11-30-2017

The Influence of Talker Expectations and Acoustic Variability on Speech Perception in ASD

Anders Hogstrom
anders.hogstrom@uconn.edu

The Influence of Talker Expectations and Acoustic Variability on Speech Perception in ASD

Anders Hogstrom

B.A., University of Chicago, 2014

A Thesis
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Sciences
at the
University of Connecticut
2017

APPROVAL PAGE

Master of Sciences Thesis

The Influence of Talker Expectations and Acoustic Variability on Speech Perception in ASD

Presented by

Anders Hogstrom, B.A.

Major Advisor _____

Inge-Marie Eigsti, Ph.D.

Associate Advisor _____

Deborah Fein, Ph.D.

Associate Advisor _____

James Magnuson, Ph.D.

University of Connecticut

2017

Acknowledgements

Many thanks to Inge-Marie Eigsti for her helpful insights throughout this project. Her mentorship throughout grad school made this project possible. Not only were her advice in planning and her comments during the preparation of this document invaluable, but her teaching has made me a better scientist, writer, and clinician.

Thanks to Deborah Fein and James Magnuson for volunteering their time and expertise to review this project. Their dedication to the pursuit of knowledge and the knowledge that they have generated throughout their careers have served as an inspiration for this and many other projects I have undertaken.

Thanks to the other members of my lab, Brian Castelluccio, Allison Canfield, Joshua Green, and Karla Rivera Figueroa, for their intellectual and emotional support. Thanks also to Ryan Morris, Tanya Rao, Meagan Robitaille, Paras Shah, and Morgan Smith for their help in collecting and organizing the data. The broad characterization of participants which was vital to this project would not have been possible without the tireless work of everyone in the lab.

Abstract

Speech perception is dependent upon the ability to map the *sensory* features of a speech signal onto the *perceptual* features which make up language (i.e., phonemes). A great deal of research over the past six decades has focused on how variability across talkers influences speech processing. Listeners are required to normalize the acoustic variability across talkers by continuously updating the mapping from the acoustic signal to phonetic representations. As such, processing speech from multiple talkers is cognitively more demanding than listening to a single talker. This processing cost appears to reflect, in part, the influence of listeners' expectations that speech is coming from multiple sources (talkers). It remains unclear whether talker normalization effects are present in individuals with autism spectrum disorder (ASD), given broad differences in social and sensory processing. The present study examined talker normalization and effects of talker expectation in adolescents with ASD and typical development. Participants were asked to respond to target words embedded in a stream of speech; the pitch of the talkers (F0) varied in half of trials. Furthermore, half of participants were told that this variability was due to fluctuations in a single talker's speech, while the other half were told that the speech was variable because it was produced by two talkers. Results indicated that participants with ASD were significantly slower to respond under conditions of acoustic variability, while typically developing participants were not. Furthermore, the degree to which participants with ASD were influenced by the variability was significantly correlated with parent-reported sensory atypicality. This relationship was not moderated by ASD symptom severity. Neither diagnostic group was influenced by the manipulation of expectations. Overall, these results suggest that sensory differences present in ASD may account in part for communication difficulties.

Introduction

Processing the Speech Signal

In order to understand speech, a listener must map *sensory* aspects of the speech signal onto the *perceptual* categories, or phonemes, which make up their language. In order to distinguish between the words *rid* and *lid*, a listener must use the appropriate acoustic cues to distinguish between the /r/ and /l/ sounds. The ease with which we can comprehend speech might make this appear to trivially easy. In fact, the challenge of operationalizing this problem is so difficult that the solution has not been found in seventy years of psycholinguistic research. Disruptions at any level of processing can have compounding effects. The focus of this thesis will be to examine how changes in sensory and information processing in autism spectrum disorder (ASD) may lead to speech processing difficulties, especially under conditions of acoustically variable speech.

While fluent speakers are able to map acoustic features of speech onto the perceptual categories of language (*phonemes)*, this mapping is not static. There exists a many-to-many relationship between sounds and phonemic categories. A given sound might be perceived as an /a/ in one phonological context, but be perceived as /ae/ in a different phonological context. Similarly, different acoustic signals may be perceived as the same phoneme, given specific contexts. For example, the formants associated with /d/ in /di/ and /du/ differ significantly, yet are perceived as the same phoneme.

