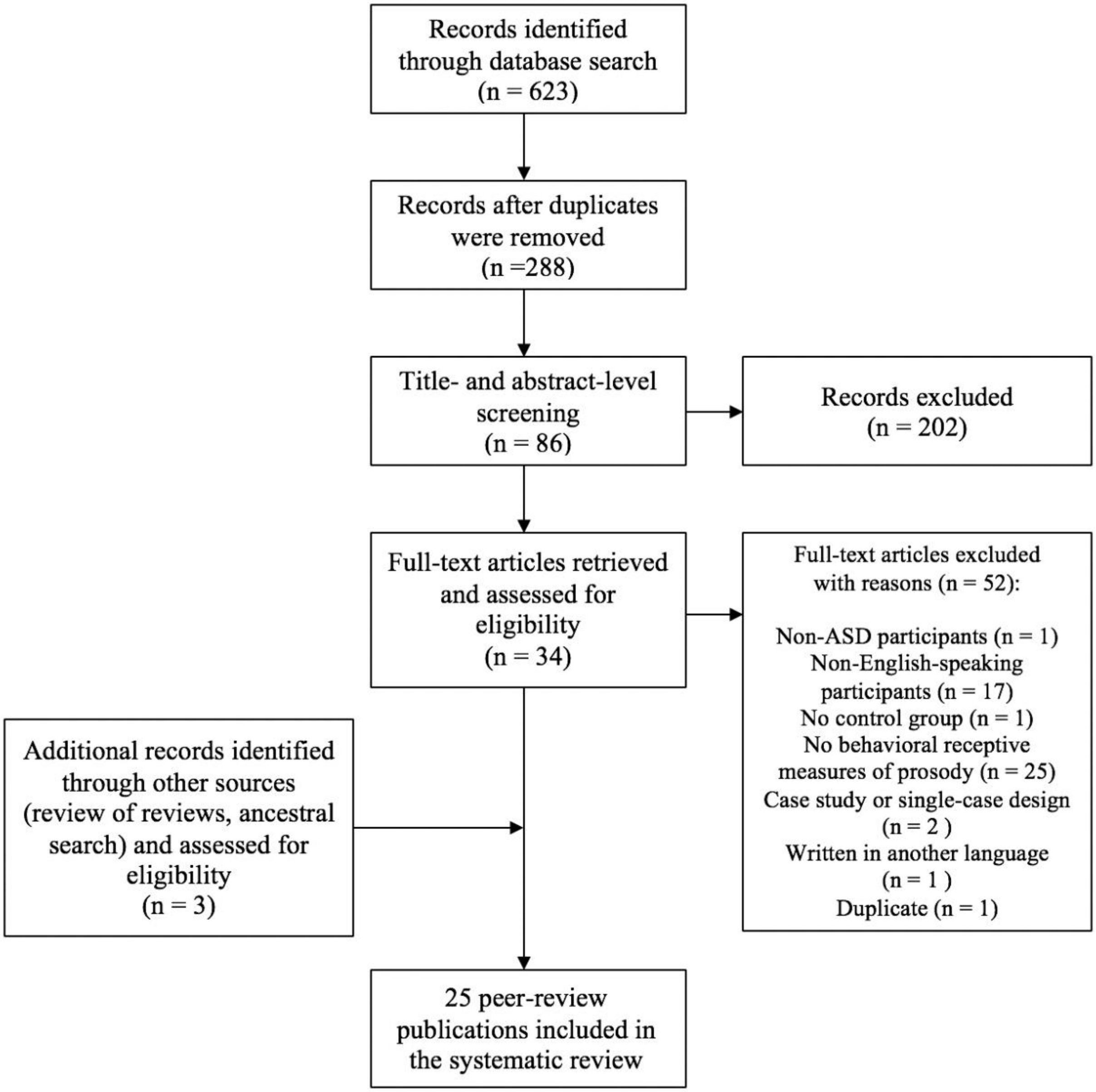
Table 2. Eligibility criteria and search terms

Aspect	Criteria
Population	Children or adults diagnosed with ASD or Asperger Syndrome
Comparison Group	Typically developing children or adults
Outcomes	At least one behavioral receptive prosody measure
Study Design	Group design Published journal articles or dissertations
Others	Published after 1980 Written in English and targeting English-speaking populations

Using these criteria, a literature search were conducted using 78 databases within ProQuest using the following syntax: ab(prosody or prosodic or intonation or suprasegmental or "pitch accent") AND ab(Autism or ASD or Autistic or Asperger). This search yielded 623 records. After duplicates were removed, 288 records remained. Next, an initial title and abstract screening reduced the number of records to 86. Then, full text study reports were retrieved for all

86 studies that passed the initial screening and were reviewed to determine eligibility. Thirty-four records passed the full-text review. Additionally, the reference lists of a previous review (McCann & Peppé, 2003) and all included studies were examined and yielded 3 additional records not identified from the database search. The final sample includes 32 unique studies (from 37 records), including 25 published journal articles and 7 dissertations and theses (Figure 1). Only published journal articles were coded and reviewed in this paper. The final sample of articles were reviewed and the following information were extracted: sample size, age, language level, cognitive level, task or paradigm used to assess receptive prosody, level of prosody assessed (form vs. function), aspect of prosody assessed (grammatical, pragmatic, or affective/emotional), and summary of findings.

Figure 1. PRISMA flow diagram for studies included in the systematic review



2.3 Results

2.3.1 Study characteristics

Table 3 provides characteristics and summarized findings of the included 25 studies. The sample size ranges from 18 to 103. Participants' mean age ranges from 5 to 36. Nineteen studies included children and six studies focused on adults. Twelve studies reported participants' language level. Eight of the twelve studies included verbal children with ASD. Four studies described their participants as having varying levels of language skills and including nonverbal children with ASD without reporting language measures. Sixteen studies reported participants' cognitive ability. Fifteen studies recruited children with ASD with cognitive functioning within the normal range and one study (Ploog, Banerjee, & Brooks, 2009) included participants with impaired cognitive functioning (i.e. an average full-scale IQ of 50 on the Stanford-Binet Intelligence Scale). Regarding the level of prosody examined, five studies focused on the form-level of prosodic processing ability, eighteen studies examined the function-level of prosody, and two studies assessed prosody at both levels.

Table 3. Summary of included studies on receptive prosody in individuals with ASD

References	N (ASD; ctrl)	Age (ASD; ctrl)	Language (ASD; ctrl)	IQ (ASD; ctrl)	Task	Form vs Function	Aspect of Prosody	Findings
Brennand, Schepman, & Rodway, 2011	15;15	14.5 (10.5 - 19.3); 13.3 (11-16.7)	NA	NA	Forced-choice paradigm: participants listened to nonsense (German) sentences and were asked to identify the emotions (anger, fear, happiness, sadness)	Function	Affective/ Emotional	The ASD and TD group showed statistically similar levels of performance identifying emotions from nonsense sentences.
Brooks, Gaggi, & Ploog, 2018	13;13	13.4 (7.1-21.3); 13.8 (7.5-21.2)	Various levels of language skills, including nonverbal children (N =7)	NA	Adapted video game paradigm: (Ploog et al., 2009): participants listened to pairs of sentences with varying contents (two different sentences) and prosody (enthusiastic vs. grouchy) in a video game paradigm and were reinforced to select one of the two in the training. Participants were then tested on stimuli with recombinations of the contents and prosodic features of training stimuli to assess participants' sensitivity to prosody and semantic contents. Participants' ability to generalize were tested using stimuli spoken by both male and female speakers.	Form	Affective/ Emotional	Performance on prosody trials was sig. lower in the ASD group than the TD group for both discrimination trials and generalization trials.

Brooks & Ploog (2013)	13;13	9 (5-17.5); 9(5.3-16)	Various levels of language skills, including nonverbal children	NA	Adapted video game paradigm (Ploog et al., 2009): Participants listened to a pair of sentences with varying content (two different sentences) and prosody (enthusiastic vs. grouchy) in a video game paradigm and were reinforced to select one of the two in the training. Participants were then tested on stimuli with recombinations of the contents and prosodic features of training stimuli to assess participants' sensitivity to prosody and semantic contents. Different from Ploog et al. (2009) which tested participant's ability to discriminate grammatical intonations, the authors tested affective prosody in this study.	Form	Affective/ Emotional	Both groups were able to discriminate between pairs of stimuli based on affective prosody. But TD children demonstrated a preference for enthusiastic stimuli over grouchy stimuli, whereas children with ASD did not demonstrate such preference.
Chevallier, Noveck, Happé, & Wilson, 2011	17;17	13.7 (11.1 - 17.8) ; 14.2 (11.6 - 16.8)	BPVS[2]: 106 (78-145); 100 (83-128)	NA	Forced-choice paradigm: Experiment 1: participants were first matched on their abilities to discriminate pitch, duration, and intensity using the Dinos task and then asked to identify speaker's manners of speech, physical states, basic emotion, social emotion, and 2nd order ToM[3] emotion in forced choices of two.	Function	Affective/Emotional	The ASD and TD group showed similar levels of performance in both accuracy and reaction time identifying all prosodic cues.

[2] BPVS = British Picture Vocabulary Scale (Dunn, Dunn, Whetton, Burley, 1997)
[3] ToM = Theory of Mind

Chevallier, Noveck, Happé, & Wilson, 2009	17;17	13.7 (11.1 - 17.8) ; 14.2 (11.6 - 16.8)	BPVS: 106 (78-145); 99 (76-128)	NA	Forced-choice paradigm: Participants were matched on their perceptual ability to discriminate pitch, duration, and intensity using the Dinos tasks and then asked to identify correct lexical stress patterns in Experiment 1, identify correct pause patterns in Experiment 2, and decide whether the speaker is sure or unsure after listening to a question or a statement in Experiment 3.	Function	Grammatical	Adolescents with AS are able to decide on the appropriate lexical stress patterns, correctly chunk compounds on the basis of rhythmic and pause cues, and differentiate questions from statements based on intonation contours with comparable accuracy and reaction time as the control group.
Diehl, Bennetto, Watson, Gunlogson, & McDonough, 2008	21;22	15.3 (11-19); 15.3(11-19)	CELF4[4] Receptive Index: 106 (86 - 128); 105 (86-119)	Measured by WISC-IV[5] or WAIS-III[6]: 112 (88 - 131); 111 (94-124)	Syntactic ambiguity paradigm: participants were instructed to complete syntactically ambiguous demands that can be disambiguated by prosody, syntax, or prosody and syntax.	Function	Grammatical	Participants with ASD were significantly worse than controls at using prosody alone to disambiguate sentence meanings but performed similarly to controls when judging meaning based on syntax only or congruent prosody and syntax.
Diehl, Friedgberg, Paul, & Snedeker, 2015	48;48	12.7 (7-17);12.7 (7-17)	Verbal (measured by CELF4 Receptive Index): 114 (85 - 151); 114 (89-136)	Measured by WISC-IV or the Differential Ability Scales: 113.3 (88-148); 113.8 (88-136)	An eye-gaze syntactic ambiguity paradigm (Snedeker & Yuan, 2008): participants were instructed to complete syntactically ambiguous demands that can be disambiguated by prosody. Different from Diehl et al., 2008 which used intermixed prosodic patterns, this paradigm uses a block design with one prosodic pattern in each of the two blocks.	Function	Grammatical	Children and adolescents with ASD succeeded at using prosodic information to resolve syntactic ambiguity in block 1. But children with ASD (aged 7-12) demonstrated chance performance on block 2, suggesting that they were less able to overcome learned prosodic patterns from block 1.

[4] CELF-4: Clinical Evaluation of Language Fundamentals – Fourth Edition

[5] WISC-IV: Wechsler Intelligence Scale for Children - Fourth Edition (Wechsler, 2008)

[6] WAIS-III: Wechsler Adult Intelligence Scale – Third Edition (Wechsler, 1997)

Diehl & Paul, 2012	24;22	12.3 (8-16); 13 (9-16)	CELF4 Core language: 97.21 (67-132); NA in the control group	Nonverbal IQ measured by the WASI[7] or the Differential Ability Scale: 103.61(75-133);NA in the control group	Form tasks from PEPS-C: Participants were asked to discriminate pairs of word, phrase, or sentence stimuli that vary in prosodic patterns.	Form	NA	The ASD group performed marginally worse at discriminating word-level prosodic features and significicantly worse at discriminating sentence-level prosodic features than the TD group. The ASD group did not differ significantly from the LD group at either tasks.
Erwin, Van Lancker, Guthrie, Schwafel, Tanguay, Buchwald, 1991	11;14	25.7 (17-39); 23.4 (20-30)	NA	97 (76-135); NA in the control group	Forced-choice paradigm: participants were presented pairs of stimuli with varying prosodic cues (raising intonation vs. falling intonation; angry vs. happy) and were asked to match the stimuli to an appropriate picture/word. Participants were also presented with sentences with varying affective prosody (anger, happiness, surprise, sadness) and instructed to identify the emotion in a forced-choice of two.	Function	Grammatical Affective/ Emotional	The ASD group performed at a comparable level on the behavioral level with the TD group.
Golan, Baron-Cohen, Hill, & Rutherford, 2007	40;26	27.5 (17-50); 24.3 (17-50)	NA	Full scale IQ measured by the WASI: 113.82 (92-138) ; 114.45 (97-138)	Forced-choice paradigm: Reading the Mind in the Voice - Revised (RMV-R) Task: participants listened to sentences with varying emotions and were asked to identify the emotion in forced-choices of four.	Function	Affective/ Emotional	Participants in the AS/ASD group performed significantly worse on the RMV-R task than the control group.

[7] WASI: Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999)

Grossman, Bemis, Skwerer, & Tager-Flusberg, 2010	16;15	12.3 (7.5-17); 12.6 (7.5-18)	PPVT-III: 107 (79-138); 111 (79-139)	Full-scale IQ measured by the K-BIT2[8]: 106.7 (87-123); 108.9 (87-123)	Forced-choice paradigm: in the affective prosody task, participants listened to sentences with varying emotions (happy, sad, neutral) and were asked to identify the emotion in forced-choices of three. In the lexical stress task, participants were asked to use lexical stress to disambiguate word/phrase pairs (e.g. hotdog vs. hot dog)	Function	Grammatical Affective/ Emotional	Participants with ASD performed similarly to the TD group on both affective prosody and lexical stress tasks.
Hubbard, Faso, Assmann, & Sasson, 2017	22;30	25.9 (18-50); 22.5 (18-50)	NA	111.3 (88-129); NA in the control group	Forced-choice paradigm: participants listened to sentences read by TD speakers and speakers with ASD and identified the emotion context (neutral, angry, happy, interested, sad).	Function	Affective/ Emotional	Participants with ASD performed worse at identifying neutral and happy sentences than TD participants
Järvinen-Pasley, Peppé, King-Smith, & Heaton, 2008	Exp1: 21;21 Exp2: 20;20	Exp 1: 12.6 (7.7-16.8); 12.21 (8.3-16.3) Exp 2: 12.0 (7.3 - 16.4); 11.9 (7.5-16.1)	Exp 1: BPVS: 84 (55-135); 87 (58-124) Exp2: BPVS: 88 (55-135); 87 (55-124)	Exp 1: RSPM[9] 89 (61-119); 88 (61-121) Exp 2: 92 (61-129); 86 (62-121)	PEPS-C: In Experiment 1, six receptive tasks in the PEPS-C were used. In Exp 2, participants listened to sentence with either raising or falling intonation and selected between *asking* and *telling*.	Form and function	Grammatical Pragmatic Affective/ Emotional	Exp 1: participants with ASD performed significantly worse at phrase- or sentence-level discrimination, affective intonation, and chunking/phrasing. Exp 2: participants with ASD were sig. less accurate at judging whether speakers sound certain from questions than TD participants.

8 KBIT-2: Kaufman Brief Intelligence Test, Second Edition (Kaufman & Kaufman, 2004)
9 RSPM: Raven Standard Progressive Matrices (Raven et al., 1992)

Kargas, López, Morris, & Reddy, 2016	21;21	30.3; 29.5	NA	Full-scale IQ measured by the WASI: 109.5; 115.9	Forced-choice paradigm: participants listened to pairs of four-syllable words with either first syllable stress (e.g. dandelion) or second syllable stress (e.g. capacity) and asked to make same-different judgement.	Form	Grammatical	The ASD group demonstrated significantly lower sensitivity on detecting lexical stress than the TD group.
Lindner & Rosén, 2006	14;16	10.21 (5-16); 10.19 (5-16)	PPVT-III[10]: 107.57(96-136);111.13 (99-136)	NA	Forced-choice paradigm: Perception of Emotion Test (PET; Egan, 1989): participants listened to or watched 160 posed scenes and were asked to label emotions (happy, angry, sad, neutral) by pointing to the appropriate picture. The 16 trials that target prosodic perception were presented in audio format and have neutral semantic contents with varying intonations to indicate affective prosody.	Function	Affective/ Emotional	Participants with ASD performed significantly worse at prosodic scenes compared to the TD participants.
Lyons, Simmons, & Paul, 2014	87;43	Preteen age subgroup: 10.73; 11.18 teen age subgroup: 15.25; 14.94	CELF4 Expressive Language: preteen age group: 100.14; 109.41 teen age group: 94; 104.5	Verbal IQ measured by the WASI or the DAS-II[11]: preteen age group: 103.9; 109.4 teen age group: 105.7; 107.9	The Prosody Protocol (PP) contains eight subtests that examine participants' ability to perceive and produce intonation (question vs. statement), stress, phrasing (pause), and affect.	Function	Grammatical Pragmatic Affective/ Emotional	Significant differences were detected in the preteen group in intonation, contrastive pitch accent task, and global receptive score (TD ~ ASD High-language > ASD Low-language).

[10] PPVT-III: Peabody Picture Vocabulary Test – Third Edition (Dunn & Dunn, 1997)
[11] DAS-II: Differential Abilities Scale-II, School-Age Version (Elliott, 1990)

Paul, Augustyn, Klin, & Volkmar, 2005	27;13	16.8; 16.7	CELF3 Receptive Score: 98.6; not reported in the TD group	Verbal IQ measured by the WISC-III: 103.9; not reported in the control group	All participants took part in 6 receptive prosody tasks (3 on grammatical prosody and 3 on pragmatic/affective prosody) that examine comprehension of intonation, stress, and phrasing.	Function	Grammatical Pragmatic Affective/ Emotional	Significant differences were detected only in the contrastive pitch accent (pragmatic stress) task (p=.004) but not in the rest five receptive prosody tasks.
Peppé et al., 2017 (McCann et al., 2007)	31;72	9.45 (6-13); 6.45, 4-11	BPVS-II: 81.4 TROG-II[12]: 79.6 CELF3: 69.8 GFTA-II[13]: 93.3 CCC[14] (raw score): 123 Language not reported in the TD group	RM[15]: 96.4 IQ not reported in the TD group	PEPS-C tasks	Form and function	Grammatical Pragmatic Affective/ Emotional	Significant differences were detected between the ASD and the TD group in word-level discrimination, phrase-/sentence-level discrimination, and affect perception, but not in turn-end, chunking, or focus.

12 TROG-II: Test for Reception of Grammar – Second Edition (Bishop, 1989)
13 GFTA-II: Goldman-Fristoe Test of Articulation – Second Edition (Goldman & Fristoe, 2000)
14 CCC: Children's Communication Checklist (Bishop, 1998)
15 RM: Raven's Matrices (Raven, Court, & Raven, 1986)

Ploog, Brooks, Scharf, & Aum, 2014	15;15	10.6(5.4-18); 10.8(5.2-16)	Various levels of language skills, including nonverbal children	NA	Adapted video game paradigm: (Ploog et al., 2009): English-speaking participants listened to a pair of nonsense sentences with varying contents and prosody (statement vs. question, enthusiastic vs. grouchy) in a video game paradigm and were reinforced to select one of the two in the training. Participants were then tested on stimuli with recombinations of the contents and prosodic features of training stimuli to assess participants' sensitivity to prosody and semantic contents.	Form	Grammatical Affective/ Emotional	Both groups discriminated reinforced stimuli based on their prosody at equivalent levels of accuracy.
Ploog, Banerjee, & Brooks, 2009	9;9	12.8 (5.3-18.3); 8 (5.3-11.9)	Medium- to low-functioning of language, including nonverbal children	Full-scale IQ measured by the Stanford-Binet: 50.43 (40-74)	Adapted video game paradigm: Participants listened to a pair of sentences with varying contents and prosody in a video game paradigm and were reinforced to select one of the two in the training. Participants were then tested on stimuli with recombinations of the contents and prosodic features of training stimuli to assess participants' sensitivity to prosody and semantic contents.	Form	Grammatical	No sig. difference was detected between the ASD and the TD group on test probes that differ only in prosody, which could be attributed to unexpectedly poor performance in the TD group.
Rutherford, Baron-Cohen, Wheelwright, 2002	19;20	29 (16-59); 36 (18-53)	NA	Full-scale IQ measured by the WAIS: 107.93 (87-133); 101 (91-116)	Forced-choice paradigm: Reading the Mind in the Voice Test: Participants listened to 2-sec segments recorded from dramatic audio books and asked to identify the speaker's mental attitude or emotion in forced-choices of two adjectives.	Function	Affective/ Emotional	Participants with ASD performed significantly worse at identifying mental states or emotions than the control group.

Singh & Harrow, 2014	10;10	10.6 (8.1 -13); 10.5 (8.7 - 12)	CELF4: 107.4;117.5 EVT[16]: 111;114 PPVT: 115;120	NA	Forced-choice paradigm: participants listened to emotional words varying in affective prosody and were asked to identify the emotion based on the tone of voice in forced choices of two (happy vs. sad).	Function	Affective/ Emotional	Participants with ASD are equally accurate in classifying emotions as TD participants. When reaction times were analyzed, an interaction was detected between group and semantic-prosodic congruence between semantic contents and prosodic pattern: TD children showed significantly slower responses to incongruent stimuli than congruent, whereas children with ASD did not.
Steward, McAdam, Ota, Peppé, Cleland, 2013	11;14	27.2 (17-39); 26.4 (21-37)	NA	Mill Hill Vocab Scale from the Raven: 14.9 (6-25); 18.1 (8-23)	Forced-choice paradigm: participants listened to vocalizations of 'mmm' and sentences with neutral prosody, congruent prosody with semantic content, or incongruent prosody with semantic contents and were asked to identify emotions in a forced-choice of five emotions (anger, fear, happiness, surprise, disgust)	Function	Affective/ Emotional	Participants with ASD performed sig. worse at identifying emotions from incongruent sentences, neutral sentences, and vocalizations of 'mmm', but not at identifying congruent sentences.