This variable relationship between sounds and perceptual categories is known as the *lack of invariance* problem, and derives in part from intra-individual variability in speaking. A given talker's articulation, and the consequent acoustic signal, varies across a number of dimensions: speaking rate (Miller & Baer, 1983), co-articulation of phonemes, in which the articulation of

one sound is influenced by a neighboring sound (Delattre, Liberman, & Cooper, 1955), affective state, etc. Inter-individual differences further complicate the mapping process. Characteristics such as gender, physical size, and accent all lead to significant variability in acoustic properties of speech (Dorman, Studdert-Kennedy, & Raphael, 1977). For example, the word *car* sounds very different when spoken by a man from Boston and a woman from California. Despite this many-to-many relationship, listeners are able to process and interpret speech rapidly. *How* listeners process this lack of invariance between the speech signal and phoneme categories has been the subject of inquiry for over six decades (Peterson & Barney, 1952).

A number of theories have been proposed to account for our ability to solve the problem of invariance. Simple, passive models of speech perception have proven clinically useful and are still often discussed. For example, Geschwind (1970) proposed that acoustic signals are processed in Wernicke's area in the superior temporal lobe to unpack linguistic meaning, then transmitted anteriorly to Broca's area in the inferior frontal gyrus to articulate a response. This model has been useful describe and predict aphasia and other clinical language deficits (Anderson, Gilmore, Roper, Crosson, & Bauer, 1999). However, such models suggest a simple, deterministic relationship between acoustic stimuli and phoneme representation, with only unidirectional information processing from Broca's to Wernicke's areas. If mapping acoustic features onto perceptual categories occurs unidirectionally, it is difficult to account for the perceptual constancy (that is, the many-to-many mapping of acoustic signals onto phonemes) experienced by listeners. In order to accomplish this feat without top-down feedback to sensory areas, one would have to rely upon a potentially large number of pattern representations to account for the lack of invariance described above (Heald & Nusbaum, 2014).

Not only is such a model not parsimonious, but it does not account for a variety of behavioral and neuroimaging data which suggest that acoustic-phonemic mapping is influenced by top-down regulation from frontal regions. The McGurk effect illustrates this kind of top-down feedback. The McGurk effect refers to the illusion created when incongruous video and audio of an individual speaking different phonemes are combined; listeners experience an auditory percept which is distinct from either of the phonemes presented in stimuli themselves. For example, if a /ba/ sound is played with video of a talker producing the syllable /ga/, individuals frequently report experiencing this multimodal input as /da/, a percept which does not correspond to either individual modality of input. This illusion suggests that auditory perception of the speech signal is influenced by the combination of multiple sensory modalities; this pattern necessarily implies that speech perception is not merely a result of bottom-up speech processing, but is the result of substantial top-down regulation. This theory is consistent with neurobiological research on the McGurk effect, which demonstrates modulation of auditory processing regions by areas which are involved in cross-modal integration, such as the inferior frontal gyrus or premotor regions (Jones & Callan, 2003), though more recent evidence suggests that this integration may occur in the superior temporal sulcus (Matchin, Groulx, & Hickok, 2014).

Furthermore, the developmental trajectory of language acquisition demonstrates the importance of top-down organization of the acoustic signal. Early in development, infants reliably differentiate between all phonemes. However, by age 12 months, as they gain more experience with their native language, they begin to ignore acoustic variability which does not meaningfully relate to phoneme contrasts (Kuhl, Williams, Lacerda, Stevens, & Jindblom, 1992). In other words, during the initial months of life, all babies will respond to the distinction between *rid* and *lid*. While English-exposed children will continue to respond to this distinction, which is

meaningful in their language, children exposed to Japanese, a language with no meaningful phonemic distinction between /r/ and /l/ sounds, will eventually stop differentiating between *rid* and *lid.*

What is the nature of this higher-order processing of speech? One model of top-down feedback in speech processing (the motor theory of speech perception) contends that speech is decoded by translating acoustic signals to the motor sequences required to produce a given sound. Within a listener, the relationship of motor sequences and acoustic signals is built over the course of his experiences as a talker—that is, as an individual produces given sounds, the relationship between the motor movement required to produce a sound, and the acoustic signal itself, is strengthened. By relying on this relationship, listeners could rely on their physical actions as talkers to decode the complex speech signal (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985; Viswanathan, Magnuson, & Fowler, 2010).