16 EVT: Expressive Vocabulary Test (Williams, 1997)

Van Lancker, Cornelius, Kreiman, 1989	28;33	Younger subgroup: 6.9 (4 - 7.9); 5.3 (3.3-7.8) Older subgroup: 11.3 (8 -22); 11.2 (8.2-17)	NA	NA	Forced-choice paradigm: participants listened to semantically neutral sentences and were asked to first select an image that represents the correct semantic content and then select an image that represents the emotion based on the tone of voice.	Function	Affective/ Emotional	Participants with ASD in both age subgroups performed sig. worse than TD controls on semantic tasks. Children with ASD in the older group also performed sig. worse at emotion recognition task.
Wang, Lee, Sigman, Dapretto, 2006	18;18	11.9; 11.9	NA	Full-scale IQ measured by the WAIS: 102; 106	Forced-choice paradigm: participants listened to short scenarios and were asked to decide whether the speaker was sincere or ironic.	Function	Affective/ Emotional	Behavioral data reveal that participants with ASD and TD participants did not differ significantly when only prosodic cues were available. However, the ASD group showed heightened recruitment of temporal regions bilaterally in the prosody-only condition.

2.3.2 Tasks, paradigms, and main findings

Studies adopted a variety of tasks and paradigms: 14 studies used a forced-choice task where participants listened to a stimulus and were asked to make a forced choice out of two to five choices; four studies used or adapted a video game paradigm with a discrimination-choice procedure embedded in the game to test prosody discrimination; three studies used the Profiling Elements of Prosodic Systems in Children (PEPS-C; Peppé & McCann, 2003); one study used an eye-tracking paradigm; and three studies used other linguistic tasks. A narrative summary and evaluation of the paradigms used and study findings are provided below to supplement the summaries provided in Table 3.

Forced-choice tasks. Fourteen studies used a forced-choice task. In some studies, participants were played one stimuli and were asked to identify the speaker's emotion among two to five emotions. In other studies, participants were played a pair of or more than two stimuli and were asked which one would correspond to a specific picture. One study examined form-level of receptive prosody and 13 studies examined function-level. Within the studies on function-level of receptive prosody, 12 studies focused on affective/emotional prosody and 3 studies examined grammatical prosody[17].

Kargas et al. (2016) is the only study that investigated perception of grammatical prosody on the *form-level* using a forced-choice task. In this study, participants listened to word pairs that differ only in the stress patterns and were instructed to make same-different judgments about each pair. The ASD group performed significantly worse at detecting lexical stress than the TD group. However, this group difference appears to be driven by a subgroup of participants with ASD (N=7) who demonstrated marked poor sensitivity to lexical stress in a relatively

[17] The number exceeds 13 because two studies examined both grammatical and affective prosody understanding (Erwin et al., 1991; Grossman et al., 2010).

homogenous sample of individuals with ASD (N = 21) with similar levels of ASD severity and cognition. The authors thus called for future studies to investigate potential subgroups within ASD to understanding the considerable variability in language and communication reported in this population.

The three studies on grammatical prosody using forced-choice tasks tested participants' ability to understand meaningful grammatical differences embodied by prosodic cues (Chevallier, Noveck, Happé, & Wilson, 2009; Erwin et al., 1991; Grossman et al., 2010). In all three studies, no significant difference was detected between TD and ASD groups, suggesting that comprehension of grammatical prosody may be a relatively intact area of receptive prosody in individuals with ASD. In one study (Erwin et al., 1991), the authors also collected event-related potential data and found that participants with ASD displayed similar P3 responses to all prosodic stimuli. This finding provided additional evidence that individuals with ASD can comprehend and process grammatical prosody at both behavioral and neurological levels.

Within the 12 studies that used forced-choice tasks to test comprehension of affective/emotional prosody, six studies reported null findings between groups (Brennand, Schepman, & Rodway, 2011; Chevallier, Noveck, Happé, & Wilson, 2011; Erwin et al., 1991; Grossman et al., 2010; Singh & Harrow, 2014; Wang, Lee, Sigman, & Dapretto, 2006) and six studies reported significant differences between TD and ASD groups (Golan, Baron-Cohen, Hill, & Rutherford, 2007; Hubbard, Faso, Assmann, & Sasson, 2017; Lindner & Rosén, 2006; Rutherford et al., 2002; Stewart, McAdam, Ota, Peppé, & Cleland, 2013; Van Lancker, Cornelius, & Kreiman, 1989). Despite inconsistent findings across these studies, several common patterns emerged. First, in two studies, group differences were detected in speed related measures rather than accuracy measures (Chevallier et al., 2011; Singh & Harrow, 2014), which suggests that individuals with ASD may be capable of processing and interpreting affective prosody but may be slower in their processing or may have difficulty using them in challenging situations. Second, two studies suggested that individuals with ASD may rely on semantic content as a compensatory strategy to detect a speaker's affect (Lindner & Rosén, 2006; Singh & Harrow, 2014). In both studies, participants performed better at identifying affects from sentences with congruent semantic content compared to neutral sentences or vocalizations without semantic contents. Additionally, two studies provided direct evidence that poor performance on receptive prosody tasks in participants with ASD were driven by low language comprehension abilities (Brennand et al., 2011; Van Lancker et al., 1989). For example, in Brennand et al. (2011), the between-group difference on affective prosody understanding was no longer significant once language was included as a covariate.

This group of forced-choice method studies also shared some common limitations. First, in several studies, participants' language ability was not measured or matched between groups. Thus, affective prosody understanding difficulties observed in the ASD group may be confounded with potential language deficits. Moreover, findings in some studies were influenced by the limits of the tasks used. For example, in some studies, the tasks had not been validated in typically developing populations. In two studies that reported null findings between groups (Brennand et al., 2011; Hubbard et al., 2017), participants in both TD and ASD groups performed close to chance level, indicating that the lack of group differences may be attributed to ambiguous experimental stimuli. In other studies, the task used pose significant cognitive or language demands on participants. In particular, in one study, participants were instructed to detect affects in sentences while counting and reporting the number of times the letter *T* was present in the sentence.

PEPS-C. Three studies[18] used the PEPS-C to examine prosody in children with autism. The PEPS-C is a semi-standardized assessment designed to assess the expressive and receptive prosody in children and adolescents (Peppé & McCann, 2003). It adopts the form vs function framework and assesses prosody on both levels. The most current version of PEPS-C (2015) consists of 14 tasks, including two tasks on the form level that measure speakers' ability to discriminate and imitate prosodic variations and 12 tasks on various functions of prosody. The 12 tasks assess six functions (turn end, affect, lexical stress, phrasal stress, chunking/phrasing, and contrastive pitch accent) each in a receptive and an expressive task. Each task consists of 16 items and a score of > 12 was used to represent competence level on each task. All three included studies used an earlier version of the PEPS-C (PEPS-C Research and Clinical version, Peppé & McCann, 2003) (Diehl & Paul, 2012; Järvinen-Pasley et al., 2008; Peppé et al., 2007). The PEPS-C Research and Clinical version includes a subset of tasks in PEPS-C 2015. Table 4 presents detailed descriptions of all tasks included in both versions of PEPS-C. Of these three studies, one study (Diehl & Paul, 2012) only reported results from two form-related tasks and two studies (Järvinen-Pasley et al., 2008, Peppé et al., 2007) reported both form- and function-related tasks.

[18] Two included publications reported on the same group of participants (McCann et al., 2007; Peppé et al., 2007). Therefore, findings in both publications were consolidated together in this study under Peppé et al., 2007.

Table 4. PEPS-C task descriptions for both PEPS-C research and clinical version and PEPS-C 2015

Task Name	Mode	Level	PEPS-C Research & Clinical Version	PEPS-C 2015	Description
Discrimination	Receptive	Form	×	×	Discriminating prosodic differences in words, phrases, and sentences. Participants heard pairs of low-pass filtered stimuli and were asked to judge whether each pair sounds the same of different.
Imitation	Expressive	Form	×	×	Imitating different forms of intonation in words, phrases, and sentences.
Turn-end Understanding	Receptive	Function	×	×	Understanding questioning vs. declarative intonation. Participants heard single words produced either in a rising or a falling intonation and were asked whether the speaker was asking or telling/reading.
Turn-end Expression	Expressive	Function	×	×	Producing single words with questioning or declarative intonations.
Affect Understanding	Receptive	Function	×	×	Understanding affect (likes vs. dislikes) on single words. Participants heard words of food items with various intonation patterns were asked whether the speaker likes or dislikes the item.
Affect Expression	Expressive	Function	×	×	Producing single words with affective intonations to suggest either liking or disliking of items.
Lexical Stress Understanding	Receptive	Function		×	Perceiving the position of stress in two-syllable words. Participants heard noun-verb homographs (e.g. REcord, reCORD) and were asked to indicate where the stress is in each word.
Lexical Stress Expression	Expressive	Function		×	Producing the correct stress-pattern in two-syllable words. Participants were shown noun-verb homographs one at a time and were asked to read them aloud.
Phrasal Stress Understanding	Receptive	Function		×	Distinguishing between two-word phrases and compound nouns. Participants heard phrases that can be disambiguated using stress (e.g. green HOUSE vs. GREENhouse) and were asked to indicate which one they heard.
Phrasal Stress Expression	Expressive	Function		×	Producing different stress patterns to indicate two-word phrases and compound nouns (e.g. green house vs. greenhouse, hot dog vs hotdog).
Boundary (Chunking) Understanding	Receptive	Function	×	×	Understanding syntactically ambiguous phrases disambiguated by prosody (e.g. fruit, salad, and milk vs. fruit salad and milk). Participants heard phrases with various pause patterns and were asked to point to a picture of what they heard.

Boundary (Chunking) Expression	Expressive	Function	×	×	Producing syntactically ambiguous phrases unambiguously using prosodic patterns. Participants were shown pictures that either show two or three items (e.g. fruit, salad, and milk vs. a bowl of fruit salad and milk) and were asked to label each picture.
Contrastive Stress (Focus) Understanding	Receptive	Function	×	×	Understanding contrastive pitch accent in context. Participants were told that the speaker went to a shopping mall to buy some socks but only realized that she forgot a pair of socks in a specific color. Participants then heard sentences such as "I wanted blue and BLACK socks" and were asked to indicate which color the speaker forgot.
Contrastive Stress (Focus) Expression	Expressive	Function	×	×	Producing contrastive pitch accent in context. Participants were presented with some incorrect statements and were asked to correct the speaker using contrastive pitch accents.

Regarding form-level task performances, all three studies using the PEPS-C provided convergent evidence that children with ASD performed worse than TD children. Specifically, in both Diehl & Paul (2012) and Järvinen-Pasley et al. (2008), significant differences were found in long-item discrimination (i.e. sentence-level prosodic patterns) but not short-item discrimination, whereas Peppé et al. (2007) detected worse performance in the ASD group in both long-item and short-item discrimination. With respect to function-level task performances, two studies reported divergent findings. in Peppé et al. (2007), children with ASD only performed significantly worse than TD children in affect identification. In contrast, Järvinen-Pasley et al.(2008) reported significant between-group differences in affect identification (affective prosody), chunking (grammatical prosody), but not in contrastive stress (semantic/pragmatic) or turn-end intonation (grammatical prosody). Across form and function tasks in Järvinen-Pasley et al. (2008), given that children with ASD had more success on short-item discrimination, contrastive stress, and turn-end intonation tasks which all assess receptive prosody on the word level, the authors suggested that children with ASD may have a specific difficulty with perceiving and comprehending prosodic changes over long linguistic stimuli. Additionally, Diehl and Paul (2012) reported that significant differences in discriminating sentence-level prosodic patterns were only found between TD and ASD but not between ASD and Learning Disabilities. This finding echoed patterns found from the forced-choice tasks that prosodic differences observed in children with ASD may be driven by language or cognitive abilities.

In summary, two patterns were observed in all three studies that used PEPS-C. First, for both form- and function-level tasks, children with ASD appeared to exhibit more difficulty perceiving and comprehending prosody in sentence-level stimuli compared to word-level stimuli. Additionally, function-level prosodic deficits in ASD may be specific to affective/emotional prosody. Comprehension of grammatical intonation and pragmatic contrastive pitch accent may be two aspects of prosody that are relatively intact in this population (Table 5).

Table 5. Comparison of findings in included studies that used PEPS-C to assess the perception and comprehension of prosody in individuals with ASD

	Diehl & Paul, 2012	**Järvinen-Pasley et al., 2008**	**Peppé et al., 2007**
Short-item Discrimination	×	×	√
Long-item Discrimination	√	√	√
Turn-end Understanding		×	×
Affect Understanding		√	√
Boundary (chunking) Understanding		√	×
Focus Understanding		×	×

Note: A check indicates that a significant difference was detected between the TD and the ASD group. A cross indicates that no significant difference was detected. Consistent findings across all studies were bolded.

Video game paradigm. In a series of four studies, Ploog, Brooks, and colleagues tested form-level receptive prosodic abilities in children with ASD (Brooks, Gaggi, & Ploog, 2018; Brooks & Ploog, 2013; Ploog et al., 2009; Ploog, Brooks, Scharf, & Aum, 2014). In these studies, a discrimination forced-choice procedure was embedded in a video game where participants listened to pre-recorded sentences varying in semantic contents and prosodic patterns. During the training phase, participants were presented with 36 trials where they were reinforced to select one of two sentences differing in both contents and prosody (e.g. "Max ate a grape" said in a rising intonation vs. "Tom threw a ball" said in a falling intonation). Once a participant responded 75% or more to the reinforced stimuli in the training session, the test phase was activated. During the test phase, participants listened to pairs of stimuli with recombinations of the semantic and prosodic features (e.g. "Max at a grape" and "Tom threw a ball" produced in both rising and failing intonations) and were prompted to choose one from each pair. Participants' performance on the task was analyzed using the accuracy rate of identifying the training stimuli features from the recombined test stimuli. All four studies recruited children with varying language functioning, including minimally-verbal children with ASD. Overall, this series of studies validated the use of a video game paradigm to test form-level prosody processing ability in a sample of children with ASD with low- to medium-levels of language abilities. Yet, in all four studies, children with ASD required twice or three times more training sessions than TD children did to reach the 75%-threshold to activate the test phase. After reaching the threshold, they were able to discriminate grammatical and affective prosody as accurately as TD controls.

Eye-tracking paradigm. Diehl, Friedberg, Paul, & Snedeker (2015) is the only included study that used an eye-gaze paradigm. However, an earlier study, Diehl, Bennetto, Watson, Gunlogson, & McDonough (2008) used the same task without the eye-tracking feature and thus these two studies were reviewed together here. Diehl and colleagues (2008) adapted a psycholinguistic paradigm, syntactic ambiguity paradigm (Kraljic & Brennan, 2005; Snedeker & Trueswell, 2003), to test grammatical prosody comprehension in age- and IQ-matched TD adolescents and adolescents with ASD. Participants were presented with ambiguous sentences that are either disambiguated by prosody (e.g. "Put the dog // in the basket on the star" or "Put the dog in the basket // on the star"), syntactic structure (e.g. "Put the dog in the basket that's on the star" or "Put the dog that's in the basket on the star"), or both (e.g. "put the dog // in the basket that's on the star" or "put the dog that's in the basket // on the star"). The first sentence in each of the examples indicates that the prepositional phrase (PP) was attached to the verbal phrase (VP attachment interpretation) and signals the destination of the verb movement (e.g. put), whereas the second sentence in each of the examples indicates that the PP attaches within the noun phrase (NP attachment interpretation) and modifies the head noun (e.g. dog).

Participants were instructed to listen to sentences and follow instructions to manipulate real objects in front of them. Participants' responses were analyzed to calculate accuracy rate in each condition (Prosody-only, Syntax-only, and Prosody + Syntax) and overall. Results indicate that adolescents with ASD were significantly worse than controls only in the Prosody-only condition, which indicates that adolescents with ASD experienced difficulties using prosodic cues for syntactic processing. One limitation suggested by the authors was that some verbs used in the study (e.g. put) may have a lexical bias toward a VP-attachment interpretation because these verbs are commonly followed by a NP and a locative PP. The authors speculated that participants with ASD may have performed poorly on this task due to difficulties overriding this lexical bias and persisting in the assumed VP attachment interpretation. If this was true, then the between-group differences would reflect a deficit in cognitive flexibility rather than prosodic comprehension deficits.

Following Diehl et al. (2008), Diehl and colleagues (2015) used less-biased stimuli in the same syntactic ambiguity paradigm and added an eye-gaze feature to the task to capture participants' efficiency at using prosodic cues. Four groups of children (TD and ASD groups each with two age subgroups, children and teens) were told that they would play a game about following instructions. Participants then listened to syntactically ambiguous commands (e.g. "You can feel the frog with the feather"). The prosody of these commands were manipulated by either placing a prosodic boundary cue (i.e. pause and boundary intonation) before the first noun phrase (e.g. "You can feel // the frog with the feather") or before the prepositional phrase (e.g. "You can feel the frog // with the feather"). In front of them were a set of toys, including a target instrument (e.g. a feather), a target animal (e.g. a frog), an animal holding an instrument (e.g. a frog with a feather), a distractor instrument (e.g. a candle which was not mentioned in the command), and a distractor animal holding a distractor instrument (e.g. a leopard holding a candle). A block design was used so that half of the participants heard one type of prosody in the first block and the other half heard the other type of prosody, which was then reversed in the second block. The block design allows the authors to test whether the difficulties children with ASD experienced in the Prosody-only condition was due to cognitive inflexibility or true prosodic deficits. Participants' eye gaze and action were coded from videotapes of the sessions by trained coders. Interestingly, different patterns emerged in Block 1 and Block 2. On Block 1, all groups demonstrated a robust effect of prosody in their responses with no significant difference between groups, which suggests that children with ASD were sensitive to prosodic boundary cues. However, on Block 2, interactions were detected between prosody and age and between prosody and diagnostic groups. Though teens with ASD performed as well as TD teens, children with ASD were at chance in the second block, which suggests that they experienced challenges interpreting the new prosodic patterns in Block 2. These findings collectively confirm that children with ASD were able to use grammatical prosodic cues to parse sentences but may have difficulty overriding learned expectations when prosodic pattern changes. This performance pattern was only evident in the younger group (7- to 12-year-olds), which may reflect immature executive function, specifically cognitive flexibility, more so than global receptive prosody deficits. Though it remains unclear the specific role that executive function plays in prosody processing in everyday spoken communication where prosody patterns vary constantly to convey syntactic, semantic, or pragmatic meanings, this study identifies and highlights the need to consider cognitive factors that lead to poor performance on prosody tasks.

Other linguistic tasks. Two studies used linguistic tasks to assess function-level of prosody. Paul et al. (2005) examined understanding of grammatical, pragmatic, and affective prosody in six receptive tasks. Lyons, Simmons, & Paul (2014) used four receptive tasks adapted from PEPS-C, Paul et al. (2005), and Hubbard & Trauner (2007) to test understanding of grammatical, affective prosody understanding. Paul et al. (2005) included high-functioning adolescents with ASD and concluded that adolescents with ASD were worse than TD adolescents in comprehension of contrastive pitch accent. In Lyon et al. (2014), participants with ASD were divided into two age groups (9-12 and 13-17), with each age group further divided into two language groups (language impaired and typical language). Notably, significant differences were only detected in grammatical intonation and contrastive pitch accent in the younger age group. Additionally, within the younger group, children with ASD with higher language levels performed similarly to TD peers whereas children with lower language levels displayed marked receptive prosodic deficits. The presence of receptive prosodic deficits only in the younger age group seems to suggest that the nature of receptive prosodic deficits in this population may be more quantitative than qualitative. Overall, these findings indicate that receptive prosodic deficits may be most evident in a subgroup of individuals with ASD.

2.3.3 Research question 1: Form-level deficits in ASD

Eight of the 25 studies reported form-level prosodic functioning findings in individuals with ASD (Table 6). Overall, five studies (63%) concluded that participants with ASD demonstrated form-level receptive prosodic deficits. Three of the four studies (Brooks & Ploog, 2013; Ploog et al., 2009, 2014) that used the video game paradigm did not find a significant difference in participants' ability to discriminate prosodic patterns (rising vs. falling intonation, enthusiastic vs. grouchy affective prosody) between groups, whereas one study (Brooks et al., 2018) did detect a significant between-group difference. All three studies (Diehl & Paul, 2012;

Järvinen-Pasley et al., 2008; Peppé et al., 2007) that used the PEPS-C found that children with ASD performed significantly worse on form-level tasks than TD children. Between-group differences were particularly evident on discrimination of prosodic patterns on phrase- or sentence-level items. One study (Kargas et al., 2016) used a forced-choice paradigm (same-different discrimination task) and found that participants with ASD were significantly worse at differentiating lexical stress patterns.

2.3.4 Research question 2: Function-level deficits in ASD

Seven of the 25 studies examined function-level prosodic abilities in individuals with ASD, with nine studies on grammatical prosody comprehension, four studies on pragmatic prosody comprehension, and 15 studies on affective/emotional prosody comprehension (some studies examined multiple functions, and thus the total number exceeds 25). The results were largely inconsistent across studies. Regarding grammatical prosody comprehension, three studies (33%) reported significant between-group differences, yet five did not. In terms of pragmatic prosody, half of the studies (N = 2, 50%) concluded that individuals with ASD demonstrated significant deficits, whereas the other half (N = 2, 50%) found comparable performances in TD and ASD groups. The majority of studies that examined function-level prosody processing focused on affective/emotional prosody comprehension. However, the findings remain contradictory: eight studies (53%) reported significant between-group differences on various emotion recognition tasks, whereas seven studies (47%) reported null findings regarding affective/emotional prosody processing.

2.3.5 Research question 3: Relations between receptive prosody and other measures

Findings from ten studies that examined associations between receptive prosody on form- or function-levels and broader linguistic or social communication functioning were also extracted (Table 7). These studies reported significant correlations between participants' performance on receptive prosody tasks and other skills and characteristics, including expressive prosody (Peppé et al., 2007), receptive language (Diehl et al., 2008; Peppé et al., 2007), speech production (Kargas et al., 2016), vocabulary (Peppé et al., 2007), overall language (Diehl & Paul, 2012; Peppé et al., 2007), social communication skills (Singh & Harrow, 2014), verbal IQ (Golan et al., 2007), full IQ (Hubbard et al., 2017; Peppé et al., 2007), age (Lindner & Rosén, 2006; Peppé et al., 2007), and autism symptomology (Golan et al., 2007). Additionally, Peppé et al. (2007) found that participants' scores on form- and function-level tasks produced different correlation patterns. Though both correlated with general language ability measured by CELF-3, participants' form-level task composite score correlated with age whereas function-level task composite score correlated with vocabulary and receptive language.

Other studies reported contradictory findings. Though Golan and collages (2017) found significant correlations between participants' performance on the emotion recognition task and Autism Quotient scores and verbal IQ measured by WASI, they did not find significant correlations between emotion recognition and age or performance IQ. Contrary to Kargas et al (2016) who found significant associations between receptive prosody and speech abnormalities, Peppé and colleagues did not find a correlation between receptive prosody and speech production. Though Golan and colleagues detected significant associations between affective prosody and verbal IQ, two other studies failed to detect any associations between affective prosody and verbal IQ (Rutherford et al., 2002; Stewart et al., 2013).

Table 6. Summary table of current evidence based on prosodic deficits in speakers with ASD at form (acoustic-perceptual) or function (linguistic) levels

Levels prosodic deficits	Types of prosodic deficits	Studies that identified significant deficits			Studies that reported null findings		
		References	N	%	References	N	Pct
Form	Acoustic-perceptual	Brooks, Gaggi, & Ploog (2018) Diehl & Paul (2012) Järvinen-Pasley, Peppé, King-Smith, & Heaton (2008) Peppé et al. (2007) Kargas, López, Morris, & Reddy (2016)	5	63%	Brooks & Ploog (2013) Ploog, Brooks, Scharf, & Aum (2014) Ploog, Banerjee, & Brooks (2009)	3	37%
Function	Gramma-tical	Diehl, Bennetto, Watson, Gunlogson, & McDonough (2008) Järvinen-Pasley, Peppé, King-Smith, & Heaton (2008) Lyons, Simmons, & Paul (2014)[b]	3	33%	Chevallier, Noveck, Happé, & Wilson (2009) Diehl, Friedgberg, Paul, & Snedeker (2015)[a] Erwin, Van Lancker, Guthrie, Schwafel, Tanguay, Buchwald (1991) Grossman, Bemis, Skwerer, & Tager-Flusberg (2010) Peppé et al. (2007) Paul, Augustyn, Klin, & Volkmar (2005)	6	67%
	Pragmatic	Lyons, Simmons, & Paul (2014)[b] Paul, Augustyn, Klin, & Volkmar (2005)	2	50%	Peppé et al. (2007) Järvinen-Pasley, Peppé, King-Smith, & Heaton (2008)	2	50%
	Affective/Emotional	Hubbard, Faso, Assmann, & Sasson (2017) Järvinen-Pasley, Peppé, King-Smith, & Heaton (2008) Golan, Baron-Cohen, Hill, & Rutherford (2007) Lindner & Rosén (2006) Peppé et al. (2007) Rutherford, Baron-Cohen, Wheelwright (2002) Stewart, McAdam, Ota, Peppé, Cleland (2013) Van Lancker, Cornelius, Kreiman (1989)	8	53%	Brennand, Schepman, & Rodway (2011) Chevallier, Noveck, Happé, & Wilson (2011) Erwin, Van Lancker, Guthrie, Schwafel, Tanguay, Buchwald (1991) Grossman, Bemis, Skwerer, & Tager-Flusberg (2010) Paul, Augustyn, Klin, & Volkmar (2005) Singh & Harrow, 2014 Wang, Lee, Sigman, Dapretto (2006)	7	47%

Note: [a] The authors detected significant differences between groups in Block 2 of the experiment and attributed significantly poorer performance in the ASD group to reduced ability to overcome learned patterns in Block 1 instead of ability to perceive prosody.
[b] This study only detected significant differences in grammatical and pragmatic functions of prosody in the preteen group but not in the adolescent group. Additionally, significant differences were only detected between the TD group and the ASD Low Language group.
[c] The paradigm used in each study was color coded in this table:
Yellow: eye-tracking; green: PEPS-C; blue: forced-choice; orange: video game paradigms; **black**: other linguistic tasks.

Table 7. Studies that examined associations between the perception of prosody and broader linguistic or social-communication functioning

References	Task	Form vs Function	Aspect of Prosody	Association with broader skills
Diehl, Bennetto, Watson, Gunlogson, & McDonough, 2008	Syntactic ambiguity paradigm	Function	Grammatical	Participants with ASD's performance on the prosody-only task was marginally correlated with receptive language scores, r = .4, p = .08.
Diehl & Paul, 2012	Form tasks from PEPS-C	Form	NA	Participants' performance on receptive form tasks on PEPS-C significantly correlated with general language abilities measured by CELF-4 Core Language Index.
Golan, Baron-Cohen, Hill, & Rutherford, 2007	Reading the Mind in the Voice - Revised (RMV-R) task	Function	Affective/ Emotional	Participants' performance on the RMV-R task correlated significantly with Autism Quotient and verbal IQ measured by WASI, but didn't correlate with age or performance IQ.
Hubbard, Faso, Assmann, & Sasson, 2017	Emotion recognition task	Function	Affective/ Emotional	Emotion recognition accuracy significantly correlated with IQ in the ASD group (r = .08, p <.001)
Kargas, López, Morris, & Reddy, 2016	Lexical stress same-different discrimination task	Function	Grammatical	A significant correlation was detected between stress perception and speech abnormalities scores (item 2 on ADOS; r =-.75, p =.001): less sensitivity on syllable stress was associated with higher speech abnormalities score on ADOS.
Linder & Rosén, 2006	Perception of Emotion Test	Function	Affective/ Emotional	Participants' performance on prosodic trials significantly correlated with age (r=.56, p<.01) and POET total score (r = .78,p<.01).

Peppé et al., 2007	All tasks from PEPS-C	Form and function	Grammatical, Pragmatic, Affective/ Emotional	In the ASD group, PEPS-C receptive scores correlated with BPVS (r =.78, p <.001), TROG (r = .59, p<.001), CELF-3 (r = .72, p<.001), RM (r = .51, p <.01), and age (r =.60, p<.001). PEPS-C receptive score did not correlate with CCC pragmatics or GTFA. PEPS-C function total score sig. correlated with BPVS, TROG, and CELF-3, whereas PEPS-C form total score correlated with CELF -3 and age. In both ASD and TD groups, PEPS-C receptive scores correlated sig. with PEPS-C expressive scores.
Rutherford, Baron-Cohen, Wheelwright, 2002	Reading the Mind in the Voice Test	Function	Affective/ Emotional	Performance on the RMV test did not correlate significantly with verbal IQ in either the ASD group or the TD group.
Singh & Harrow, 2014	Emotion Recognition task	Function	Affective/Em otional	Participants' congruence effect (i.e. faster responses to congruent stimuli compared to incongruent stimuli) was directly related to social communication skills (measured by SCQ), suggesting that greater autistic symptomatology was associated with a reduced sensitivity to congruence based on semantic contents and prosodic cues.
Stewart, McAdam, Ota, Peppé, Cleland, 2013	Emotion recognition task	Function	Affective/ Emotional	Participants' performance on the tasks did not correlate with the Mill Hill (verbal IQ) test.

2.4 Discussion

There have been more attempts to understand receptive prosody in the ASD population in the past two decades. In this systematic review, 25 studies that examined form- and function-level of receptive prosody in ASD were identified and summarized. A review of current literature suggests that conflicting evidence has been reported on both form- and function-levels. Compared to function-level prosodic abilities, form-level prosodic deficits in ASD have received relatively more convergent support. Five studies (63%) reported that children and adults with ASD performed significantly worse on form-level tasks such as discriminating lexical stress and prosodic intonation patterns than TD controls. Regarding function-level prosody processing abilities, studies reported as many significant differences as null findings between ASD and TD groups across grammatical, pragmatic, and affective/emotional prosody. Though the results of these studies have not painted a conclusive picture on form- or function-level prosody processing in ASD, this review identifies four themes that warrant further discussion: auditory processing in ASD in general, prosodic deficits in ASD as a function of stimuli complexity and task demands, prosodic deficits in ASD as a function of sample characteristics, and potential dissociation between form- and function-level prosody.

2.4.1 Auditory processing in ASD

Given that the main theme of this review was receptive prosody as an acoustic and linguistic phenomenon on form- and function-levels, only studies that included at least one behavioral measure of receptive prosody was included. However, in addition to form-level processing ability as a prerequisite of function-level processing, prior studies have also argued that basic auditory processing skills need to be examined before assessing form- or function-level receptive prosody considering that prosodic phenomena are conveyed by subtle variations in acoustic parameters such as F0, intensity, and duration (Chevallier et al., 2009, 2011).

In a recent review, O'Conner (2012) provided a comprehensive account on auditory processing in ASD and summarized current findings on pitch and intensity perception in detecting changes in pitch (Heaton, 2005; Heaton, Pring, & Hermelin, 1999; Mottron, Peretz, & Ménard, 2000) and identifying the pitch of isolated pure-tone stimuli in discrimination and categorization tasks (Bonnel et al., 2003; O'Riordan & Passetti, 2006). Studies that tested the ability to discriminate different sound intensity levels in individuals with ASD have found similar findings: children and adults with ASD were as accurate as TD controls at discriminating intensity levels in pure-tones (Bonnel et al., 2010; Jones et al., 2009). Two studies (Chevallier et al., 2009, 2011) that were included in this current systematic review used the Dino Task as a control task to assess participants' ability to discriminate pitch, intensity, and duration before assessing their ability to perceive or comprehend prosody (Sutcliffe & Bishop, 2005). Consistent with previously reviewed studies, no group difference was detected between TD and ASD groups in intensity, duration, or pitch discrimination. Thus, these findings collectively suggest that children with ASD are equipped with sufficient basic auditory processing skills needed to process prosody.

To reconcile these findings on intact or enhanced auditory processing abilities in ASD with the findings that suggested receptive prosodic deficits in ASD, O'Conner concluded that individuals with ASD may exhibit intact or even enhanced discrimination of low-level acoustic information (e.g. pitch and intensity in pure-tone stimuli) but manifest impaired performance when processing high-level linguistic information (e.g. pitch contours in speech stimuli). This hypothesis is consistent with both the Enhanced Perception Functioning theory (Mottron, Dawson, Soulières, Hubert, & Burack, 2006) which proposes that individuals with ASD have enhanced discrimination ability of low-level perceptual information and the Neural Complexity theory (Bertone, Mottron, Jelenic, & Faubert, 2005; Samson, Mottron, Jemel, Belin, & Ciocca, 2006) which predicts that enhanced and reduced processing ability in individuals with ASD depend on the complexity of the stimuli. In summary, evidence from auditory processing literature suggests that prosodic deficits reported in individuals with ASD are more likely due to a combination of overly focused selective attention to perceptual information in ASD and the complexity of speech stimuli used in prosody research rather than inadequate basic auditory processing ability.

2.4.2 Prosodic deficits in ASD as a function of stimuli complexity and task demands

There is some evidence to suggest that individuals with ASD demonstrate more evident deficits when processing sentence-level stimuli compared to word-level stimuli. Two studies (Diehl & Paul, 2012; Järvinen-Pasley et al., 2008) that used PEPS-C form-level tasks found that children with ASD only displayed deficits in discrimination of prosodic patterns at phrase or sentence-level but not at word-level. Järvinen-Pasley and colleagues also found this pattern in function-level tasks: children with ASD performed as well as TD controls on the turn-end task in PEPS-C that uses single words with either rising or falling intonations but were significantly less accurate in a subsequent grammatical intonation task that used sentences. A closer examination of the findings regarding affective prosody comprehension reveals similar patterns. Six out of the eight studies that reported null findings on affective prosody either used word stimuli (Grossman et al., 2010; Singh & Harrow, 2014), short sentences such as “Way to go” or “You look great”(Wang et al., 2006), pseudo-sentences (Brennand et al., 2011), or low-pass filtered sentences without semantic contents (Grossman et al., 2010). In contrast, all but two studies that detected significant between-group differences mostly used full unfiltered sentences as stimuli (Erwin et al., 1991; Golan et al., 2007; Hubbard et al., 2017; Lindner & Rosén, 2006; Rutherford et al., 2002; Stewart et al., 2013). Together, these findings suggest that receptive prosodic deficits may only manifest in tasks that involve long linguistic units.

In addition to stimuli complexity, uneven task demands across studies could also explain inconsistent findings in the reviewed studies. For example, task demands in the included studies range from a forced-choice of two after listening to one word to identifying speakers' emotional state in full sentences while reporting back the number of times participants heard the letter T in sentences. While the former type of task could mask important prosodic processing differences due to ceiling effects, the latter likely assesses abilities beyond prosodic processing. These drastic differences in task demands render it challenging to compare results directly across studies.

2.4.3 Prosodic deficits in ASD as a function of sample characteristics

Another theme identified throughout this review is that receptive prosodic deficits may only be evident in a subgroup of individuals with ASD. In Lyons et al. (2014), children with ASD were further divided into a group with typical language functioning (ASD-Hi) and a group with low language ability (ASD-Lo) based on their CELF-4 standard scores. Results reveal that children in the ASD-Lo group performed significantly worse on receptive prosody tasks than children in both ASD-Hi and TD groups. Similarly, Diehl and Paul (2012) only detected significant differences between TD and ASD groups but not between children with ASD and age, language, and cognitive-matched children with learning disabilities. These findings imply two important points. First, observed prosodic deficits could be an artifact of underlying cognitive or language deficits in individuals with ASD. Indeed, a subset of the included studies examined associations between receptive prosody and language or cognitive abilities and found significant correlations between them. Second, prosodic deficits are not unique to the ASD population. Studies that examine receptive prosody across clinical conditions could contribute interesting insights into the underlying processes needed to successfully process prosody. For instance, a study that compared age, receptive vocabulary, and cognitive-matched participants with Williams Syndrome and intellectual disabilities found that children with WS showed enhanced performance in affective prosody processing than children with intellectual disabilities but both groups showed deficits in grammatical prosody processing tasks (Skwerer, Schofield, Verbalis, Faja, & Tager-Flusberg, 2007). Future studies could compare children with ASD with children with hearing loss, Williams Syndrome, or Specific Language Impairment to decipher the impact of perceptual, linguistic, and cognitive abilities on the perception and comprehension of various aspects and functions of prosody.

2.4.4 Resolving potential dissociation between form- and function-level prosody

Regarding the hypothesis that form-level auditory-perceptual discrimination ability is a prerequisite for function-level linguistic processing ability, the combined findings from some studies suggest a dissociation between form- and function-levels of receptive prosody ability (i.e. impaired form-level ability and unimpaired function-level ability). However, as discussed above, results from different studies are not directly comparable due to uneven stimuli complexity, task demands, and sample characteristics. In the two studies that examined both form- and function-level of receptive prosody (Järvinen-Pasley et al., 2008; Peppé et al., 2007) in similar samples of children with ASD, deficits at both levels were detected in children with ASD. Neither study examined the relations between form- and function-level of receptive prosody.

2.4.5 Clinical implications

Although current evidence on prosody deficits in individuals with ASD did not illuminate the relations between form- and function-level of receptive prosody in this population, several studies did highlight that form-level of prosody ability should not be the only focus in prosody intervention (Diehl & Paul, 2012; Järvinen-Pasley et al., 2008). Given that the *ultimate* goal of speech-language intervention is to improve an individual's *functional* social communication, focusing predominantly on form-level auditory-perceptual discrimination ability presumably is not sufficient to translate into functional gains. A comparison of form-level and function-level tasks in PEPS-C illustrates this point. In form-level tasks, participants listened to pairs of low-pass filtered words and sentences and were asked to judge whether the pair sounds the same or different. In contrast, in function-level tasks, participants listened to unfiltered sentences and were asked to interpret the meanings of various prosodic patterns. Essentially, prosody has been decontextualized in form-level tasks. In reality, prosody is always presented with linguistic information and is often laden with social information. Therefore, even though one may start with form-level discrimination training, it seems critical that prosodic patterns be taught in functional tasks to improve interpretation of prosody in social communication.

To date, there has been no published study on intervention of receptive prosody deficits in ASD. Diehl and Paul (2012) suggested that it might be helpful to develop interventions that target expressive and receptive prosody simultaneously. For example, individuals with ASD could be instructed to make prosodic changes in communicative or non-communicative ways as a strategy to help them focus their attention on meaningful changes in prosody and interpret prosody in an interactive context.

2.4.6 Limitations and future directions

This reviewed identified several common limitations in the published studies to date that examined receptive prosody in ASD. First, in several studies, the stimuli used in the task had not been validated, which may have compromised the rigor of the experiments. For instance, as reviewed above, one study (Brennand et al., 2011) used stimuli created by a previous study to test affective prosody. Yet the stimuli used to indicate certain specific emotions (fearful) were originally recorded to indicate different emotions (e.g. anxious) and were not validated by a separate group of listeners. Second, ceiling effects were common in many studies, particularly in the ones that used a binary forced-choice paradigm. Although one study argued that using a binary choice reduces demands on auditory memory (Peppé et al., 2007), it also reduces the ecological validity of the task and likely masks potential prosodic processing deficits in individuals with ASD. Third, adequate group equivalence was not established in some studies. In some studies, participants between groups have significantly different language or cognitive functioning. Many studies did not report language or cognitive abilities of participants. On a related note, in a few studies, the tasks used were loaded with language demands. For example, in one study, participants were expected to understand vocabulary such as *brooding* and *lured*, but participants' comprehension of such vocabulary were not assessed beforehand. These methodological limitations imply that observed prosody differences could be confounded with language or cognitive deficits.

There are also several limitations to this systematic review in addition to varying degrees of methodological rigor in included primary studies. First, due to the scope of this paper, only published articles were coded and summarized in this paper, which could lead to publication bias. Additionally, the external validity of these results is compromised because all but four included studies recruited high-functioning individuals with ASD. Thus, it remains unclear whether trends observed in this review would generalize to a different subgroup of individuals with ASD.

Finally, during the abstract screening phase of this systematic review, more than twenty studies on receptive prosody in a non-English speakers were identified. Due to the scope and the inclusion criteria of this review, those studies were excluded. However, given that prosodic structures vary in different languages (Ohala, 1983), cross-linguistic studies on receptive prosody in ASD may provide novel insights into the nature and the extent of prosodic deficits in this population. For instance, in Mandarin, a unique linguistic device that serves similar functions as prosodic cues in English is termed sentence-final particle (Sun, 2006). As its name suggests, this device always appears at the end of an utterance and can serve grammatical, pragmatic, and affective functions even though it does not add additional semantic content to the utterance. To illustrate, the question marker, “ma” (吗), at the end of an sentence in combination with a rising tone would indicate a question. “Ne” (呢), a different sentence-final particle, can be used to either mark questions or can be used ic combination with various pitch contours to convey a range of emotions, including affection, appreciation, surprise, or even sarcasm (Sun, 2006). Studying perception of intonation contours and the comprehension of sentence-final particles in Mandarin-speaking individuals with ASD could be an innovative attempt to test theories on underlying mechanisms of receptive prosody.

CHAPTER 3

The Perception and Comprehension of Prosody in Children with ASD

3.1 Introduction

Language deficits are a common feature in children and adults with Autism Spectrum Disorder (ASD)(Magiati et al., 2014) and are likely related to the core features of the disorder. Pragmatic language, defined as the social use of language, stands at the intersection of the social and language domains and could provide important insights into both social and language deficits in ASD (Prutting, 1982; Tager-Flusberg, 1999). In particular, prosody, one crucial aspect of pragmatic skills, is essential to social participation and interaction and, accordingly, is a reasonable candidate for mechanistic studies designed to advance current understanding of social and language impairments in ASD. Though structural language skills vary in this population, individuals with ASD demonstrate significant difficulties with at least one aspect of prosody (McCann & Peppé, 2003; McCann et al., 2007). Furthermore, prosodic deficits observed in this population tend to persist and affect long-term social and communicative competence even when other aspects of language improve (Paul, Shriberg, et al., 2005). These findings suggest that accurate identification and effective interventions that target prosodic deficits are necessary to improve long-term social and communication outcomes for individuals with ASD. This study examined the perception and comprehension of prosody in children with ASD and associations among receptive prosody and broader social communication skills.

3.1.1 Prosody and its significant role in language and communication

Spoken language conveys not only words and sentences but also a wide range of other information such as intonation, rhythm, stress, timing, tone, etc. These features are collectively referred to as "prosody" and are defined as suprasegmental features of language because they can span across multiple phonetic segments (Lehiste, 1970; Wagner & Watson, 2010). Given that in English prosody generally does not have orthographic representation other than punctuations, the terminology of prosody and the scope of the topic tend to vary across studies (Peppé, 2009; Xu, 2015). More recent studies of prosody have adopted the form vs. function theoretical framework to investigate prosodic deficits in ASD (Järvinen-Pasley et al., 2008; Peppé et al., 2007). In these studies, the form level of prosody involves the ability to perceive and produce prosodic forms (i.e. auditory discrimination and vocal imitation); whereas the function level requires the ability to cognitively and linguistically process communicative functions that prosody serves in spoken communication, including grammatical, semantic/pragmatic, and affective/emotional functions (Crystal, 1986; McCann & Peppé, 2003; Paul et al., 2005).

Prosody contributes significantly to spoken language comprehension and successful social communication (Cutler & Swinney, 1987; Frazier, Carlson, & Clifton, 2006; Mehler et al., 1988; Paul et al., 2005). One of the most important functions of prosody is to enhance or modulate information beyond the literal meaning of an utterance, such as speaker intention or emotion (Couper-Kuhlen, 1986; Shattuck-Hufnagel & Turk, 1996). For instance, contrastive pitch accent can be used by speakers to highlight information that's new or important to conversation partners (Selkirk, 1995; Watson, 2010). Intonation can be used to convey a speaker's general emotional state or even contradict the literal meaning of an utterance, as in the case of sarcasm (Bolinger, 1982, 1983; Capelli, Nakagawa, & Madden, 1990). Given that social communication involves not only exchanging the content of messages but also inferring the intentions of others (Brothers, 1990), successful communication depends on one's ability of perceive and comprehend subtle prosodic cues.