In contrast to simple feed-forward models of speech processing, this motor feedback framework emphasizes the importance of top-down feedback from frontal to posterior regions of the brain during speech processing. Such a system is not uncommon amongst complex cognitive tasks, as the recruitment of the motor system has been implicated in a number of cognitive processes (Galantucci, Fowler, & Turvey, 2006). Some have also suggested that the motor system is a part of the so-called "mirror neuron" system, which activates when individuals merely view others performing a motor task. Iacoboni & Dapretto (2006) suggest that this motor processing underlies a number of important social processes with connections to visual and tactile stimuli on and near the body, as well as auditory stimuli.

Some have suggested that even these approaches under-estimate the amount of active processing needed to map the acoustic signal accurately onto categorical representations. Several studies have demonstrated top-down effects of expectation, knowledge, or attention on the processing of auditory information. For example, Galbraith & Arroyo (1993) demonstrated that selective attention to one ear influenced processing in the auditory brainstem, a subcortical brain area early in the auditory processing pathway which reliably responds to stimulation from the cochlea. The fact that attentional manipulations influence these early responses to acoustic stimulation necessitate cortical modulation of early sensory/perceptual processing. Furthermore, effects of attention have also been demonstrated to modulate responses in the cochlea. Researchers have found that evoked otoacoustic emissions (signals emitted by the cochlea in response to acoustic stimulation) differ depending on whether listeners are instructed to attend to one ear or another. The fact that a selective attention manipulation elicited differences in evoked otoacoustic emissions suggests that cortical processing influences the auditory processing stream as *early* as the cochlea (Giard, Collet, Bouchet, & Pernier, 1994).

Other research has demonstrated that explicit knowledge or expectations about talkers can influence the processing of acoustic properties of multiple talkers (this processing is termed "talker normalization"). Magnuson & Nusbaum (2007) demonstrated that a priori expectations about talkers influence speech perception. Participants in their task were significantly slower to react to speech when it was acoustically variable, as if it had been produced by two talkers, consistent with prior research. Participants also listened to blocks of speech with more subtle acoustic variability, with only a 10Hz difference in pitch across words. In this condition, there were no significant effects on reaction time. However, when participants had an explicit expectation that this subtle variability was meaningful, i.e., that the variability reflected the

presence of multiple talkers, they were once again significantly slower compared to acoustically homogenous speech. The identical variability had differential impact on processing (RT), depending on high-level expectations about the number of talkers.

Nusbaum and Magnuson (2007) suggested that this pattern of results indicated that listeners utilize active, knowledge-mediated mechanisms to quickly adjust to changes in acoustic-phonetic mappings. To the extent that top-down active control mechanisms are influential in speech processing, individual differences should lead to differential effects of these expectations. Furthermore, do differences in sensory perception broadly influence bottom-up processing mechanisms that influence acoustic-phonetic mapping? While cognitive mechanisms support this mapping process, differences in the quality of input may also influence one's ability to decode the speech signal. In order to answer these questions, the present study turns to autism spectrum disorder (ASD). Among the core symptoms of ASD are deficits in socio-communicative processing that might influence sensitivity to inter-individual differences in input, as well as sensory differences that might influence bottom-up processing of variable input. The following section discusses the nature of communicative differences in ASD, and how sensory processing differences may interactive with or lead to these challenges.

Language Deficits in ASD

Autism spectrum disorder (ASD) is a neurodevelopmental disorder defined by impairments in social interaction and communication and the presence of stereotyped behaviors or interests and atypical sensory responses (APA, 2013). ASD is often accompanied by delays in early receptive and expressive language acquisition (Gamliel, Yirmiya, Jaffe, Manor, & Sigman, 2009). These gaps are most apparent in childhood, and children with ASD often make language gains over the course of development; despite these gains, language deficits in ASD can persist

throughout the lifespan, even in individuals who acquire fluent language (Eigsti & Bennetto, 2009; Howlin, Goode, Hutton, & Rutter, 2004).

Language pragmatics are a life-long challenge for essentially all individuals with ASD. "Pragmatics" refers to *how* language is used to achieve social communication and encompasses domains such as negotiating turn-taking, register (i.e., altering speech as a result of social context or interlocutor), and integrating speech with eye contact, body language, and facial expression (Eigsti, De Marchena, Schuh, & Kelley, 2011). Individuals with ASD often fail to respond fully to questions or social comments (Capps, Kehres, & Sigman, 1998), and may miss implied requests or humor which depend on understanding subtle aspects of communication (Ozonoff & Miller, 1996).