3.1.2 Prosodic deficits in ASD and their functional implications

Atypical prosody has been frequently identified in individuals with ASD and was among one of the original features that Kanner described in his seminal paper (American Psychiatric Association, 2013; Kanner, 1943; McCann & Peppé, 2003). Clinical descriptions of prosodic deficits in ASD include flat, exaggerated, or “sing-song” intonation, and wide variation in intensity (Peppé et al., 2007). A review paper (McCann & Peppé, 2003) revealed that individuals with ASD present with a variety of deficits in both expressive and receptive modalities of prosody, across all elements of prosody, and impacting various functions that prosody serves in communication. Paul and colleagues examined both production and perception of a range of prosodic elements in a group of children with ASD and found that participants with ASD performed significantly worse than the TD control group in production and perception of stress regardless of whether it was used to serve grammatical functions or pragmatic functions (Paul, Augustyn, et al., 2005). Participants with ASD had significantly more utterances coded by trained listeners as inappropriate across prosodic elements (Shriberg et al., 2001). Additionally, naïve untrained listeners rated speech from individuals with ASD with a higher rate of awkward rate, volume, and intonation (Bone, Black, Ramakrishna, Grossman, & Narayanan, 2015).

These prosodic deficits may pose tremendous barriers to language learning and everyday social participation for children with ASD. Deficits in production or perception of prosody can lead to communication breakdowns, negatively impact one’s social competence, and are considered as one of the most significant challenges to social participation and integration (McCann, Peppé, Gibbon, O’Hare, & Rutherford, 2007; Paul et al., 2005). The aforementioned finding that untrained listeners perceived speech from speakers with ASD as "awkward and odd" highlights the stigmatizing effect of prosodic deficits on social acceptance of individuals with ASD (Bone et al., 2015; Shriberg et al., 2001). Children with ASD are more likely to be marginalized within the peer group (Pepler, Schroeder, Weiss, Cappadocia, & Bebko, 2014), develop less friendships (Chamberlain, Kasari, & Rotheram-Fuller, 2007), and experience a higher rate of bullying (Van Roekel, Scholte, & Didden, 2010).

3.1.3 The need to understand receptive prosody ability in children with ASD

The vast majority of existing research on prosody has primarily focused on expressive abilities, leaving perception of prosody poorly understood (O'Connor, 2012). Existing studies that examined receptive prosody have reported contradicting findings. As shown in the previous chapter, conflicting evidence has been reported on both form- and function-levels of receptive prosody in children with ASD. While some studies have reported that children with ASD demonstrated difficulty in perception of prosodic forms and comprehension of contrastive pitch accent, intonation, and lexical stress (Paul, Augustyn, et al., 2005; Peppé et al., 2007), other studies have found that children with ASD perform similarly to typically developing peers when detecting affects from intonation (Grossman et al., 2010), disambiguating words based on lexical stress, and deciding whether a sentence is a statement or a question based on intonation contours (Chevallier et al., 2009).

Given the likelihood of language impairments in children with ASD, it is important to consider the linguistic demands of tasks used to measure perception or comprehension of prosody in ASD. The majority of previous studies in this area used behavioral protocols where participants were expected to comprehend instructions and follow specific prompts to demonstrate their ability to perceive and comprehend prosodic cues in linguistic tasks. However, these tasks rely on participants' language comprehension and metalinguistic knowledge (e.g. understanding the difference between a sentence and a question) in addition to the ability to perceive and comprehend prosodic cues. These approaches also depend on participants' ability to respond consistently to task instructions, which can be challenging for children with ASD (Hudry et al., 2010). Thus, there is a need for studies that use objective measures to further current understanding of receptive prosodic ability in children with ASD.

3.1.4 Using eye-tracking to measure receptive prosody

The visual-world paradigm (VWP) has emerged to be a powerful approach for investigating spoken language processing in typically developing individuals (Huettig, Rommers, & Meyer, 2011; Tanenhaus et al., 1995). In a typical VWP task, participants looks as an experimental display as they hear an utterance that either instructs them to locate a target item (e.g. "Look at the cat") or describes or comment upon a scene (e.g. "The boy will move the cake"). The utterance or the display can be manipulated and participants' eye movements are analyzed to understand the impact of the experimental manipulation on participants' real-time language comprehension. Previous studies have used VWP to study various aspects of spoken language comprehension, including speech perception, syntactic processing, and pragmatic inferencing (Huettig et al., 2011).

Using VWP as a task to measure prosody processing has several advantages over other existing behavioral paradigms. First, this task presents relatively low task demands and only requires that participants look at a target item in a display as they hear the item. Second, it provides continuous information about a participant's linguistic processing and thus can be used to not only examine whether a participant comprehends a linguistic stimulus but also how a linguistic stimulus is processed. Finally, previous studies have found that fine-grained acoustic information can affect speed of linguistic processing in this paradigm (Huettig et al., 2011; Meyer & Damian, 2007).

Despite its potential to be used as a novel objective method to study the impact of subtle prosodic cues in the speech signal on linguistic processing in both typical and clinical populations, to date, only one study has used a VWP task to examine comprehension of grammatical prosody in children with ASD (Diehl et al., 2015). In this study, four groups of children (TD and ASD groups each with two age subgroups, children and teens) were told that they would play a game about following instructions. Participants then listened to syntactically ambiguous commands (e.g. "You can feel the frog with the feather"). The prosody of these commands were manipulated by either placing a prosodic boundary cue (i.e. pause and boundary intonation) before the first noun phrase (e.g. "You can feel // the frog with the feather") or before the prepositional phrase (e.g. "You can feel the frog // with the feather"). The results indicated that children with ASD were able to use grammatical prosodic cues to parse sentences but had difficulty overriding learned expectations when prosodic pattern changes. This performance pattern was only evident in the younger group (7- to 12-year-olds), which may reflect immature executive function, specifically cognitive flexibility, more so than global receptive prosody deficits.

3.1.5 Rationale for focusing on contrastive pitch accent in ASD

Contrastive pitch accent is used to mark a word as prominent (Pierrehumbert & Hirschberg, 1990) and is acoustically associated with longer duration, greater amplitude, and pitch movement on the stressed syllable (Watson, Gunlogson, & Tanenhaus, 2006). It is often used to direct listener's attention to a salient information in a discourse context or to contrast a piece of information with possible alternatives (Bolinger, 1961, 1972; Pierrehumbert, 1980). This study focuses on perception of contrastive pitch accent for two reasons. First, previous studies have suggested that contrastive pitch accent may be a specific area of deficit for children with ASD (Paul, Augustyn, et al., 2005; Peppé et al., 2007; Simmons & Baltaxe, 1975); however, these studies did not use an objective approach to help overcome potential linguistic or cognitive task demands. Second, the ability to perceive and comprehend an contrastive pitch accent is essential for social communication. It has been considered as a semantic or pragmatic aspect of prosody because speakers often use contrastive pitch accent to convey or infer subtle information beyond a sentence's literal meaning (Peppé, 2009). Because prosody is a complex acoustic and linguistic phenomenon that serves various functions in communication, isolating one aspect of prosody and assessing it in a controlled setting could provide a window into disruptions in perception of prosody in ASD. Given the key role of contrastive pitch accent in social communication, focusing on contrastive pitch accent could serve as a first step to fill existing gaps in prosody research in ASD.

Previous work on contrastive pitch accent understanding in TD children has consistently demonstrated that contrastive pitch accent, when used contextually appropriately, can facilitate listeners' comprehension and accelerate their visual search for a target referent. This effect has been termed as an anticipatory effect of contrastive pitch accent (Ito & Speer, 2008). In contrast, contextually inappropriate use of contrastive pitch accent has been found to mislead listener's processing and delay their referent search, which has been also termed as a garden-path effect (Arnold, 2008; Dahan et al., 2002; Ito & Speer, 2008). Ito and colleagues investigated children's ability to perceive and comprehend contrastive pitch accent in a VWP task and found that TD children demonstrated both the facilitative effect and garden-path effect when listening to sentences with manipulated contrastive pitch accents (Ito et al., 2014). These results indicate that typically developing children can rapidly perceive and comprehend the semantics of contrastive pitch accent during spoken language processing.

3.1.6 The goals of this study

In the present study, our goals were to first compare the perception and comprehension of contrastive pitch accent using an eye-tracking task in TD children and children with ASD and then investigate associations between eye-tracking task-related measures of contrastive pitch accent comprehension and more clinically-oriented measures of broader social communication skills within the ASD sample.

Specifically, this study was guided by two research questions:

1. Are children with ASD less likely to comprehend the semantics of contrastive pitch accent during spoken language comprehension than TD children as measured by an eye-tracking visual-world task?
2. Are visual-world task-related measures of the comprehension of contrastive pitch accent within the ASD group related to clinical measures of their receptive prosody ability, pragmatic language ability, social communication functioning, and ASD symptom severity?

3.2 Methods

3.2.1 Overview of study design

The Vanderbilt University Institutional Review Board approved the study protocol. Parental consent and the child's assent were obtained for all study procedures. Twenty-four TD children and 18 children with ASD between 8 and 14 years who met inclusion criteria participated in two experimental tasks. The first task was an AX same-different task (Gerrits & Schouten, 2004) designed to assess participants' ability to perceive and discriminate contrastive pitch accent forms. The second task was an eye-tracking VWP task (Tanenhaus et al., 1995) adapted to assess participants' ability to comprehend the contrastive semantics of contrastive pitch accent. Participants were also tested on a battery of clinical assessments on receptive prosody, pragmatic language, social communication, and ASD symptom severity to investigate relations between participants' ability to process contrastive pitch accent during online spoken language comprehension and broader skills.

3.2.2 Participants

Forty-eight children between 8 and 14 were recruited (N_{TD} = 24, N_{ASD} = 24). The inclusion criteria for TD children are as follows: a) native English speaker; and b) no existing diagnosis of neurological, hearing, visual, or cognitive impairment. For children with ASD, diagnoses were confirmed using the Autism Diagnostic Observational Schedule – Second Edition (ADOS-2; Lord et al., 2012). Children with ASD were eligible for this study if they have a confirmed diagnosis of ASD, an Intelligence Quotient (IQ) of above 70 according to the Stanford-Binet Intelligence Scale – Fifth Edition (SB-5; Roid, 2003), and no existing diagnosis of hearing or visual impairment. Six participants with ASD were excluded because they did not meet the cognitive criterion.

Participants in both groups were matched on age, overall IQ, non-verbal IQ, and verbal IQ. Participants' language abilities were measured using the Clinical Evaluations of Language Fundamentals – Fifth Edition (CELF-5; (Wiig, Semel, & Secord, 2013). Additional descriptive characteristics of the participants are available in Table 8.

Table 8. Mean (and SD) of demographic and clinical measures for the ASD and the TD groups

	TD	ASD	*p*
Age (M, SD)	11.61 (1.97)	11.04 (1.92)	.35
IQ (M, SD)	106.61 (11.21)	100.83 (14.44)	.15
NVIQ[a] (M, SD)	11.26 (2.87)	10.28 (2.42)	.25
VIQ[b] (M, SD)	10.96 (1.55)	10.00 (3.50)	.24
Language[c] (M, SD)	108.04 (12.96)	91.67 (15.65)	<.001***

Note: [a] NVIQ = non-verbal IQ measured by Stanford-Binet Intelligence Scale – Fifth Edition (SB-5; Roid et al., 2013); [b] VIQ = verbal IQ measured by SB-5; [c] Language measured by Clinical Evaluation of Language Fundamentals – Fifth Edition (CELF-5; Wiig et al., 2013).

3.2.3 Assessment Measures

Profiling Elements of Prosody in Speech – Communication (PEPS-C). The PEPS-C (Peppé & McCann, 2003) is a semi-standardized computerized prosody assessment. It consists of 14 tasks with seven receptive tasks and seven expressive tasks. Each subtest contains 16 items. The authors of the PEPS-C suggest that a subtest score above 12 (75%) indicates that a participant reached competence level in a subtest. Two scores from the PEPS-C were used in this study: Contrastive Pitch Accent Understanding subtest score and receptive prosody composite. Because both the visual-world task-related measures and the Contrastive Pitch Accent Understanding were conceptualized to measure the same construct, we used the Contrastive Pitch Accent Understanding to confirm the construct validity of the visual-world task-related measures before assessing relations between visual-world task related measures and clinical measures. Additionally, receptive prosody composite from the PEPS-C, which is the sum of all receptive subtests and represents comprehension of all aspects of prosody including but not limited to contrastive pitch accent, was used as an index of participants' overall receptive prosody ability.

Clinical Evaluation of Language Fundamentals - Fifth Edition Metalinguistics (CELF-5 Metalinguistics). The CELF-5 Metalinguistics (Semel, Wiig, & Secord, 2014) is a standardized test that assesses participants' ability to make inferences, engage in discourses, and understand ambiguous or figurative language. Participants' performance on the Making Inferences and Conversational Skills subtests can be used to derive a Meta-Pragmatic Index score, which was used as a measure of participants' pragmatic language ability.

Social Responsiveness Scale (SRS). The SRS (Constantino, 2012) is a parent-reported measure with a 0 - 3 Likert scale, designed to identify the presence and measure severity of social deficits in both the general population and clinical settings. The score from Social Communication Subscale was used as a measure of participants' social communication skills.

Autism Diagnostic Observation Scale, Second Edition (ADOS-2). The ADOS-2 (Lord et al., 2012) is a standardized observational measure used to elicit the behaviors that are characteristic of ASD in a standardized setting. Total score from the ADOS-2 logarithm was used as a measure for ASD symptom severity for children in the ASD group.

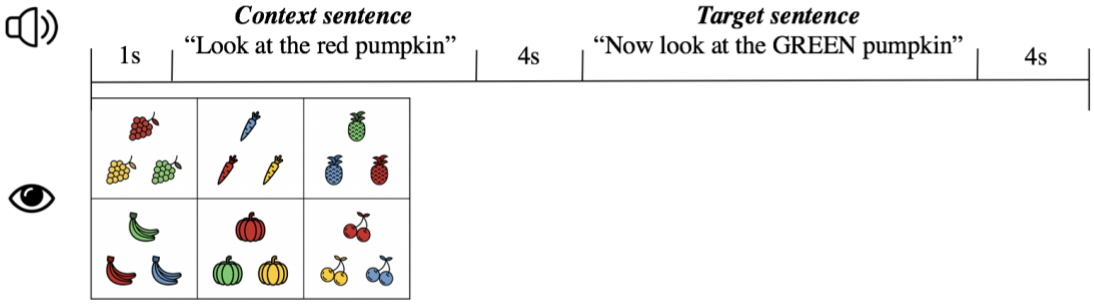
3.2.4 Experimental tasks

AX same-different discrimination task. We adapted an AX same-different discrimination task as a control task (Gerrits & Schouten, 2004) to confirm that participants in this study meet the auditory-perceptual prerequisite before assessing high-level prosody processing ability. This task was programmed and implemented using the PsychoPy software packages (Peirce, 2007, 2008). Participants listened to 16 test trials with two acoustic stimuli and be instructed to press two keys on the computer keyboard to indicate whether two acoustic stimuli are the same or different. All acoustic stimuli follow the same structure: “the” + adjective + noun (e.g. the sunny morning, the hot summer). Each pair contains the same phrases but the prenominal adjective was manipulated so that 8 trials had a pair with identical contrastive pitch accent patterns (e.g. the SUNNY morning and the SUNNY morning, capitalized words denoting the presence of a contrastive pitch accent.) and 8 trials had a pair with different patterns (e.g. the SUNNY morning vs. the sunny morning). Participants were provided with two examples and four practice trials before test trials. Participants’ accuracy score was used as a measure of their ability to discriminate contrastive pitch accent. Participants in both groups reached an accuracy rate of 100%, indicating that participants with ASD are able to detect the presence or absence of contrastive pitch accent on words as well as TD participants.

Eye-tracking visual-world task. To assess participants’ ability to comprehend contrastive pitch accent during online spoken language processing, we adapted a visual-world paradigm from Ito et al. (2014). The visual-world paradigm (VWP) is an established technique in psycholinguistics to study language processing (Tanenhaus et al., 1995). In a typical VWP task, participants look at visual scenes and listen to spoken language containing references to objects in the scene. In our task, participants watched a 19-minute video with a total of 72 trials. Each trial in this paradigm consists of one visual scene and two sentences (see

Figure 2 for an example). On each trial, each participant was instructed to look at two items in two sentences. The assignment of contrastive pitch accent in the instructions was manipulated to create either appropriate or inappropriate contexts for contrastive pitch accents.

Figure 2. Schematic diagram for a sample trial in the visual-world task



Visual stimuli. The visual stimuli consist of 72 static slides. The 72 test trials were prepared by combining 12 unique items from 6 categories (i.e. clothing items, household items, animals, furniture, office supplies, and fruit & vegetables) in four colors. Each slide was divided into six cells and each cell contains one unique item in three colors (see

Figure 2 for an example). The six items on each slide are always drawn from the same category. Items were carefully chosen to avoid any items that associate with high autism interest (Sasson, Dichter, & Bodfish, 2012; Sasson, Turner-Brown, Holtzclaw, Lam, & Bodfish, 2008) as these items may bias attention in the ASD group over and above the effect of the experimental conditions. Item images were first tested in a pilot study with children and adults who did not take part in this study to confirm that selected images are recognizable and familiar to participants. In the pilot study, six children between 8 and 14 were shown all selected images and were asked to name each one. Only images that were correctly labeled by all participants were included in the task. The combination of items, colors, and positions were counterbalanced so that the number of appearance for each item, each color, and each of the six positions within a slide were the same across the entire set of stimuli design.

Auditory stimuli. The auditory stimuli were recorded by a female native speaker of Standard American English at 44.1 KHz using Praat (Boersma & Weenink, 1992-2017). The auditory stimuli for each trial consist of a pair of sentences: the context sentence and the target sentence. The context sentence always contains a prenominal adjective with a neutral accent, whereas the pronominal adjective in the target question was either assigned a contrastive pitch accent or a neutral accent. Acoustic analyses confirmed that accented prenominal adjectives correspond to significantly longer duration ($M_{contrastive}$ = 475.08 ms, $M_{neutral}$ = 282.66 ms , $p < .001$), higher F0 mean peak ($M_{contrastive}$ = 228 Hz, $M_{neutral}$ = 166 Hz , $p < .001$), and higher F0 peak ($M_{contrastive}$ = 306 Hz, $M_{neutral}$ = 187 Hz , $p < .001$) than neutral adjectives. The Tone and Break Index coding of recorded stimuli also confirmed that prenominal adjectives with contrastive pitch accents correspond to a L+H* annotation and prenominal adjectives that are not accented correspond to an H* annotation (Watson, Tanenhaus, & Gunlogson, 2008).

Once recorded, the sentences were edited so that the pre-target material (i.e. “Look at the”) and the target material (i.e. prenominal adjective and noun) were spliced out of their original context to create critical sentences. The same pre-target material was used across conditions for each target item. This step was necessary to ensure that any visual search patterns detected in this paradigm were solely due to the difference in contrastive pitch accent patterns.

Experimental conditions. This task has four critical conditions and one filler condition as follows: Appropriate-Accented (A), Appropriate-Neutral (B) , Inappropriate-Accented (C), and Inappropriate-Neutral (D), and Filler (F; see

Table 9 for all conditions and examples of sentences). In two critical conditions A and C, a contrastive pitch accent was assigned to the adjective in the target sentence, whereas no contrastive pitch accent was assigned to the adjectives in the target sentence in control conditions B and D. The filler trials were included to avoid participants' anticipation of the pitch accent patterns. The 72 test trials consist of 36 critical trials with nine critical trials for each condition for each participants and 36 filler trials. Half of the items in each of six categories were randomly selected and assigned as targets in filler trials. The assignment of the rest 36 items to four critical conditions was counterbalanced across four lists using a Latin Square design. Every list contained 72 unique items. The order of the trials was randomized in creating each list but was fixed for every use of that list. Participants were randomly assigned to view one of four presentation lists.

The key comparisons of interest are condition A vs. B and condition C vs. D. Both conditions A and B contain the same noun across the context sentence and the target sentence, creating an appropriate context to use contrastive pitch accents in condition A. An anticipatory effect would be present if participants look at the target item faster in condition A compared to B. Both conditions C and D contain different nouns across two sentences, creating an inappropriate context to use a contrastive pitch accent in condition C. A garden-path effect would be present if participants look at the target item slower or look at the competitor item (i.e. the incorrect item primed by the inappropriate use of a contrastive pitch accent) faster in condition C compared to condition D.

Table 9. Visual-world task conditions, with capitalized words denoting the presence of contrastive pitch accent

Condition	Example of a context sentence	Example of a target sentence
A. Appropriate-Accented	Look at the red pumpkin	Now look at the GREEN pumpkin
B. Appropriate-Neutral	Look at the red pumpkin	Now look at the green pumpkin
C. Inappropriate-Accented	Look at the red grapes	Now look at the GREEN pumpkin
D. Inappropriate-Neutral	Look at the red grapes	Now look at the green pumpkin
F. Filler	Look at the blue cherries	Now look at the yellow carrot

Procedure. For the visual-world task, participants sat in front of a Tobii X2 eye-tracker monitor and a set of speakers. Their eyes were first calibrated using the Tobii Clear View 5-point calibration program. They were then instructed to look at pictures while listening to sentences that would ask them to look for specific items in each picture. Participants' eye movements during the task will be sampled at 60 Hz.

Data preparation. A 250 x 250 pixels square around each target item was used as the target area of interest (AOI). Participants' eye movement data at each time sample was coded as either 1 (on) or 0 (off) for each given AOI. The analysis window was decided *a priori* as a 1200-ms time window, beginning 300 ms after the onset of the prenominal adjective in the target sentence, This window was offset by 300 ms because programming and executing an eye movement typically takes 200 ms in adults and 300 ms in children (Arnold, 2008; Hallet, 1986).

Statistical analysis. For the first research question, preliminary analyses were first completed to confirm the feasibility of the VWP task. Then, mixed-effect logistic regression models were used to examine the comprehension of contrastive pitch accent in TD participants and participants with ASD. This approach was used because it accommodates to both binomially distributed fixation data and also accounts for the clustered nature of observations from visual-world paradigm (e.g. trials nested in subjects and items) (Barr, 2008; Jaeger, 2008). The dependent variable in all mixed-effect logistic regression models were binary fixation responses (i.e. yes or no) to the correct target or the incorrect competitor during the 300ms to 1500 ms analysis window post target adjective onset. This 1200-ms time window includes 72 equally spaced, 16.67 ms (given a sampling rate of 60Hz from Tobii X2 eye-tracker) time points of binary data that indicate whether or not a participant was fixating on the target or competitor during that time point. All statistical models included crossed random intercepts and slopes for participants and items which allow estimates of subject and item variability in addition to fixed effects of condition and/or group (Baayen, Davidson, & Bates, 2008).

Additionally, given that a significant difference was detected between diagnostic groups on participants' language ability. Two sets of post-hoc analyses were conducted for Research Question 1 to understand the impact of language on comprehension of contrastive pitch accent. First, a mixed-effect logistic regression model with language and conditions as fixed effect, their interaction, and crossed random effects of subject and item was conducted to examine the impact of language on online processing of contrastive pitch accent continuously. Additionally, subgroup analyses comparisons were performed by assigning children in the ASD group into two subgroups using a cutoff of a score greater than 1 standard deviation below the mean on the CELF-5: children with a standard score lower than 85 were placed in the ASD with language impairment group (ASD+LI; N = 11) and those with a score of 85 or above were placed in the ASD with typical language group (ASD+TL; N = 7). Participants' fixation patterns in the TD group and the two ASD language subgroups were examined visually and tested statistically in mixed-effect logistic regression models.

For the second research question, before testing relations between individual differences in the visual-world task performance, correlation analyses were conducted between visual-world task-related measures and relevant clinical measures to confirm the construct validity of visual-world task-related measures. Two task-related measures of contrastive pitch accent processing and one measure of general language processing were derived from the visual-world task as follows: a) contrastive pitch accent comprehension (random by-participant slopes from mixed-effect logistic regression models), b) speed of contrastive pitch accent processing (latency of first fixation in condition C - i.e. inappropriate-accented), and c) speed of general linguistic processing (latency of first fixation in condition D - i.e. inappropriate-neutral). These three task-related measures were respectively correlated with the Contrastive Pitch Accent Understanding subtest from PEPS-C and a clinical measure of language ability (i.e. standard score from the CELF-5 language assessment) within the ASD group to provide a test of the construct validity of the visual-world task-related measures. Additionally, correlation analyses were conducted to examine the associations between these three task-related measures and four clinical measures: a) receptive prosody as measured by the receptive prosody composite from PEPS-C; b) pragmatic language as measured by the Meta-Pragmatic Index from CELF-5 Metalinguistics; c) social communication as measured by the Social Communication Subscale from SRS; and d) ASD symptom severity as measured by the total score from ADOS-2.

3.3 Results

3.3.1 Research question 1

3.3.1.1 Feasibility

Data cleaning and data loss. To confirm feasibility of using visual-world paradigm to test prosody processing in children with ASD, data loss during analysis window were analyzed first. For some timepoint samples, the eye tracker may not be able to capture eye movement either due to blinks or excessive movements. On average, percentage of tracked sample was 90% for TD participants and 87% for participants with ASD. Percentage of tracked sample during the analysis window did not differ across conditions (p = .56) or groups (p = .98). These results confirmed that this paradigm provided sufficient data to assess comprehension of contrastive pitch accent in TD participants and participants with ASD.

In addition, given that context sentences were embedded in each trial to create either an appropriate or inappropriate context for contrastive pitch accent in target sentences, a trial where a participant looked at the target item yet failed to look at the context item does not provide meaningful information regarding participants' comprehension of contrastive pitch accent. Thus, we excluded trials where participants failed to look at the context item following the onset of the prenominal adjective. This step removed 29 trials out of 1,728 trials from TD participants (2%) and 80 trials out of 1,296 trials from participants with ASD (6%). In the final analysis sample, on average, each TD participant contributed 71 trials whereas each participant with ASD contributed 68 trials.

Demonstration of comprehension of contrastive pitch accent in TD participants. Before assessing group differences, we first examined comprehension of contrastive pitch accent in TD participants. In previous studies (Ito, Bibyk, Wagner, & Speer, 2013; Ito & Speer, 2008), comprehension of contrastive pitch accent was demonstrated by two effects: anticipatory effects, where an appropriate use of a contrastive pitch accent accelerates participants' visual search for the correct target item, and garden-path effects, where an inappropriate use of a contrastive pitch accent misleads participants' visual search so that they increase their looks to the incorrect competitor item and delays their looks to the correct item.

Participants' fixation proportion to the target AOI during the target sentence were binned in time bins of 100 ms and then presented graphically in continuous time course. Figure 3-5 depict participants' fixation proportions to target item in conditions C and D with an inappropriate context for contrastive pitch accents (

Figure 3), competitor item in conditions C and D (

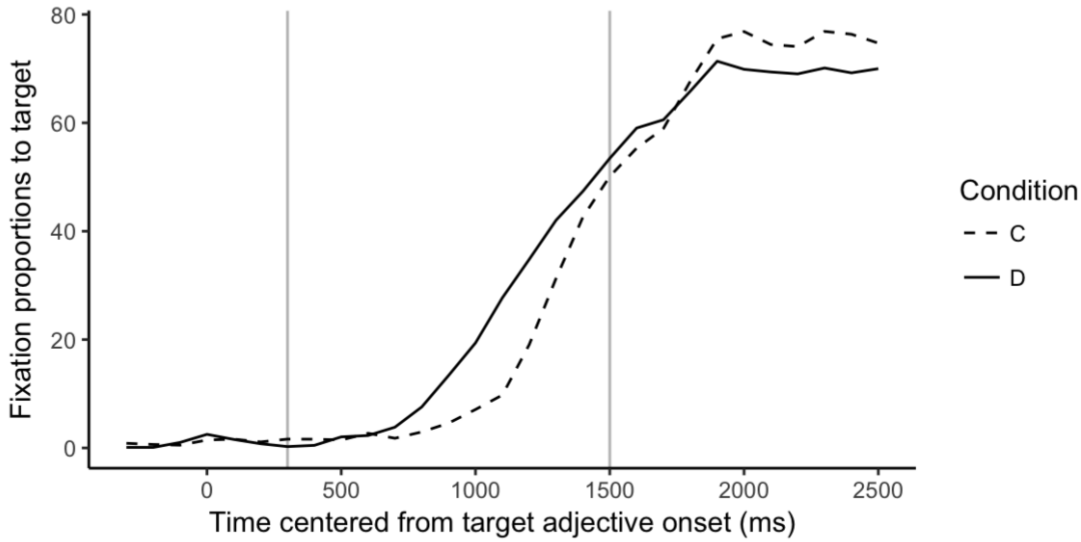
Figure 4), and target item in conditions A and B which presents an appropriate context for contrastive pitch accent (

Figure 5). In both

Figure 3 and

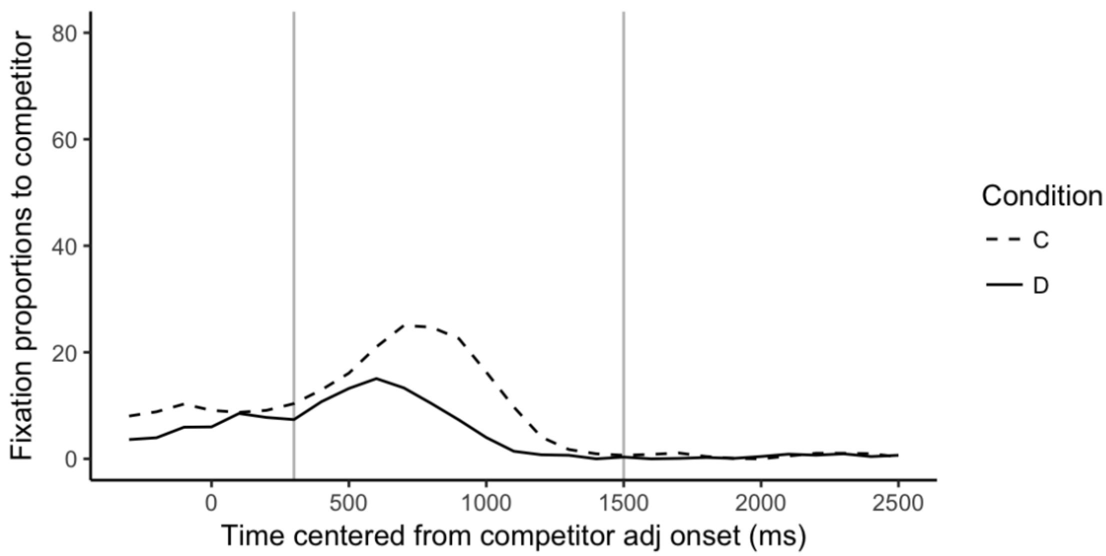
Figure 4, a clear separation was observed between two lines during the analysis window, which indicates a robust garden-path effect in TD participants. When participants heard an inappropriately placed contrastive pitch accent in condition C, their fixations to the competitor increased and their fixations to the target was delayed. The results of mixed-effect logistic regression confirmed the garden-path effect in TD participants. Two separate models were fitted with a fixed effect of condition and crossed random effects of subjects and items to predict participants' fixations to the target or competitor. A significant fixed effect of condition in both models confirmed the delayed fixation toward the target (β = 0.76, *SE* = 0.24, Wald's z = 3.01, p =.002) and the increased fixation toward the competitor (β = -1.03, *SE* = 0.26, Wald's z = -3.89, $p < .001$) in condition C compared to D. The odds of looking at the correct target was 53% times less (odds ratio = 0.47, 95% CI: [0.29, 0.76]) in condition C when a contrastive pitch accent was presented in an inappropriate context compared to the neutral condition D than in. The odds of looking at the incorrect competitor was 186% times (odds ratio = 2.86, 95% CI: [1.64, 4.76]) greater in condition C than condition D.

Figure 3. Mean fixation proportion to the correct target in conditions C (Inappropriate – Accented) and D (Inappropriate – Neutral) in TD participants



Note: The vertical lines indicate the analysis time window.

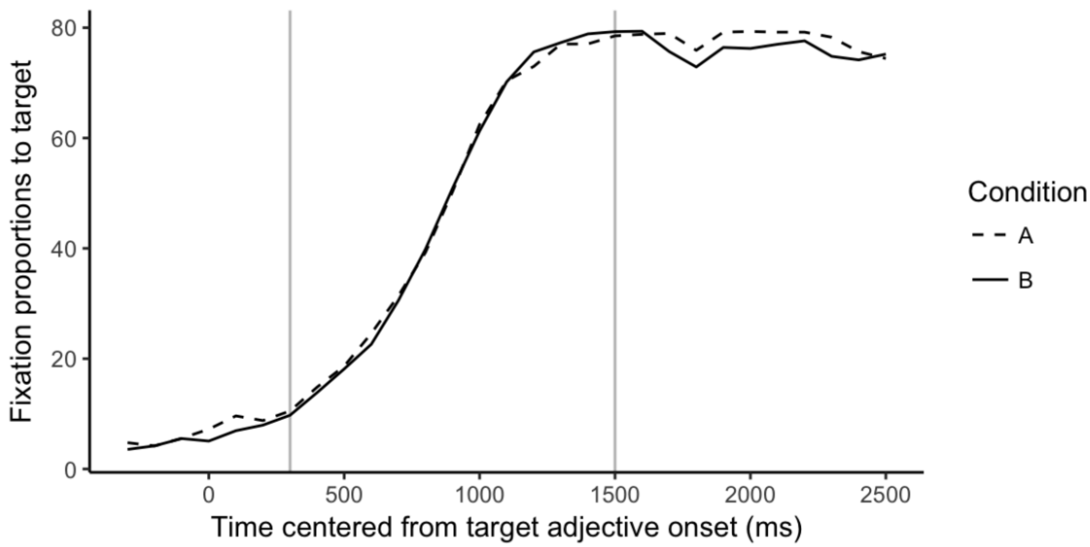
Figure 4. Mean fixation proportion to the incorrect competitor in conditions C (Inappropriate – Accented) and D (Inappropriate – Neutral) in TD participants



Note: The vertical lines indicate the analysis time window.

In contrast, we did not detect an anticipatory effect in TD participants. As shown in Figure 5, participants' fixations to the target in conditions A and B align with each other and did not differ significantly based on condition in a mixed-effect logistic regression model (β = .02, *SE* = .2 , Wald's z = .09, *p* = .93). Given that we replicated only the garden path effect but not the anticipatory effect of contrastive pitch accent using this paradigm, these findings support limiting the use of this paradigm for the comparison of ASD to TD to analyses of the garden patch effect as an objective and dimensional measure of comprehension of contrastive pitch accent in participants with ASD.

Figure 5. Mean fixation proportion to the correct target in conditions A (Appropriate – Accented) and B (Appropriate – Neutral) in TD participants

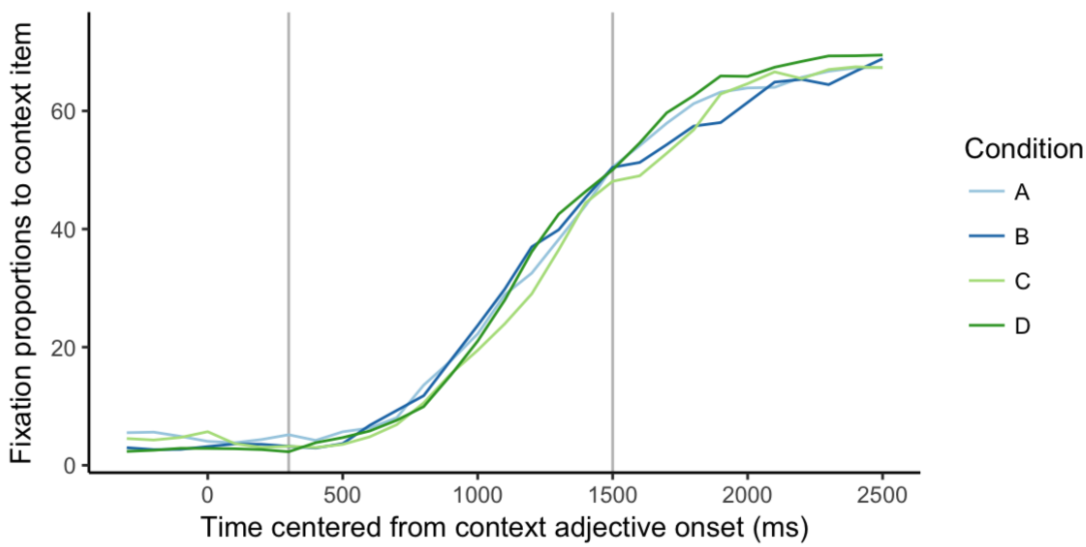


Note: The vertical lines indicate the analysis time window.

Fixations to context Item in TD participants. Finally, as a control analysis, participants' looks to context items during the response window of the context sentence were analyzed. Given that the context sentences in all conditions were not manipulated and contain neutral accented adjectives, participants' looks to context items should not differ. As shown in

Figure 6, participants' looks to context items in during the context response window in all four conditions align with each other and did not differ significantly based on condition (*p* = .67).

Figure 6. Mean fixation proportion to context items in all four conditions (A, B, C, D) in TD participants



Note: The vertical lines indicate the analysis time window.

3.3.1.2 Between-group comparison in comprehension of contrastive pitch accent

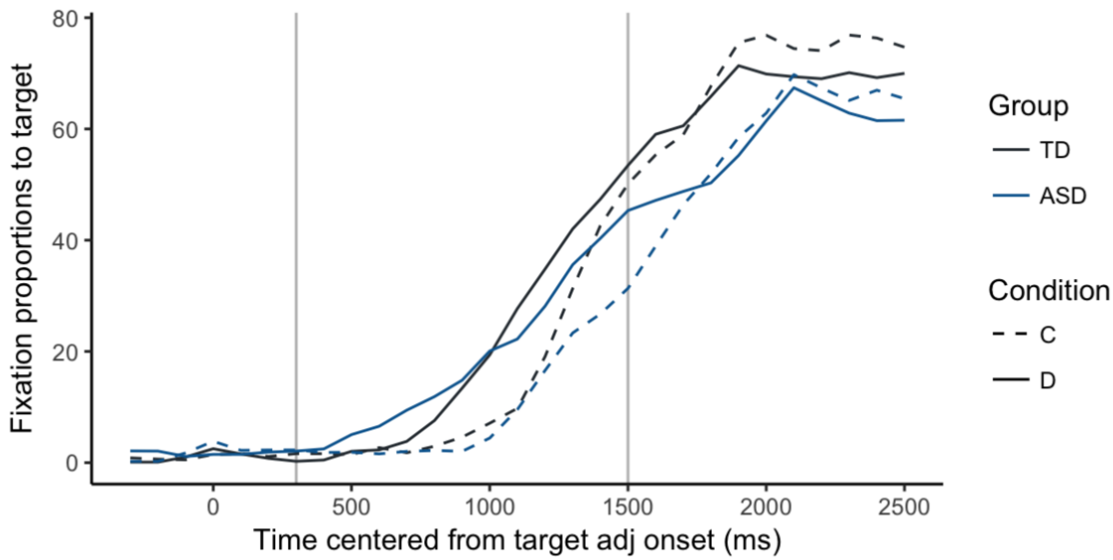
Figure 7 and

Figure 8 depict the mean fixation proportions to target (

Figure 7) and competitor item (

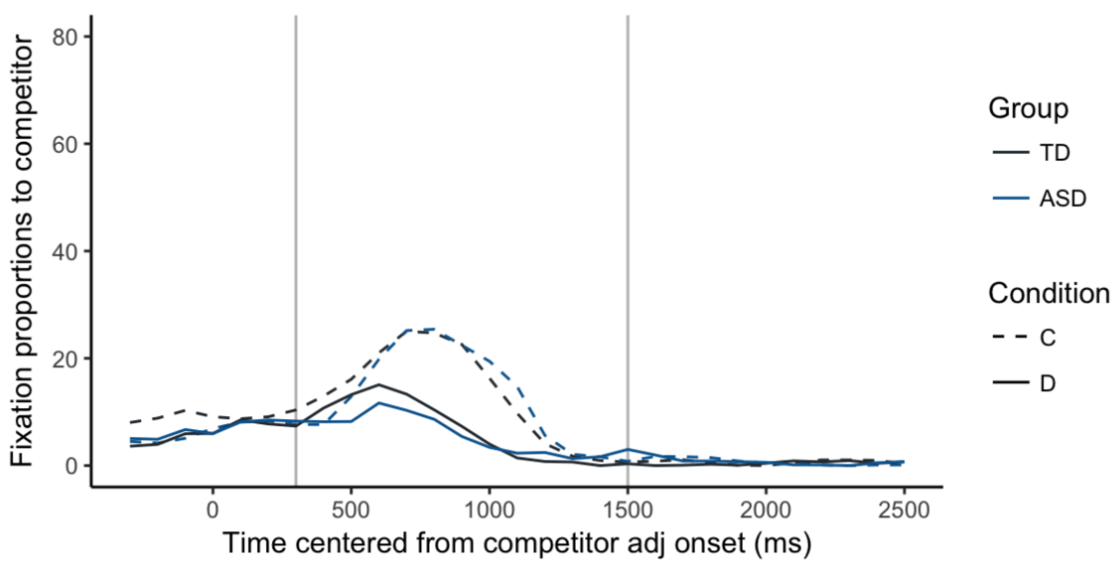
Figure 8) in conditions C and D for both TD and ASD groups. In both graphs, a clear separation was observed between the solid line and the dashed line during the analysis window for both groups, indicating a robust garden-path effect in both groups. In both groups, participant fixations to targets started rising later in condition C (fixations following the context-inappropriate prosodic cue) compared to condition D (fixations following a neutral sentence cue). Participants' fixations to competitor items followed the opposite pattern: fixations to competitor items started rising around 300 ms post target adjective onset in condition C. These early steep increases in fixations to the competitor item suggest that participants immediately comprehend the semantics of contrastive pitch accent on a prenominal adjective and anticipated the next item to be the same item as the context item they just heard before they received and processed the noun information that specified the correct target item.

Figure 7. Mean fixation proportion to the correct target in conditions C (Inappropriate – Accented) and D (Inappropriate – Neutral) in TD participants and participants with ASD



Note: The vertical lines indicate the analysis time window.

Figure 8. Mean fixation proportion to the incorrect competitor in conditions C (Inappropriate – Accented) and D (Inappropriate – Neutral) in TD participants and participants with ASD



Note: The vertical lines indicate the analysis time window.

Statistical analyses were conducted using two mixed-effect logistic regression models that include a fixed effect of group, a fixed effect of condition, their interaction, and crossed random effects of subjects and items. Results from both models showed a significant effect of condition with no significant group effect or group × condition interaction (Table 10).