Why do pragmatics remain a challenge for adults with ASD despite the fact that they have mild to minimal impairments in other language domains? Some researchers have proposed cognitive and behavioral accounts for these deficits. For example, Baron-Cohen, Leslie, & Frith, (1985) suggested that Theory of Mind deficits could account for the range of ASD symptoms, including pragmatic deficits. For example, it may difficult to understand sarcasm or humor if you have difficulty understanding the intentions of another person and may instead interpret comments literally.

Others have suggested that social motivation underlies ASD symptoms (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012). The Social Motivation theory suggests that individuals with ASD do not receive pleasurable, rewarding feedback from social interactions, with others, which can account for deficits in social orientation or maintaining social interactions. If individuals with ASD have disturbed motivation to participate in social interactions, they may withdraw from conversations or answer questions inappropriately. This

withdrawal may further harm their skills by eliminating opportunities to practice pragmatic skills for situations in which they are motivated to interact. Furthermore, it may impede their ability to efficiently process speech from multiple talkers. Individuals with ASD may ignore or fail to utilize social information which can assist them in decoding speech. Furthermore, having fewer social interactions may allow them fewer opportunities to practice engaging this system.

Domain-general cognitive differences may ~~also~~ contribute to delays in language acquisition in ASD. Frith & Happe (1994) suggested that "weak central coherence" could play a role in communication deficits. They suggest that a fundamental cognitive deficit in ASD is the extraction of a gestalt, which leads to a greater focus on sensory details. For example, typically developing individuals often find it difficult to mimic perceptually coherent block designs, because they find it difficult to isolate pieces from the coherent whole; in contrast, individuals with ASD show a greatly reduced effect of perceptual coherence when completing these tasks (Caron, Mottron, Berthiaume, & Dawson, 2006). Weak central coherence also accounts for findings that individuals with ASD show poor reading of homographs (i.e., two separate words that share a spelling); accurate decoding of homographs requires the integration of information across a sentence (Happe, 1997). This effect may be due to poor top-down feedback from regions responsible for integrating information into a perceptually coherent whole back to sensory processing regions.Russo et al. (2008) discuss the possibility that weaker top-down feedback may also be responsible for differences in auditory brainstem response to pitch contour in ASD.

These data may also reflect differences in bottom-up sensory processing, as proposed in Enhanced Perceptual Processing accounts of ASD (Mottron, Dawson, Souliéres, Hubert, & Burack, 2006). The proposal that sensory differences in ASD are due to poor top-down

regulation is also consistent with biological evidence of weakened long-distance functional connectivity with the frontal lobes in ASD (Courchesne et al., 2007). Deviations in top-down regulation of sensory systems may also underlie some of the clinical symptoms of ASD. Individuals with ASD frequently report clinical differences in sensory sensitivity or reactivity, though there is broad heterogeneity. Some individuals with ASD are unresponsive to stimuli in their environment, for example someone calling their name. On the other hand, many children with ASD are easily distracted by sensory stimuli and may engage in self-stimulatory behaviors ("stimming").

These behavioral differences may not merely reflect social or attentional differences in ASD, but may be fundamentally tied to differences in perceptual experiences. Within the auditory domain, numerous studies have demonstrated greater sensitivity to small differences in pitch. Bonnel et al. (2003) demonstrated heightened processing of pitch in individuals with ASD compared to typically developing controls on both pitch discrimination (same/different) and categorization (high/low) tasks. Children with ASD also perform superior to TD controls when asked to learn associations between absolute pitches and pictures of animals, suggesting that these differences in audition may extend to the long-term encoding of pitch information (Heaton, Hermelin, & Pring, 2016).

While *superior* auditory performance may superficially appear at odds with *delayed* language acquisition in this group, it may be the precursor of these deficits. Eigsti & Fein (2013) demonstrated a negative relationship between pitch discrimination abilities and retrospective reports of early language milestone acquisition in a sample of adolescents with ASD; this relationship was not present in an age-matched, typically developing sample. This relationship may suggest that enhanced pitch discrimination abilities *impair* rather than assist speech

processing. The authors posited that enhancements in pitch discrimination skills make it difficult for individuals with ASD to develop categories because they attend more than their peers to small differences across utterances or talkers, even when those differences are not meaningful in the language. Differences in the top-down regulation of auditory processing may enhance the perceptual distinctions between talkers, making it more difficult to develop an acoustic-phonemic map which is tuned to characteristics of the native language.