Table 10. Summary of mixed-effect logistic regression analyses (fixed effects only) for binary fixation response to target data and for binary fixation response to competitor data

Variable	Binary fixations to target				Binary fixations to competitor			
	Estimate	***SE***	***z***	***p***	**Estimate**	***SE***	***z***	***p***
(Intercept)	-2.20	.22	-10.17	<.001***	-2.37	.25	-9.45	<.001***
Group	-0.32	.17	-1.89	.06	.02	.20	.09	.93
Condition	**0.52**	**.21**	**2.51**	**.01***	**-1.11**	**.26**	**-4.25**	**<.001***
Group × Condition	.22	.19	1.19	.24	-.05	.27	-.18	.86

Note: * p <. 05, *** p <. 001

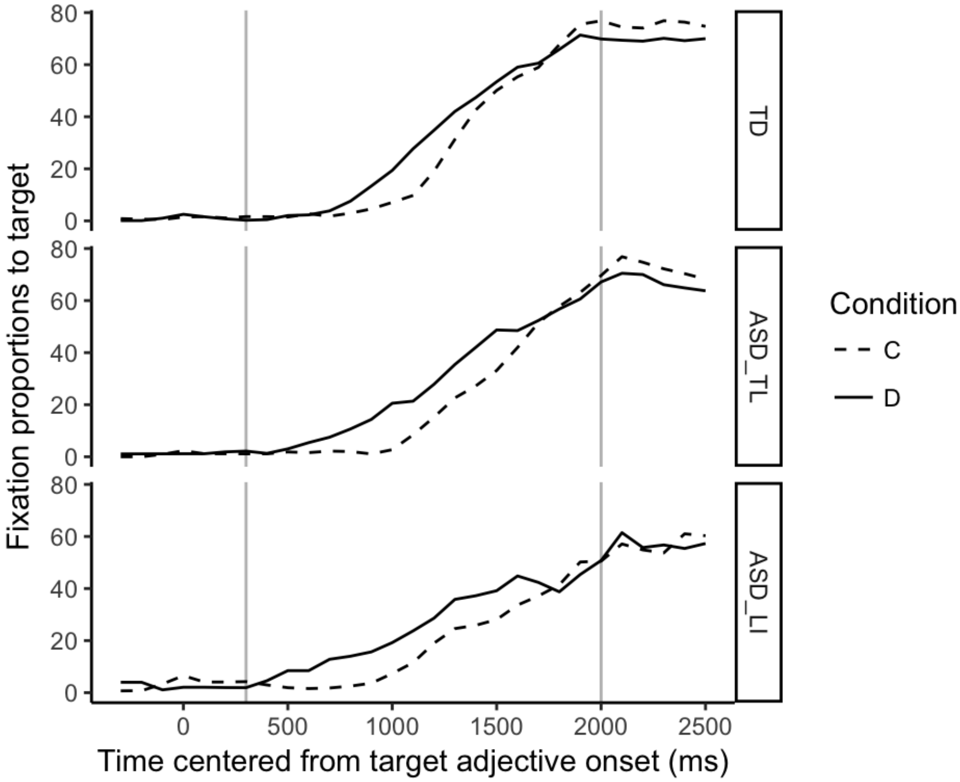
3.3.1.3 Relation of visual-world task performance and language ability

Given the significant difference in language ability between TD and ASD groups, we conducted post-hoc analyses to examine the impact of language ability on comprehension of contrastive pitch accent as measured in the visual-world task. Two sets of analyses were conducted. First, we fit a mixed-effect model with language and conditions as fixed effect, their interaction, and crossed random effects of subject and item. Results revealed a significant fixed effect of language (β = .01, *SE* = .00, Wald's *z* = 1.95, *p* =.05), a significant fixed effect of condition (β = .24, *SE* = .12, Wald's *z* = 2.03, *p* = .04), with no interaction (*p* = .51).

Additionally, subgroup analyses were conducted by submitting subgroup membership and conditions as fixed effects, their interaction, and crossed random effects of subject and item. As shown in

Figure 9, children in all three subgroups demonstrated a robust garden-path effect of contrastive pitch accent by displaying a clear separation between the two lines for conditions C and D. In addition, there is an apparent difference in the slope of the change in fixation proportion over time between the three groups (TD > ASD+ TL > ASD+ LI). Post-hoc analyses using binary fixation to target as the dependent measure revealed a significant pairwise contrast only between TD and ASD+LI groups (β = .-.47, *SE* = .16, Wald's *z* = -2.89, *p* = .004), but not between TD and ASD+TL (*p* = .17) or between the two ASD subgroups (*p* =.12). When translated into odds ratio, these results indicate that participants with ASD with LI is associated with a lower odds of looking at target than TD participants (odds ratio = .625). While participants with ASD with typical language is also associated with a lower odds of looking at target (odds ratio = .83), this difference is not statistically significant.

Figure 9. Mean fixation proportion to the target item in conditions C and D for the TD group, the ASD+TL subgroup, and the ASD+LI subgroup



Note: The vertical lines indicate the analysis time window. A wider analysis window from 300 to 2000 ms was used for this analysis to accommodate the processing speed in children in the ASD+LI group.

3.3.2 Research question 2

3.3.2.1 Construct validity of visual-world task-related measures

Three task-related measures were selected from the visual-world task: a) contrastive pitch accent comprehension (random by-participant slopes); b) speed of contrastive pitch accent processing (latency of first fixation in condition C - i.e. inappropriate-accented), and c) speed of general linguistic processing (latency of first fixation in condition D - i.e. inappropriate-neutral). To examine the construct validity of these three task-related measures, correlation analyses were conducted to measure the association between these task-related measures with the Contrastive

Pitch Accent Understanding subtest scores from the PEPS-C and the CELF-5 standard score within the ASD group. As shown in Table 11, the task-related contrastive pitch accent comprehension measure did not correlate with either the specific measure of prosody perception (PEPS-C subtest) or the general measure of language ability (CELF-5). In contrast, a significant and differential pattern of associations was found for the two speed of performance measures. The task-related measure of speed of contrastive pitch accent processing was significantly correlated with the Contrastive Pitch Accent Understanding subtest but not language measured by CELF-5. The opposite pattern of associations was found for the task-related measure of speed of general linguistic processing (significant correlation with language, but not the specific measure of prosody perception).

Table 11. Correlations of visual-world task-related measures and standardized test measures of contrastive pitch accent understanding and language in the ASD group

Construct	Comprehension of CPA		Speed of CPA processing		Speed of General Linguistic Processing	
Measure	Random By-Participant Slope		Latency of First Fixation to Target in Condition C		Latency of First Fixation to Target in Condition D	
	r	*p*	*r*	*p*	*r*	*p*
PEPS-C CPA Understanding Subtest	-.18	.47	**-.33***	**.03**	-.22	.16
Language (CELF-5 SS)	.13	.56	-.22	.16	**-.32***	**.04**

Note. CPA = Contrastive Pitch Accent; **p* <.05.

3.3.2.2 Relations between visual-world task-related measures and clinical measures in ASD

Correlation analyses were conducted within the ASD group to examine the relation between task-related measures (comprehension of pitch accent, speed of processing of pitch accent, speed of processing of linguistic cues) and clinical measures of impairment in ASD, including: a) receptive prosody (receptive prosody composite from PEPS-C); b) pragmatic language ability (CELF-5 Metalinguistic Meta-Pragmatic Index); c) social communication impairment (Social Responsiveness Scale); and d) ASD overall symptom severity (ADOS-2 Total Score). As shown in Table 12, significant correlations between visual-world task measures and clinical measures were only found for the two task-related speed of processing measures. No significant correlations were found between the task-related comprehension of pitch accent measure and the clinical measures. Pragmatic language ability was significantly correlated with both specific (contrastive pitch accent) and general (linguistic cues) task-related speed of performance measures. Autism severity was significantly correlated with only the task-related speed of pitch accent performance.

Table 12. Correlations of visual-world task-related measures and standardized test measures of clinical impairment for the ASD group

Construct	Comprehension of CPA		Speed of CPA processing		Speed of General Linguistic Processing	
Measure	Random By-Participant Slope		Latency of First Fixation to Target in Condition C		Latency of First Fixation to Target in Condition D	
	r	*p*	*r*	*p*	*r*	*p*
PEPS-C Receptive Prosody Composite	.15	.56	-.21	.18	-.15	.34
Pragmatic Language (CELF-5 Metalinguistic Meta-Pragmatic Index)	-.09	.74	**-.34***	**.02**	**-.40****	**.008**
Social Communication (SRS Communication)	-.02	.95	-.02	.93	.18	.47
ASD symptom severity (ADOS-2 Total Score)	.01	.98	**.62****	**.006**	-.08	.76

Note. CPA = Contrastive Pitch Accent; **p* <.05; ***p*<.01.

3.4 Discussion

3.4.1 Summary of findings

Our findings provide evidence that the visual-world paradigm is a feasible and valid approach for measuring comprehension of specific prosodic cues (i.e. contrastive pitch accent) in children with ASD. We found evidence that our sample of children with ASD can perceive contrastive pitch accents on an acoustic-perceptual level (AX task) and also can demonstrate comprehension of the semantic content of contrastive pitch accent during spoken language comprehension (visual-world task). We did not find evidence for significant differences between the ASD and TD groups in participant's ability to comprehend contrastive pitch accent. However, the results of post-hoc analyses indicated that the subgroup of children with ASD and comorbid language impairments demonstrated adequate comprehension of pitch accent cues but were slower in their processing of both contrastive pitch accent specific and general linguistic cues relative to children with ASD with typical language ability. In line with this, we also found that only speed-related performance measures from the visual-world task correlated significantly with measures of clinical impairment in participants with ASD.

One key finding that emerged is that children with ASD in this study were found to be sensitive to contrastive pitch accent as evidenced by a robust garden-path effect. To our knowledge, four previous studies have examined comprehension of contrastive pitch accent in ASD (Järvinen-Pasley et al., 2008; Lyons et al., 2014; McCann et al., 2007; Paul, Augustyn, et al., 2005). Our finding is consistent with Järvinen-Pasley et al. (2008), McCann et al. (2007), and Lyons et al. (2014) but inconsistent with Paul et al. (2005). Both Järvinen-Pasley et al. (2008) and McCann et al. (2007) used Contrastive Pitch Accent Understanding subtest from PEPS-C and reported no significant differences between diagnostic groups. The divergence between our study and Paul et al. (2005) could be explained by methodological differences. Paul and colleagues presented participants with pairs of sentences on paper (e.g. "Do you want vanilla?" and "Do you want chocolate cake?") and asked participants to listen to another sentence on the tape (e.g. "I want CHOCOLATE ice-cream"). Participants were asked to check off the sentence that should come before the one they just heard. Given that this task involves longer and more complicated instructions than the visual-world paradigm, the differences observed groups may be driven by either poor language comprehension, poor auditory memory, or cognitive deficits in children with ASD.

Our findings most closely aligned with findings reported by Lyon et al. (2014). In Lyons and colleagues' work (2014), the authors first reported no significant difference in contrastive pitch accent understanding across diagnostic groups. However, once participants with ASD were categorized into subgroups of high-language (ASD-Hi) and low-language (ASD-Lo) based on their performance on standardized language assessment, a pairwise difference between TD and ASD-lo subgroup emerged in ability to comprehend contrastive pitch accent. In our study, although we found that children with ASD with LI showed clear evidence of comprehending and using contrastive pitch accents in the visual-world task (i.e. a robust garden path effect), they performed significantly worse than children with ASD with typical language and TD children in two other aspects. First, they fixated on the target significantly slower in both accented and neutral conditions during the response window. Second, though not the focus of this study, we also observed that the ASD-LI group also demonstrated lower peak amount of fixations after the response window.

Previous studies using the visual-world paradigm outside of autism have reported evidence that individual differences in performance on the task may be driven by the presence or absence of language impairment. For example, McMurray and colleagues used the visual-world paradigm to evaluate spoken word recognition in groups of typically developing adolescents who varied in language and cognitive abilities (McMurray, Munson, & Tomblin, 2014). They noted that participants with LI rather than cognitive impairment showed a delayed fixation and a lower peak of fixation. Similarly, another visual-world paradigm study that examined language processing in children with ASD reported no significant difference between ASD and TD but detected significant difference in children with LI with or without ASD (Brock, Norbury, Einav, & Nation, 2008). Though few in numbers, these studies provide convergent evidence for delayed and reduced fixations for target words in individuals with LI. Taken together, the patterns we observed in the group of children with ASD and LI appear to be consistent with previous studies in this area and together these studies suggest that children with ASD with a comorbid language impairment should be considered as a separate and distinct subgroup of children with ASD.

Finally, we found evidence for both the construct validity and the criterion-related validity of two speed of performance measures derived from our visual-world task within our sample of ASD children. Specifically, task-related speed of contrastive pitch accent processing was found to associate with another receptive measure of contrastive pitch accent from the PEPS-C and task-related speed of general linguistic processing was found to be associated with overall language ability as measured by a standardized language assessment. These two task-related speed of performance measures also correlated with clinical impairment measures in the expected direction. Task-related speed of contrastive pitch accent processing correlated with pragmatic language deficits and with overall ASD symptom severity. Further, task-related speed of general linguistic processing correlated with general language ability and pragmatic language ability. To our knowledge, this is the first study that shows a significant correlation between processing speed of specific prosodic cues and ASD symptom severity. These findings suggest that the visual-world task-related speed of performance validly measures individual differences in persons with autism in their efficiency of using contrastive pitch accent and processing general linguistic information. Additionally, these individual differences in contrastive pitch accent and linguistic processing are associated more broadly with measures of a variety of clinical impairments in autism. Importantly, the correlations between task performance and clinical measures found in this study were modest in magnitude and clearly indicate that a variety of other and as yet unknown factors is related to individual differences in pitch accent performance in children with ASD.

An unexpected finding was that we found evidence for both construct and criterion validity for two task-related speed of performance measures but not for a task-related measure of comprehension of contrastive pitch accent cues. When considered along with the fact that we found no evidence for group differences between the ASD and TD groups on comprehension of contrastive pitch accent cues in the visual world task (i.e. lack of group × condition interaction), the bulk of the evidence from this study indicates that it is likely speed of processing deficits rather than comprehension deficits that contribute most to receptive prosody ability in children with ASD.

3.4.2 **Limitations**

One limitation of the current study is that we were only able to replicate the garden-path effect of contrastive pitch accent but not the anticipatory effect reported by Ito et al. (2014). Although this is expected based on previous literature that the anticipatory effect tends to have smaller effect size than the garden-path effect (Ito et al., 2013; D. G. Watson, 2010), it seems likely that methodological differences between our study and Ito et al. (2014) can explain the lack of anticipatory effect in our study. We used a more specific item-wise area of interest (AOI) instead of a bigger cell-wise AOI as in Ito et al. (2014). Sentences used to examine the anticipatory effect include the same item with different colors as context item and target item across context and target sentences. It was assumed in Ito et al. (2014) that participants would scan outside the cell that contains both context and target items once they locate the context item and then come back to the same cell to look for the target item. Trials where a participant kept fixating on the context item within the cell were removed before analysis. When we re-analyzed our data using the bigger cell-wise AOIs, we found that within the pause following the context sentence, our participants' fixations stayed within the cell after locating the context item and thus only needed to move a small amount to look at the target item. The short distance between the context item and the target item may not be sufficient to show an anticipatory effect in our design. A longer pause following the context sentence may give participants sufficient time and encourage them to scan outside the cell to explore other items in the display.

A second piece of evidence relevant to our lack of anticipatory effects in our task comes from the subgroup of children with ASD with LI. Of interest, the anticipatory effect is most evident in this group with reduced linguistic processing speed. It appears that in a group of children with poor language processing abilities, participants were able to show a faster response to target item when a context-appropriate contrastive pitch accent was used. In other words, this observation provides additional support that for TD children and children with ASD without LI, the short distance between context and target items may not be sufficient to demonstrate anticipatory effects.

Additionally, our analysis sample only included children with ASD without intellectual disability. We included a cognitive criterion (i.e. an IQ above 70) to ensure that participants in TD and ASD groups were matched on age and cognitive ability. Six children with ASD were recruited to participate but were excluded in the final analyses because they did not meet the cognitive criterion. Though language was not an inclusion criterion, participants with ASD in this sample all have fluent and flexible use of language. One caveat of such a sample is that findings from this current study may not generalize to other subgroups of children with ASD, namely minimally verbal children with ASD. It is important to note that all six children with ASD who were excluded sat through and completed both the control AX task and the visual-world paradigm. Additionally, all six children demonstrated ability to perceive contrastive pitch accent by reaching ceiling on the AX task. Though these children yielded significantly more trackloss samples and more inattentive trials compared to children with ASD without intellectual disability or TD children, they still provided on average 46 usable trials (64% of total number of trials). These data suggest that this paradigm has the potential to be used with this subgroup of children with ASD. Future studies may test this paradigm with a larger sample of children with ASD with intellectual disability to confirm the feasibility of this paradigm.

Finally, though we are interested in understanding individual differences in prosody processing in children with ASD and have attempted to establish construct validity and criterion-related validity for task-related measures, the reliability of the parameters from our experimental prosody perception task needs to be confirmed in future studies. Although recent research has begun to use online measures to study meaningful clinical difference during spoken language processing in clinical populations (Brock et al., 2008; McMurray, Samelson, Lee, & Bruce Tomblin, 2010), stability and reliability of these online measures need to be determined first.

Farris-Trimble and McMurray have eloquently summarized that in order to be considered as a measure of individual difference, a visual-world paradigm measure needs to assess a consistent aspect of an individual's behavior rather than a general pattern that's meaningful only as a group measure (Farris-Trimble & McMurray, 2013). This measure also needs to be sensitive enough to reveal subtle effects during spoken language processing. Only one previous study have examined reliability of visual-world paradigm measures: Farris-Trimble and McMurray examined test-retest reliability of parameters from a visual-world paradigm for spoken word recognition in a group of typical adults and found that fixations to target parameters (slope and asymptote) were highly reliable. Despite these initial evidence on the reliability of visual-world paradigm parameters as measures of individual differences, more studies replicating these findings across various visual-world paradigm designs in clinical populations are needed to determine aspects of visual-world paradigm that can accurately capture clinically relevant individual differences.

3.4.3 Future directions

The results of our study indicate that visual-world paradigm can be an objective, informative, and feasible approach to study comprehension of prosody in children with ASD. Future studies can use this approach to investigate comprehension of prosodic elements other than contrastive pitch accent, such as intonation and pause, to detect potential additional barriers to language comprehension and social communication.

Additionally, given our finding that children with ASD with LI showed significantly different processing patterns in this paradigm, it would be interesting to test comprehension of prosody in children with LI without ASD to further elucidate the impact of ASD versus LI on comprehension of prosody.

3.4.4 Clinical implications

One interesting observation from this study was that the children with ASD with LI in this study showed a robust garden-path effect in the visual-world paradigm, which indicated sufficient comprehension of contrastive pitch accent during spoken language comprehension, yet failed the Contrastive Pitch Accent Understanding subtest in the PEPS-C, which is a standardized test of prosody. This inconsistency may be explained by two possibilities. One possibility is that given that PEPS-C requires more language comprehension than the visual-world paradigm, the poor performance on the PEPS-C may be driven by language ability rather than the specific ability to comprehend prosodic cues. A second plausible explanation is that even though children with ASD with LI are sensitive to contrastive pitch accent during spoken language comprehension, they are slower at displaying the effect compared TD children and children with ASD without LI. A 500 ms delay in resolving garden-path effect during online processing may translate into marked deficits on a standardized assessment of prosody comprehension given how quickly spoken language evolves over time in the context of usual discourse. Thus, from a clinical perspective, it is important to note that prosodic deficits may manifest in children with ASD in a processing efficiency or speed-related manner beyond the broader categorical distinction of whether an individual can comprehend or use prosodic cues.

Our findings also highlight the possibility that it is children with ASD and comorbid LI, but not those without LI, who will manifest the greatest evidence for prosody related deficits. This indicates the importance of assessing language ability in children with ASD. Although it is unclear if language deficits in children with ASD are best conceptualized as a continuum of heterogeneity that can be explained by additional factors (e.g. cognitive, environmental, genetic) or as a comorbidity of language impairments (Boucher, 2012), it is clear that a subset of children with ASD present with significant language concerns including deficits in the processing of prosodic cues. Our findings support the idea that better identification of this subset of ASD in future studies is likely to help the broader effort to better understand heterogeneity in ASD etiology, mechanism, course, and intervention outcome.

CHAPTER 4

General Discussion

4.1 Summary of this Work

This work was originally motivated by my interest in child-directed speech, which is a type of speech with distinctive and exaggerated prosodic characteristics that caregivers use when interacting with young children (Cristia, 2013; Ferguson, 1964; Fernald & Simon, 1984). An original dissertation topic idea was to investigate the extent to which children with ASD learn language from prosodic input such as child-directed speech. However, a closer literature review revealed a critical gap in the autism field regarding the perception and comprehension of prosodic input. The vast majority of existing research on prosody in children with ASD has primarily focused on expressive abilities (Peppé & McCann, 2003), leaving receptive prosody poorly understood. Gaining a solid understanding of the perception and comprehension of prosody in children with ASD is a necessary first step toward studying the extent to which children with ASD learn from prosodic input such as child-directed speech. In other words, through serendipity, my interests in child-directed speech has guided me to the topic of prosody in ASD and the fascinating challenge of creating novel paradigms to measure receptive prosody in a clinical population.

In Chapter 1, I reviewed the role prosody serves in language development for language learners and spoken language communication for proficient language users. From early on in life, infants are sensitive to various prosodic elements and take advantage of prosodic cues to facilitate language learning. Extensive previous work have shown that prosody supports word segmentation (Jusczyk et al., 1999), word learning (Herold et al., 2011), syntactic segmentation

(Gleitman & Wanner, 1982; Jusczyk et al., 1992), and emergent literacy (Beattie & Manis, 2014). For adults who are proficient language users, prosodic cues facilitates parsing and disambiguating syntactic information (Kjelgaard & Speer, 1999; Snedeker & Trueswell, 2003), resolving referent in discourse contexts (Arnold, 2008; Terken & Nooteboom, 1987), and improving recognition memory (Fraundorf, Watson, & Benjamin, 2010b; Shintel et al., 2014). Prosody is unique in a sense that it is an integrated part of language yet it conveys information about the linguistic signal itself at the same time. Because of this unique characteristic of prosody, deficits in perceiving or comprehending prosody can have significant consequences on language development and social communication. For infants or young children who are language learners, reduced attention to or difficulty in perceiving prosodic cues may cost them valuable bootstraps that facilitate language processing and result in cascading effects on later language development. For adults, difficulty in perceiving or comprehending prosody can have a direct functional impact in everyday living. To illustrate, the importance of receptive prosody can be evident when an individual with ASD misses a job opportunity because he or she misinterpreted a question with raising intonation as a statement and did not answer, or when a child with ASD fails to make friends at school because he or she constantly misunderstands other children's meaning due to difficulties differentiating sarcastic tones from genuine tones. The narrative review in this chapter highlighted the complexity and significance of prosody.

Chapter 2 presented a systematic review on receptive prosody deficits in individuals with ASD. Prosodic deficits have been a clinical symptom of ASD throughout the history of this syndrome. In Kanner's landmark 1943 paper where he first characterized ASD, he described some among his patients as exhibiting "odd intonation". In the latest revision of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013), prosody was still listed under persistent deficits in social communication and social interaction across multiple contexts:

> Deficits in nonverbal communicative behaviors used for social interaction are manifested by absent, reduced, or atypical use of eye contact, gestures, facial expressions, body orientation, or speech intonation... Interaction may be relatively subtle within individual modes (e.g. someone may have relatively good eye contact when speaking) but noticeable in poor integration of eye contact, gesture, body posture, prosody, and facial expression for social communication (APA, 2013, p. 54).

Therefore, this systematic review was motivated by the significance of prosody in language development and social communication and persistent disruptions in prosody reported in previous literature and observations from my own clinical practice. Using inclusion criteria developed *a priori*, we searched for studies that focused on comparison of receptive prosody ability between TD individuals and individuals with ASD published between 1980 and 2019. The final sample of reviewed articles includes 25 peer-reviewed publications. A detailed review of these studies revealed several important findings. First, semantic or pragmatic aspect of receptive prosody has received the least attention. Only 16% of existing studies in individuals with ASD examined semantic or pragmatic aspect of receptive prosody (e.g. contrastive pitch accent). Second, previous studies in ASD that have measured receptive prosody were limited by the methods they used. The majority of studies used a linguistic task that involves extensive linguistic and cognitive demands on participants. Therefore, the findings from these study may relate more to language or cognitive deficits rather than prosody-specific deficits. Third, receptive prosodic deficits in individuals with ASD may vary as a function of stimuli complexity and task demands. Fourth, there are some evidence suggest that receptive prosodic deficits may only be evident in a subgroup of individuals with ASD, namely individuals with ASD with comorbid language impairments. Lastly, limited studies have examined associations between receptive prosody and broad clinical measures. This systematic review identified existing gaps regarding receptive prosodic deficits in individuals with ASD and limitations in current methods used to measure receptive prosody in individuals with ASD and motivated the focus and the design of the experimental study reported in Chapter 3.

Chapter 3 described the development of a visual-world paradigm adapted from Ito et al. (2014) and the application of this adapted task paradigm to examine the comprehension of contrastive pitch accent in children with ASD. I added a set of methodological adaptations to the original Ito et al. (2014) visual-world paradigm to tailor to specific needs associated with children with ASD. For instance, a control task was added to confirm that participants possess the ability to perceive and discriminate contrastive pitch accent before assessing their ability to comprehend contrastive pitch accent during online processing. Also, unique items for all trials were added to the task to avoid possible learning effects driven by repeated items. All images used in the task were tested in a pilot study with a separate sample of children and adults to confirm that images are recognizable and familiar. Using this improved study design, I replicated the garden path effect but not the anticipatory effect in typically developing children between 8 and 14. These findings confirmed the feasibility of using this paradigm in children and support limiting the use of the visual-world paradigm to analyses of the garden path effect to measure comprehension of contrastive pitch accent.

I then used this adapted visual world task to compare comprehension of contrastive pitch accent between TD children and children with ASD between 8 and 14 and examined associations between task parameters and clinical measures. Overall findings from this study demonstrated that the visual-world paradigm is feasible for children with ASD and that children with ASD in this age range are sensitive to contrastive pitch accent during spoken language comprehension. Additionally, consistent with one previous study (Lyons et al., 2014), we found that one specific subgroup of children with ASD with language impairment showed delayed processing that's not specific to prosodic cues despite a robust garden-path effect that indicates their ability to comprehend contrastive pitch accent. Interestingly, this is also the group that showed marked deficits on a standardized assessment of receptive prosody (data not reported within this dissertation, but available from the author). Regarding association between task-related measures from the visual-world task and clinical measures, our results provided evidence for both the construct and the criterion-related validity for task-related speed-of-processing measures (latency of first fixation). Specifically, speed of contrastive pitch accent processing correlated with another measure of contrastive pitch accent understanding from PEPS-C, pragmatic language, and ASD symptom severity. Speed of general linguistic processing correlated with a clinical measure of general language and pragmatic language abilities.

Overall, the present body of work represents the first eye-tracking paradigm to assess comprehension of semantic/pragmatic prosody in individuals with ASD. The new knowledge generated by this work achieved the goal of developing an objective measure of comprehension of specific prosodic cues. Findings from this study provide a crucial first step in accurately capturing specific prosodic deficits in individuals with ASD and pave the way for future study that seeks to understand the mechanisms that underlie language deficits and social communication difficulties in ASD.

4.2 Future Directions

This body of work can fuel future studies in three directions as follows: a) perception ad comprehension of elements of prosody other than contrastive pitch accent; b) perception of and attention to infant-directed speech in infancy and early childhood; and c) potential ways to improve sensitivity to prosodic cues in children with ASD.

To begin with, this work confirmed that the visual-world paradigm is an objective, informative, and feasible approach to study comprehension of contrastive pitch accent during spoken language processing in children with ASD. A natural extension of this current work is to use this approach to investigate comprehension of prosodic elements other than contrastive pitch accent, such as intonation and pause, to detect potential additional barriers to language comprehension and social communication

A second possible direction is to use the visual-world eye-tracking approach to understand attention to and learning from prosodic input in infants and toddlers with ASD or at-risk for ASD. Extensive research on typical development has consistently shown that infant-directed speech, characterized by unique prosodic features, facilitates early language development. A recent meta-analysis has confirmed that the distinctive, exaggerated prosodic features in infant-directed speech primes infants' attention and facilitates communicative development (Spinelli et al., 2017). Yet, previous studies in ASD have also suggested that children with ASD display reduced attention to infant-directed speech and accordingly have fewer access to opportunities to benefit from the facilitate effect of infant-directed speech (Paul, Chawarska, Fowler, Cicchetti, & Volkmar, 2007; Watson, Baranek, Roberts, David, & Perryman, 2010). A reasonable next step in this line of research is to use the visual world eye-tracking approach developed in the proposed study with a simplified scene to investigate attention to and learning from infant-directed speech in infants and toddlers at-risk for ASD. Attention to infant-directed speech has been found to correlate concurrently and longitudinal with communication skills in children with ASD (Watson et al., 2010; Watson, Roberts, Baranek, Mandulak, & Dalton, 2012). Therefore, studying the extent to which infants and toddlers with ASD or at-risk for ASD attend to and learn from prosodic cues in infant-directed speech may shed light on early language learning experience in this population and provide insights into avenues to effective early intervention tailored for children with ASD.

Finally, in addition to understand perception and comprehension of prosody in this population, it is critical for future research to study potential ways to improve one's sensitivity to prosodic cues. In this study, we found that a subgroup of children with ASD showed delayed processing during the visual-world paradigm and marked deficits on a standardized prosody assessment. Thus, it would be informative to study the extent to which sensitivity to prosody or efficiency of prosody processing is malleable to change. An intervention study could examine whether explicit instructions or increasing the amount of input with prosodic cues could improve receptive prosody in this population.

4.3 Conclusion and Implications

Overall, finding from this dissertation can be summarized by three themes: a) findings established the feasibility of using a visual-world paradigm to test receptive prosody in children with ASD; b) as a group, children with ASD demonstrate ability to perceive and comprehend a specific type of prosody, contrastive pitch accent, as well as TD children; c) a subgroup of children with ASD with language impairment showed delayed processing during online spoken language comprehension and also marked deficits in a standardized receptive prosody assessment. The significance of this work is related to potential extensions to both basic and translational research. The novel visual-world paradigm allows close examinations of intermediate states during prosody processing and general linguistic processing and has the potential to reveal mechanisms that contribute to heterogeneous language and social communication outcomes that we see in individuals with ASD. This work also carries clinical significance. Although previous work have shown that children with ASD can develop pronounced expressive and receptive prosody, to our knowledge, prosody has rarely been incorporated into work on intervention development for individuals with ASD. Given that "what gets measured gets treated," developing a way to objectively measure prosodic deficits is a critical first step towards expanding intervention efforts in ASD to include a focus on remediating prosodic deficits. Studying comprehension of a specific type of prosody in a controlled task allowed us to quantify the type and the magnitude of receptive prosodic deficits and thus has the potential to guide future studies to develop specific and effective interventions to improve language and social outcomes in children with ASD.

REFERENCE

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(3), 714–735. https://doi.org/10.1037/0278-7393.22.3.714

American Psychiatric Association. (2013). *DSM-V. American Journal of Psychiatry*. https://doi.org/10.1176/appi.books.9780890425596.744053

Arnold, J. E. (2008). THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition*, *108*(1), 69–99. https://doi.org/10.1016/j.cognition.2008.01.001

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56. https://doi.org/10.1177/00238309040470010201

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Beattie, R. L., & Manis, F. R. (2014). The relationship between prosodic perception, phonological awareness and vocabulary in emergent literacy. *Journal of Research in Reading*, *37*(2), 119–137. https://doi.org/10.1111/j.1467-9817.2011.01507.x

Beckman, M. (1992). Evidence for speech rhythms across languages. In *Speech Perception,* *Production and Linguistic Structure* (pp. 457–463).

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*. https://doi.org/10.1080/016909696387213

Begelsonm, E., Bergmann, C., Byers-Heinlein, K., Cristia, A., Cusack, R., Dyck, K., … Soderstrom, M. (2017). Quantifying sources of variability in infancy research using the infant-directed speech preference. *PsyArXiv*, 1–46. https://doi.org/https://doi.org/10.31234/osf.io/s98ab

Bertone, A., Mottron, L., Jelenic, P., & Faubert, J. (2005). Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity. *Brain*, *128*(10), 2430–2441. https://doi.org/10.1093/brain/awh561

Birch, S., & Clifton Jr, C. (1995). Focus, accent, and argument structure: effects on language comprehension. *Language and Speech*, *38*, Pt 4/. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/002383099503800403

Bolinger, D. (1961). Contrastive accent and contrastive stress. *Language*, *37*(1), 83–96. https://doi.org/10.1371/journal.pone.0005772

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language*, *48*(3), 633. https://doi.org/10.2307/412039

Bolinger, D. (1982). Intonation and Its Parts. *Language*, *58*(3), 505–533. https://doi.org/10.1371/journal.pone.0005772

Bolinger, D. (1983). Intonation and Gesture. *American Speech*, *58*(2), 156–174. Retrieved from https://www.jstor.org/stable/pdf/455326.pdf?casa_token=JAeETixojNUAAAAA:616ZP8gk4wA61DDFBOtX0WxaS4pPmcjuUyS4PNgkwcVhBeieSde4lxED1lnjZwFIr-SxamPWp071ZFKd-rKaksPeknMPEGW6WRuNDi7zsU1bIh7s3yjM

Bolinger, D. (1989). *Intonation and its uses : melody in grammar and discourse*. Stanford, California: Stanford University Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=xFuUFtDyqJsC&oi=fnd&pg=PA1&dq=bolinger+1989&ots=naY08c9U0p&sig=MYO1e1_hQ3_m7fvWVJ0OUHa-hmM#v=onepage&q=bolinger 1989&f=false

Bone, D., Black, M. P., Ramakrishna, A., Grossman, R. B., & Narayanan, S. (2015). Acoustic-prosodic correlates of "awkward" prosody in story retellings from adolescents with autism. In *Proceedings of the Annual Conference of the International Speech Communication Association* (Vol. 2015-Janua, pp. 1616–1620). Retrieved from http://

Bonnel, A., McAdams, S., Smith, B., Berthiaume, C., Bertone, A., Ciocca, V., … Mottron, L. (2010). Enhanced pure-tone pitch discrimination among persons with autism but not Asperger syndrome. *Neuropsychologia*, *48*(9), 2465–2475. https://doi.org/10.1016/j.neuropsychologia.2010.04.020

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A.-M. (2003). Enhanced pitch sensitivity in individuals with Autism: a signal detection analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226–235. https://doi.org/10.1162/089892903321208169

Brennand, R., Schepman, A., & Rodway, P. (2011). Vocal emotion perception in pseudo-sentences by secondary-school children with Autism Spectrum Disorder. *Research in Autism Spectrum Disorders*, *5*(4), 1567–1573. https://doi.org/http://dx.doi.org/10.1016/j.rasd.2011.03.002

Brock, J., Norbury, C., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition*, *108*(3), 896–904. https://doi.org/10.1016/j.cognition.2008.06.007

Brooks, P. J., Gaggi, N. L., & Ploog, B. O. (2018). Generalization of content and emotional prosody across speakers varying in gender in youth with autism spectrum disorder. *Research in Developmental Disabilities*, *83*, 57–68. https://doi.org/http://dx.doi.org/10.1016/j.ridd.2018.08.004

Brooks, P. J., & Ploog, B. O. (2013). Attention to emotional tone of voice in speech perception in children with autism. *Research in Autism Spectrum Disorders*, *7*(7), 845–857. https://doi.org/http://dx.doi.org/10.1016/j.rasd.2013.03.003

Brothers, L. (1990). The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci*, *1*, 27–51. https://doi.org/citeulike-article-id:2826191

Bull, D., Eilers, R. E., & Oller, D. K. (1984). Infants' discrimination of intensity variation in multisyllabic stimuli. *The Journal of the Acoustical Society of America*, *76*(1), 13–17. https://doi.org/10.1121/1.391110

Bull, D., Eilers, R. E., & Oller, D. K. (1985). Infants' discrimination of final syllable fundamental frequency in multisyllabic stimuli. *The Journal of the Acoustical Society of America*, *77*(1), 289–295. https://doi.org/10.1121/1.392270

Butler, S. C., O'Sullivan, L. P., Shah, B. L., & Berthier, N. E. (2014). Preference for infant-directed speech in preterm infants. *Infant Behavior and Development*, *37*(4), 505–511. https://doi.org/10.1016/J.INFBEH.2014.06.007

Capelli, C. A., Nakagawa, N., & Madden, C. M. (1990). How children understand sarcasm: the role of context and intonation. *Child Development*, *61*(6), 1824–1841.

Carlson, K., & Tyler, J. C. (2018). Accents, not Just prosodic boundaries, influence syntactic attachment. *Language and Speech*, *61*(2), 246–276.

https://doi.org/10.1177/0023830917712282

Chamberlain, B., Kasari, C., & Rotheram-Fuller, E. (2007). Involvement or Isolation? The Social Networks of Children with Autism in Regular Classrooms. *Journal of Autism and Developmental Disorders*, *37*(2), 230–242. https://doi.org/10.1007/s10803-006-0164-4

Chevallier, C., Noveck, I., Happé, F., & Wilson, D. (2009). From acoustics to grammar: Perceiving and interpreting grammatical prosody in adolescents with Asperger Syndrome. *Research in Autism Spectrum Disorders*, *3*(2), 502–516. https://doi.org/10.1016/j.rasd.2008.10.004

Chevallier, C., Noveck, I., Happé, F., & Wilson, D. (2011). What's in a voice? Prosody as a test case for the Theory of Mind account of autism. *Neuropsychologia*, *49*(3), 507–517. https://doi.org/10.1016/j.neuropsychologia.2010.11.042

Clifton, C., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, *45*(Pt 2), 87–114. https://doi.org/10.1177/00238309020450020101

Cohen, D., Cassel, R. S., Saint-Georges, C., Mahdhaoui, A., Laznik, M.-C., Apicella, F., … Chetouani, M. (2013). Do parentese prosody and fathers' involvement in interacting facilitate social interaction in infants who later develop Autism? *PLoS One*, *8*(5), e61402. https://doi.org/http://dx.doi.org/10.1371/journal.pone.0061402

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior & Development*, *20*(4), 477–488. Retrieved from http://www.psyc.vt.edu/sites/default/files/inline_files/Page185/cooper_et_al_1997.pdf

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant directed-speech in the first month after birth. *Child Development*, *61*, 1584–1595. https://doi.org/10.2307/1130766

Cooper, W. E., & Sorensen, J. M. (1977). Fundamental frequency contours at syntactic boundaries. *The Journal of the Acoustical Society of America*, *62*(3), 683–692. https://doi.org/10.1121/1.381556

Couper-Kuhlen, E. (1986). *An introduction to English prosody*. Edward Arnold. Retrieved from https://scholar.google.com/scholar?hl=en&q=Couper-Kuhlen%2C+1986&btnG=&as_sdt=1%2C43&as_sdtp=

Cristia, A. (2013). Input to language: the phonetics and perception of infant-directed speech. *Linguistics and Language Compass*, *7*(3), 157–170. https://doi.org/10.1111/lnc3.12015

Crystal, D. (1986). Prosodic development. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 33–48). Cambridge, UK: Cambridge University Press. Retrieved from http://www.musikochsprak.se/wp-content/uploads/2014/02/Crystal-D-1979-Prosodic-Development.pdf

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America*, *83*(4), 1553–1573. https://doi.org/10.1121/1.395911

Curtin, S. (2009). Twelve-month-olds learn novel word–object pairings differing only in stress pattern. *Journal of Child Language*, *36*(05), 1157. https://doi.org/S0305000909009428 [pii]\r10.1017/S0305000909009428 [doi]

Curtin, S. (2010). Young infants encode lexical stress in newly encountered words. *Journal of Experimental Child Psychology*, *105*(4), 376–385. https://doi.org/10.1016/j.jecp.2009.12.004

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. *Cognitive Models of Speech Processing*, 105–121. Retrieved from https://repository.ubn.ru.nl/bitstream/handle/2066/15711/6107.pdf

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech Und Language*, *2*, 133–175. Retrieved from https://pure.mpg.de/rest/items/item_68762_7/component/file_68763/content

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*(2), 141–201. https://doi.org/10.1177/002383099704000203

Cutler, A., & Swinney, D. A. (1987). Prosody and the development of comprehension. *Journal of Child Language*, *14*(1), 145–167. https://doi.org/10.1017/S0305000900012782

Dahan, D., Tanenhaus, M. K., & Chambersc, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314. https://doi.org/10.1016/S0749-596X(02)00001-3

Diehl, J. J., Bennetto, L., Watson, D. G., Gunlogson, C., & McDonough, J. (2008). Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and Language*, *106*(2), 144–152. https://doi.org/10.1016/j.bandl.2008.04.002

Diehl, J. J., Friedberg, C., Paul, R., & Snedeker, J. (2015). The use of prosody during syntactic processing in children and adolescents with autism spectrum disorders. *Development and Psychopathology*, *27*(3), 867–884. https://doi.org/http://dx.doi.org/10.1017/S0954579414000741

Diehl, J. J., & Paul, R. (2012). Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *6*(1), 123–134. https://doi.org/10.1016/j.rasd.2011.03.012

Dunst, C. J., Gorman, E., Hamby, D. W., & Hamby, D. W. (2012). Preference for infant-directed speech in preverbal young children. *Center for Early Literacy Learning*, *5*(1), 1–13. Retrieved from www.earlyliteracylearning.org

Eilers, R. E., Bull, D. H., Kimbrough, D., & Lewis, D. C. (1984). The discrimination of vowel duration by infants ' for publication. *Journal of Acoustical Society of America*, *75*(4), 1213–1218. Retrieved from http://asadl.org/journals/doc/ASALIB-home/info/terms.j

Erwin, R. J., Van Lancker, D., Guthrie, D., Schwafel, J., Tanguay, P., & Buchwald, J. S. (1991). P3 responses to prosodic stimuli in adult autistic subjects. *Electroencephalography & Clinical Neurophysiology: Evoked Potentials*, *80*(6), 561–571. https://doi.org/http://dx.doi.org/10.1016/0168-5597(91)90139-O

Farris-Trimble, A., & McMurray, B. (2013). Test-retest reliability of eye tracking in the visual world paradigm for the study of real-time spoken word recognition. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1328–1345. https://doi.org/10.1044/1092-4388(2012/12-0145)

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist*, *66*(6), 103–114. https://doi.org/10.1525/aa.1964.66.suppl_3.02a00060

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Development*, *60*(6), 1497–1510. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2612255

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In *Nonverbal vocal communication: Comparative and developmental approaches* (p. 262). Retrieved from https://books.google.com/books?hl=en&lr=&id=lEVQt9gzAFcC&oi=fnd&pg=PA262&dq=meaningful+differences+in+mothers%27+speech+to+infants&ots=4WqVJJ7Yfp&sig=kyjy0PKUa7crnFO6uYGIZTjTDBI

Fernald, A., & McRoberts, G. (1986). Prosodic bootstrapping: A critical analysis of the argument and the evidence. In *Signal to Syntax: Bootstrapping From Speech To Grammar in Early Acquisition*. Retrieved from https://books.google.com/books?hl=en&lr=&id=HWaYAgAAQBAJ&oi=fnd&pg=PA365&dq=prosodic+bootstrapping+hypothesis&ots=YxoBl2FTtd&sig=JBTfbmo97C7RPhIAQLAxWAaM2uw

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113. https://doi.org/10.1037/0012-1649.20.1.104

Fisher, C., & Tokura, H. (1996). Prosody in speech to infants: direct and indirect acounstic cues to syntactic structure. In *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 343–364). Retrieved from https://books.google.com/books?hl=en&lr=&id=HWaYAgAAQBAJ&oi=fnd&pg=PA343&dq=fisher+tokura+1996&ots=Yxpxj0JNxi&sig=cmSbeMEuNcEKThPFpZZBOkAOnyA

Frankel, F., Simmons, J. Q., Richey, V. E., Others, A., Simmons, J. Q., & Richey, V. E. (1987). Reward value of prosodic features of language for autistic, mentally retarded, and normal children. *Journal of Autism and Developmental Disorders*, *17*(1), 103–113. https://doi.org/10.1007/BF01487263

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010a). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, *63*(3), 367–386. https://doi.org/10.1016/j.jml.2010.06.004

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010b). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, *63*, 367–386. https://doi.org/10.1016/j.jml.2010.06.004

Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*(6), 244–249. https://doi.org/10.1016/j.tics.2006.04.002

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, *27*(4), 765–768. https://doi.org/10.1121/1.1908022

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). Is voice a marker for autism spectrum disorder? A systematic review and meta-analysis. *Autism Research*, *10*(3), 384–407. https://doi.org/http://dx.doi.org/10.1002/aur.1678

Gerken, L. A., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition*, *51*(3), 237–265. https://doi.org/10.1016/0010-0277(94)90055-8

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363–376. https://doi.org/10.3758/BF03194885

Gervain, J. (2018). Gateway to language: The perception of prosody at birth. In H. Bartos, M. den Dikken, Z. Bánrénti, & T. Váradi (Eds.), *Boundaries Crossed, at the Interfaces of Morphosyntax, Phnolog, Pragmatics, and Semantics* (pp. 373–384). Cham: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-90710-9_23

Gleitman, L., & Wanner, E. (1982). Language acquisition: the state of the state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3–48). Cambridge, England: Cambridge University Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=Dnc6AAAAIAAJ&oi=fnd&pg=PR7&dq=gleitman+wanner+1982&ots=ceLrtBBHf-&sig=EZUbLhdVkLK_bJ__wwB-OjqFeoc

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2007). The "Reading the Mind in the Voice” test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders*, *37*(6), 1096–1106. https://doi.org/10.1007/s10803-006-0252-5

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, *15*(2), 103–113. https://doi.org/10.1177/002383097201500201

Grossman, R. B., Tager-flusberg, H., Bemis, R. H., Plesa Skwerer, D., Tager-flusberg, H., Skwerer, D. P., & Tager-flusberg, H. (2010). Lexical and affective prosody in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*, *53*(3), 778–793. https://doi.org/10.1044/1092-4388(2009/08-0127)

Grosz, B., & Hirschberg, J. (1992). Some intonational characteristics of discourse analysis. *International Conference on Spoken Language Processing*, 1–9. Retrieved from http://www.iscaaspeech.org/archive

Hallet, P. E. (1986). Eye movement. In K. Buff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*. New York: Wiley.