The extent to which auditory discrimination abilities continue to influence speech perception in ASD into adolescence and adulthood is unclear. Once individuals have achieved fluency (and therefore have relatively stable internal representations of phonetic categories), to what extent do perceptual differences continue to play a role in language deficits in ASD? While adolescents and adults with ASD may be able to communicate effectively and understand speech, there may still exist subtle deficits in the speed or accuracy of speech perception. These deficits may be especially salient in the context of high talker variability (e.g., in conversation with multiple interlocutors).

In typically developing individuals, talker normalization effects are not merely the result of sensory variability. Magnuson & Nusbaum (2007) demonstrated that typically developing individuals exhibited slower reactions in conditions with acoustically variable speech, but the effect of talker variability was moderated by the *expectation* of multiple talkers. As noted above, when participants believed that they would hear *one* computer-synthesized talker whose voice went up and down, there was a lesser effect of acoustic variability. These results demonstrate that minimal acoustic variability do not necessarily elicit a cognitive cost. However, when participants believe that variability is socially meaningful (i.e., that the variability signifies *two* talkers), this same acoustic variability does produce a processing cost. Such an effect

demonstrates that typically developing listeners use active, top-down strategies in order to make sense of acoustic variability in the speech signal.

The present study sought to answer two questions. Consistent with previous literature, we hypothesized that individuals with TD would be slower to respond under conditions of small acoustic variability, but only when they were instructed to expect multiple talkers. In contrast, given ASD-associated differences in auditory processing, we predicted subtle deficits in processing of acoustically variable speech, reflected in slower reaction times. Talker variability was hypothesized to differentially affect speech processing in individuals with ASD. A second primary aim probed whether individuals with ASD would integrate expectations about talkers to moderate the processing of acoustically variable speech. Individuals with ASD fail to appropriately use social information and may have difficulty with top-down regulation of sensory experiences and behaviors. We therefore hypothesized that individuals with ASD show a reduced influence of expectations about talkers on speech processing, and that the degree of difference would correlate with ASD severity. A third exploratory aim was to understand how individual differences in sensory processing may lead influence talker normalization effects.

Methods

Participants

Sixteen adolescents with ASD and 15 adolescents with TD participated in this study. Participants did not differ on chronological age, gender, or verbal IQ as measured by the Stanford-Binet Intelligence Scales, Fifth Edition(Roid, 2003); see Table 1. Participants were recruited through fliers in the community, participation in previous studies, and word of mouth. For inclusion, participants were required to have a full-scale IQ above 85 and be native English speakers. Participants with a history of significant neurological impairment (including seizures

and concussions) or any hearing problems were excluded. Participants with comorbid learning or psychiatric disorders were not excluded from participation, in order to reflect typical demographic variability. Participants were excluded from the TD group if they had any first-degree relatives with ASD. One participant from the ASD group was excluded due to failure to complete all necessary tasks. The final sample therefore includes 15 participants in the ASD group and 15 participants in the TD group. Demographic information and scores for each group are included in Table 1.

Table 1. Demographic Information for ASD and TD groups.

	ASD *Mean (SD)* *Range*	**TD** *Mean (SD)* *Range*	χ^2 **or** F	p
N (M:F)	15 (11:4)	15 (10:5)	.16	.69
Chronological Age (Years)	15.6 (2.0) 12.9-18.8	14.6 (1.8) 12.2-17.8	1.96	.17
Stanford-Binet				
Non-Verbal	11.7 (2.1) 8-15	10.9 (2.1) 8-15	.92	.35
Verbal*	10.1 (2.5) 5-12	12.6 (2.0) 9-16	9.33	.005
Total	105.2 (10.3) 85-118	110.6 (8.6) 97-121	2.4	.13
SCQ (Total Score)*	17.6 (8.1) 4-26	2.1 (1.6) 0-5	46.2	<.001
Sensory Profile*	143 (30) 67-187	166 (23) 123-186	5.56	0.03
BRIEF*	68.8 (11.0) 44-85	55.5 (10.9) 40-72	19.5	<.001
ADOS				
Communication	7.1 (2.7) 2-12			
Social Reciprocity	2.6 (2.6) 0-9			
Communication + Social Reciprocity	9.7 (1.9) 7-13			

Note: ASD, Autism Spectrum Disorder; TD, typically developing; SCQ, Social Communication Questionnaire; BRIEF, Behavior Rating Inventory of Executive Functioning; ADOS, Autism Diagnostic Observation Scale.