Hawthorne, K., & Gerken, L. A. (2014). From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition*, *133*(2), 420–428. https://doi.org/10.1016/j.cognition.2014.07.013

Hawthorne, K., Rudat, L., & Gerken, L. A. (2016). Prosody and the acquisition of hierarchical structure in toddlers and adults. *Infancy*, *21*(5), 603–624. https://doi.org/10.1111/infa.12130

Heaton, P. (2005). Interval and contour processing in autism. *Journal of Autism and Developmental Disorders*, *35*(6), 787–793. https://doi.org/10.1007/s10803-005-0024-7

Heaton, P., Pring, L., & Hermelin, B. (1999). A pseudo-savant: A case of exceptional musical splinter skills, *5*, 503–512. https://doi.org/10.1080/135547999908402745

Herman, R. (2000). Phonetic markers of global discourse structures in English. *Journal of Phonetics*, *28*(4), 466–493. https://doi.org/10.1006/jpho.2000.0127

Herold, D. S., Nygaard, L. C., Chicos, K. A., & Namy, L. L. (2011). The developing role of prosody in novel word interpretation. *Journal of Experimental Child Psychology*, *108*(2), 229–241. https://doi.org/10.1016/j.jecp.2010.09.005

Hirschberg, J. (2017). Pragmatics and prosody. *The Oxford Handbook of Pragmatics*, 532–549. https://doi.org/10.1017/cbo9781139022453.031

Hirschberg, J., & Nakatani, C. H. (1996). A prosodic analysis of discourse segments in direction-giving monologues. In *34th Annual Meeting of the Association for Computational Linguistics* (pp. 286–293). https://doi.org/10.3115/981863.981901

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*(3), 269–286. https://doi.org/10.1016/S0010-0277(87)80002-1

Hubbard, D. J., Faso, D. J., Assmann, P. F., & Sasson, N. J. (2017). Production and perception of emotional prosody by adults with autism spectrum disorder. *Autism Research*, *10*(12), 1991–2001. https://doi.org/10.1002/aur.1847

Hudry, K., Leadbitter, K., Temple, K., Slonims, V., McConachie, H., Aldred, C., … Charman, T. (2010). Preschoolers with autism show greater impairment in receptive compared with expressive language abilities. *International Journal of Language and Communication Disorders*, *45*(6), 681–690. https://doi.org/10.3109/13682820903461493

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Ito, K., Bibyk, S. A., Wagner, L., & Speer, S. R. (2013). Interpretation of contrastive pitch accent in six- to eleven-year-old English-speaking children (and adults). *Journal of Child Language*, *41*(1), 84–110. https://doi.org/10.1017/S0305000912000554

Ito, K., Bibyk, S. A., Wagner, L., & Speer, S. R. (2014). Interpretation of contrastive pitch accent in six- to eleven-year-old English-speaking children (and adults). *Journal of Child Language*, *41*(1), 82–108. https://doi.org/10.1017/S0305000912000554

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*(2), 541–573. https://doi.org/10.1016/j.jml.2007.06.013

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Järvinen-Pasley, A., Peppé, S., King-Smith, G., & Heaton, P. (2008). The relationship between form and function level receptive prosodic abilities in autism. *Journal of Autism and Developmental Disorders*, *38*(7), 1328–1340. https://doi.org/10.1007/s10803-007-0520-z

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*, 548–567. https://doi.org/10.1006/jmla.2000.2755

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, *12*(1), 131–141. https://doi.org/10.1111/j.1467-7687.2008.00740.x

Jones, C. R. G., Happé, F., Baird, G., Simonoff, E., Marsden, A. J. S., Tregay, J., … Charman, T. (2009). Auditory discrimination and auditory sensory behaviours in autism spectrum disorders. *Neuropsychologia*, *47*(13), 2850–2858.

https://doi.org/10.1016/j.neuropsychologia.2009.06.015

Jusczyk, P. W. (1992). Developing phonological categories from the speech signal. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonoloigcal development: Models, research, implications* (pp. 17–64). Parkton, MD: York Press.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23. https://doi.org/10.1006/COGP.1995.1010

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*(2), 252–293. https://doi.org/10.1016/0010-0285(92)90009-Q

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*(3–4), 159–207. https://doi.org/10.1006/cogp.1999.0716

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*. https://doi.org/10.1105/tpc.11.5.949

Kargas, N., López, B., Morris, P., & Reddy, V. (2016). Relations among detection of syllable stress, speech abnormalities, and communicative ability in adults with Autism Spectrum Disorders. *Journal of Speech, Language, and Hearing Research*, *59*(2), 206–215. https://doi.org/http://dx.doi.org/10.1044/2015_JSLHR-S-14-0237

Karzon, R. G., & Nicholas, J. G. (1989). Syllabic pitch perception in 2- to 3-month-old infants. *Perception & Psychophysics*, *45*(1), 10–14. https://doi.org/10.3758/BF03208026

Kehoe, M. (2013). Definition and framework of prosody. In *The Development of Prosody and Prosodic Structure* (pp. 3–11). New York: Nova Science Publishers, Inc.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosody facilitation and interference in the resolution of temporary syntactic closure ambiguity. Retrieved from http://www.idealibrary.com

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, *3*(3), 129–140. Retrieved from https://eric.ed.gov/?id=EJ124315

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*, *50*(2), 194–231. https://doi.org/10.1016/j.cogpsych.2004.08.002

Kutik, E. J., Cooper, W. E., & Boyce, S. (1983). Declination of fundamental frequency in speakers' production of parenthetical and main clauses. *The Journal of the Acoustical Society of America*, *73*(5), 1731–1738. https://doi.org/10.1121/1.389397

Lehiste, I. (1970). *Suprasegmentals* (First Edit). Cambridge, MA: The MIT Press.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *The Journal of the Acoustical Society of America*, *53*(1), 380–380. https://doi.org/10.1121/1.1982702

Liberman, M., & Sproat, R. (1992). Modified Noun Phrases in English. In *Lexical Matters* (pp. 131–181). Retrieved from https://books.google.com/books?hl=en&lr=&id=vIy7eEqEqKkC&oi=fnd&pg=PA131&dq=liberman+sproat+1992&ots=iCsR1shLSY&sig=2xOds4uOUTXOcvb_6POzy3Dl2Fc

Lieberman, P. (1967). Intonation, perception, and language. *M.I.T. Research Monograph*, *38*(xii), 210. Retrieved from https://psycnet.apa.org/record/1967-16665-001

Lindner, J. L., & Rosén, L. A. (2006). Decoding of emotion through facial expression, prosody and verbal content in children and adolescents with Asperger's Syndrome. *Journal of Autism and Developmental Disorders*, *36*(6), 769–777. https://doi.org/http://dx.doi.org/10.1007/s10803-006-0105-2

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule: ADOS-2. *Los Angeles, CA: Western Psychological Services*.

Lyons, M., Simmons, E. S., & Paul, R. (2014). Prosodic development in middle childhood and adolescence in high-functioning autism. *Autism Research*, *7*(2), 181–196. https://doi.org/http://dx.doi.org/10.1002/aur.1355

Magiati, I., Tay, X. W., & Howlin, P. (2014). Cognitive, language, social and behavioural outcomes in adults with autism spectrum disorders: A systematic review of longitudinal follow-up studies in adulthood. *Clinical Psychology Review*, *34*(1), 78–86. https://doi.org/10.1016/j.cpr.2013.11.002

Marcus, M., & Hindle, D. (1990). Description theory and intonation boundaries. *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, 483–512. Retrieved from https://psycnet.apa.org/record/1991-97555-018

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders*, *38*(4), 325–350. https://doi.org/http://dx.doi.org/10.1080/1368282031000154204

McCann, Peppé, S., Gibbon, F. E., O'Hare, A., Rutherford, M. D., Fiona, E., … Hare, A. O. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *International Journal of Language & Communication Disorders*, *42*(6), 682–702. Retrieved from http://www.qmu.ac.uk/casl/pubs/McCann et al 2006 WP6 Prosody-Language Relationship.pdf

McMurray, B., Munson, C., & Tomblin, J. B. (2014). Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research, 57*(4), 1344–1362.

https://doi.org/10.1044/2014_JSLHR-L-13-0196

McMurray, B., Samelson, V. M., Lee, S. H., & Bruce Tomblin, J. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178. https://doi.org/10.1016/0010-0277(88)90035-2

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory and Cognition*, *35*(3), 494–503. https://doi.org/10.3758/BF03193289

Morgan, J. L. (1986). *From simple input to complex grammar*. Cambridge, MA: MIT Press.

Morgan, J. L., & Demuth, K. (Eds.). (1996). *Signal to syntax*. Mahwah, NJ: Lawrence Erlbaum Associates.

Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Psychology*, *14*, 477–492.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006, January 2). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*. Springer US. https://doi.org/10.1007/s10803-005-0040-7

Mottron, L., Peretz, I., & Ménard, E. (2000). Local and global processing of music in high-functioning persons with Autism: beyond central coherence? *Journal of Child Psychology and Psychiatry*, *41*(41), 1057–1165. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.489.6067&rep=rep1&type=pdf

Nakatani, C. H., Hirschberg, J., & Grosz, B. J. (1995). Discourse structure in spoken language : Studies on speech corpora. In *Working notes of the AAAI spring symposium on empirical methods in discourse interpretation and generation* (pp. 106–112). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.50.5272&rep=rep1&type=pdf

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756–766. https://doi.org/10.1037/0096-1523.24.3.756

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development*, *21*(4), 779–784. https://doi.org/10.1016/S0163-6383(98)90044-3

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by english-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, *43*(1), 1–19. https://doi.org/10.1006/jmla.2000.2698

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, *33*(1), 127–146. https://doi.org/10.1111/j.1551-6709.2008.01007.x

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2011.11.008

O'Riordan, M., & Passetti, F. (2006). Discrimination in autism within different sensory modalities. *Journal of Autism and Developmental Disorders*, *36*(5), 665–675. https://doi.org/10.1007/s10803-006-0106-1

Ohala, J. (1983). Cross-language use of pitch: An ethological view. *Phonetica*, *40*, 1–18. Retrieved from https://www.karger.com/Article/PDF/261678?casa_token=CAfgomWe-

Q8AAAAA:3oA14mbkQrXPnCUjT6T10xvPfqnXh5lUs5rhapt1UoJ8orprzS9mpkYq9DqGbKFjov5qxRAQbbs

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220. https://doi.org/10.1007/s10803-004-1999-1

Paul, R., Augustyn, A., Volkmar, F., & Cohen, D. (2000). Grammatical and pragmatic prosody perception in high-functioning autism. Paper presented in the Symposium for Resaerch in Child Language Disorder. In *Madison, WI, USA.*

Paul, R., Chawarska, K., Fowler, C., Cicchetti, D., & Volkmar, F. (2007). "Listen my children and you shall hear": Auditory preferences in toddlers with Autism Spectrum Disorders. *Journal of Speech Language and Hearing Research*, *50*(5), 1350. https://doi.org/10.1044/1092-4388(2007/094)

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. https://doi.org/10.1007/s10803-005-0031-8

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*. https://doi.org/10.3389/neuro.11.010.2008

Pepler, D. J., Schroeder, J. H., Weiss, J. A., Cappadocia, M. C., & Bebko, J. M. (2014). Shedding light on a pervasive problem: A review of research on bullying experiences among children with Autism Spectrum Disorders. *Journal of Autism and Developmental*

Disorders, 44(7), 1520–1534. https://doi.org/10.1007/s10803-013-2011-8

Peppé, S. (2009). Why is prosody in speech-language pathology so difficult? *International Journal of Speech-Language Pathology, 11*(4), 258–271. https://doi.org/10.1080/17549500902906339

Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics and Phonetics, 17*(4–5), 345–354. https://doi.org/10.1080/0269920031000079994

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children With high-functioning Autism. *Journal of Speech Language and Hearing Research, 50*(4), 1015–1028. https://doi.org/10.1044/1092-4388(2007/071)

Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. Retrieved from https://dspace.mit.edu/bitstream/handle/1721.1/16065/07492108-MIT.pdf?sequence=2

Pierrehumbert, J. (1984). English phonetics and phonology: a practical course. By Peter Roach. Cambridge: Cambridge University Press, 1983. Pp. x, 212. *RELC Journal*, *15*(1), 117–118. https://doi.org/10.1177/003368828401500113

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. *Intentions in Communication*, 271–311.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: MIT Press.

Ploog, B. O., Banerjee, S., & Brooks, P. J. (2009). Attention to prosody (intonation) and content in children with autism and in typical children using spoken sentences in a computer game. *Research in Autism Spectrum Disorders, 3*(3), 743–758.

https://doi.org/10.1016/j.rasd.2009.02.004

Ploog, B. O., Brooks, P. J., Scharf, A., & Aum, S. (2014). Perception of the prosody and content of sentences in an unfamiliar language in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *8*(7), 775–787. https://doi.org/http://dx.doi.org/10.1016/j.rasd.2014.03.014

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America*, *90*(6), 2956–2970. https://doi.org/10.1121/1.401770

Prutting, C. A. (1982). Pragmatics as social competence. *Journal of Speech and Hearing Disorders*, *47*(May), 123–134. https://doi.org/10.1044/jshd.4702.123

Roid, G. H. (2003). *Stanford-Binet intelligence scale*.

Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: a study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders*, *32*(3), 189–194. https://doi.org/10.1023/A:1015497629971

Samson, F., Mottron, L., Jemel, B., Belin, P., & Ciocca, V. (2006). Can spectro-temporal complexity explain the autistic pattern of performance on auditory tasks? *Journal of Autism and Developmental Disorders*, *36*(1), 65–76. https://doi.org/10.1007/s10803-005-0043-4

Sasson, N. J., Dichter, G. S., & Bodfish, J. W. (2012). Affective responses by adults with autism are reduced to social images but elevated to images related to circumscribed interests. *PLoS ONE*, *7*(8), e42457. https://doi.org/10.1371/journal.pone.0042457

Sasson, N. J., Turner-Brown, L. M., Holtzclaw, T. N., Lam, K. S. L., & Bodfish, J. W. (2008). Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. *Autism Research*, *1*(1), 31–42. https://doi.org/10.1002/aur.4

Schachner, A., & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Developmental Psychology*, *47*(1), 19–25. https://doi.org/10.1037/a0020740

Schafer, A. (1997). *Prosodic parsing: The role of prosody in sentence comprehension*. Retrieved from http://www2.hawaii.edu/~aschafer/Schafer1997.pdf

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *The Journal of the Acoustical Society of America*, *71*(4), 996–1007. https://doi.org/10.1121/1.387581

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, *57*(1), 24–48. https://doi.org/10.1016/j.jml.2006.10.004

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In *The handbook of phonological theory* (1st ed., pp. 550–569).

Semel, W., Wiig, E. H., & Secord, W. A. (2014). *Clinical Evaluation of Language Fundamentals—Metalinguistics*. San Antonio, TX: Pearson Assessments.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 193–247. https://doi.org/10.1007/BF01708572

Shintel, H., Anderson, N. L., & Fenn, K. M. (2014). Talk this way: The effect of prosodically conveyed semantic information on memory for novel words. *Journal of Experimental Psychology: General*, *143*(4), 1437–1442. https://doi.org/10.1037/a0036605

Shriberg, L. D., Paul, R., McSweeney, J. L., Klin, A., Cohen, D., Volkmar, F. R., & McSweeny, J. L. (2001). Speech and Prosody Characteristics of Adolescents and Adults with High-

Functioning Autism and Asperger Syndrome. *Journal of Speech, Language and Hearing Research*, *44*(5), 1097–1115. https://doi.org/10.1044/1092-4388(2001/087)

Simmons, J. Q., & Baltaxe, C. (1975). Language patterns of adolescent autistics. *Journal of Autism and Childhood Schizophrenia*, *5*(4), 333–351. https://doi.org/10.1007/BF01540680

Singh, L., & Harrow, M. S. (2014). Influences of semantic and prosodic cues on word repetition and categorization in autism. *Journal of Speech, Language, and Hearing Research*, *57*(5), 1764–1778. https://doi.org/http://dx.doi.org/10.1044/2014_JSLHR-L-13-0123

Skwerer, D. P., Schofield, C., Verbalis, A., Faja, S., & Tager-Flusberg, H. (2007). Receptive prosody in adolescents and adults with Williams Syndrome. *Language and Cognitive Processes*, *22*(2), 247–271. https://doi.org/10.1080/01690960600632671

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*. https://doi.org/10.1016/S0749-596X(02)00519-3

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities - speer_kjelgaard_dobroth1996.pdf. *Journal of Psycholinguistic Research*, *25*(2). Retrieved from https://link.springer.com/content/pdf/10.1007/BF01708573.pdf

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Stewart, M. E., McAdam, C., Ota, M., Peppé, S., & Cleland, J. (2013). Emotional recognition in Autism Spectrum conditions from voices and faces. *Autism*, *17*(1), 6–14. https://doi.org/http://dx.doi.org/10.1177/1362361311424572

Sun, C. (2006). *Chinese: A linguistic introduction*. *Chinese: A Linguistic Introduction*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511755019

Sutcliffe, P., & Bishop, D. V. M. (2005). Psychophysical design influences frequency discrimination performance in young children. *Journal of Experimental Child Psychology*, *91*(3), 249–270. https://doi.org/10.1016/J.JECP.2005.03.004

Tager-Flusberg, H. (1999). A psychological approach to understanding the social and language impairments in autism. *International Review of Psychiatry (Abingdon, England)*, *11*(4), 325–334. https://doi.org/10.1080/09540269974203

Tanenhaus, M. K., Spivey, M. J., Eberhard, K. M., Sedivy, J. C., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Terken, J., & Nooteboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes*, *2*(3–4), 145–163. https://doi.org/10.1080/01690968708406928

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Van Lancker, D. R., Cornelius, C., & Kreiman, J. (1989). Recognition of emotional-prosodic meanings in speech by autistic, schizophrenic, and normal children. *Developmental Neuropsychology*, *5*(2–3), 207–226.

https://doi.org/http://dx.doi.org/10.1080/87565648909540433

Van Roekel, E., Scholte, R. H. J., & Didden, R. (2010). Bullying among adolescents with autism spectrum disorders: Prevalence and perception. *Journal of Autism and Developmental Disorders*, *40*(1), 63–73. https://doi.org/10.1007/s10803-009-0832-2

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7), 905–945. https://doi.org/10.1080/01690961003589492

Wang, A. T., Dapretto, M., Hariri, A., Sigman, M. D., & Bookheimer, S. Y. (2001). Processing affective and linguistic prosody in autism: an fMRI study. *NeuroImage*, *13*(6), S621.

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: the role of prosody and context. *Brain*, *129*(4), 932–943. https://doi.org/http://dx.doi.org/10.1093/brain/awl032

Watson, D. G. (2010). The Many Road to Prominence: Understanding Emphasis in Conversation. In *Psychology of Learning and Motivation* (Vol. 52, pp. 163–183). https://doi.org/10.1016/S0079-7421(10)52004-8

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2006). Acoustic Prominence and Reference Accessibility in Language Production. In *Speech Prosody*. Retrieved from http://www.isca-speech.org/archive

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, *106*(3), 1548–1557. https://doi.org/10.1016/j.cognition.2007.06.009

Watson, D. G., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, *19*(6), 713–

755. https://doi.org/10.1080/01690960444000070

Watson, D. G., Gunlogson, C. a, & Tanenhaus, M. K. (2006). Online methods for the investigation of prosody. *Methods in Empirical Prosody Research*, 259–282. Retrieved from http://s3.amazonaws.com/academia.edu.documents/41751264/Online_methods_for_the_investigation_of_20160129-29089-129acbj.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1500333657&Signature=94b3c2FVrUcgnBDkAjMVvDgwAE4%253D&response-content-disposition=inlin

Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science*, *32*(7), 1232–1244. https://doi.org/10.1080/03640210802138755

Watson, L. R., Baranek, G. T., Roberts, J. E., David, F. J., & Perryman, T. Y. (2010). Behavioral and physiological responses to child-directed speech as predictors of communication outcomes in children With Autism Spectrum Disorders. *Journal of Speech, Language, and Hearing Research*, *53*(4), 1052–1064. https://doi.org/10.1044/1092-4388(2009/09-0096)

Watson, L. R., Roberts, J. E., Baranek, G. T., Mandulak, K. C., & Dalton, J. C. (2012). Behavioral and physiological responses to child-directed speech of children with autism spectrum disorders or typical development. *Journal of Autism and Developmental Disorders*, *42*(8), 1616–1629. https://doi.org/10.1007/s10803-011-1401-z

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197–234. https://doi.org/10.1207/s15473341lld0102_4

Wiig, Semel, W., & Secord, W. A. (2013). *Clinical Evaluations of Language Fundamentals – Fifth Edition (CELF-5)*. Bloomington, MN: NCS Pearson.

Xu, Y. (2015). Speech prosody: Theories, models, and analysis. In A. R. Meireles (Ed.), *Courses on speech prosody* (pp. 146–177). Newcastle upon Tyne, UK: Cambridge Scholars Publishing. Retrieved from http://www.homepages.ucl.ac.uk/~uclyyix/yispapers/Xu_Theories_Models_authorV.pdf

Yule, G. (1980). Speakers' topics and major paratones. *Lingua*, *52*(1–2), 33–47. https://doi.org/10.1016/0024-3841(80)90016-9