Diagnoses for the ASD group were verified by trained clinicians using the Autism Diagnostic Observation Scales, 2nd edition (ADOS; Lord et al., 2012) and Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003). The ADOS is a semi-structured assessment measure used to diagnose ASD. The ADOS was administered only to members of the ASD group in order to confirm the diagnosis of ASD. Depending on each participant's maturity, Module 3 or Module 4 was administered. The Social Communication Questionnaire is a 40-item parent-report measure designed as an autism screening tool. All participants' parents completed the Lifetime version, which probes whether autism-related symptoms have ever been present for a child. Data from 27 participants is included. Two parents in the TD group did not return the measure; one parent in the ASD group had many ambiguous responses which could not be scored.

All participants in the ASD group scored above the ADOS cutoff score of 7 for autism spectrum; eleven participants in the ASD group scored above the ADOS cutoff score of 9 for autism. Furthermore, the ASD group scored significantly higher on the SCQ, indicating greater impairment (see Table 1). While four participants scored below the SCQ cutoff of 15, we judged them to have ASD, given their reported history of an ASD diagnosis and expert clinical judgement on the ADOS.

Informed written consent was obtained from parents and participants prior to testing. This research was approved by the University of Connecticut Institutional Review Board.

Measures

Stanford-Binet Intelligence Scales, Fifth Edition (Roid, 2003). The Stanford-Binet Intelligence scales is a measure of cognitive ability. Participants completed two subtests of the Stanford-Binet: Matrices and Vocabulary, which together provide a reliable estimate of full-scale IQ (FSIQ). Performance on these subtests was used to calculate non-verbal (NVIQ) and verbal intelligence quotient (VIQ) respectively.

Short Sensory Profile. (McIntosh, Miller, Shyu, & Dunn, 1999). The Short Sensory Profile is a 38-item parent-report measure which examines clinically relevant, sensory-related difficulties. Items describe sensory seeking/avoiding behaviors across all sensory modalities, including items such as "Is distracted or has trouble functioning if there is a lot of noise around" or "Will only eat certain tastes," which parents rate on a five-point scale from Always to Never. Higher scores on the Short Sensory Profile represent more typical sensory experiences, while lower scores suggest more likely differences in sensory experiences.

BRIEF. The Behavior Rating Inventory for Executive Function (BRIEF) is an 86-item, parent report questionnaire (Gioia, Guy, Isquith, & Kenworthy, 1996). The BRIEF provides a Behavioral Regulation Index, which represents a child's ability to modulate emotions and behavior appropriately, and a Metacognition Index, which represents a child's ability to organize and plan for the future. The BRIEF also provides a Global Executive Composite which summarizes a child's executive functioning across both indices.

Hearing Screen. Intact hearing was confirmed using a GS1-61 audiometer (Grason-Stadler, Inc.), which presented tones at 20dB at 500, 1000, 2000, 4000, and 8000Hz to each ear. Hearing testing was performed in a standard laboratory room, rather than a soundproof booth. Five participants (4 TD) failed screening at 500Hz in one or both ears. One participant in the

ASD group failed testing at 500 and 1000Hz in the right ear. One participant in the TD group failed testing at 500Hz in the left ear and 8000 Hz in the right ear. These failures may have been due to environmental noise (e.g., air conditioner) rather than hearing deficits. None of the participants reported a history of hearing deficits, and none experienced difficulty in comprehension of task instructions or procedures. Therefore no participants were excluded on the basis of the hearing screener.

Procedure

Participants completed testing in a quiet room at the University of Connecticut. The measures included in this study were part of a larger study of communication in ASD. Testing was completed over approximately five hours across one or two sessions.

Stimuli. The stimuli in the present study were provided by James Magnuson and are a subset of stimuli described by Magnuson & Nusbaum (2007). Stimuli consisted of nineteen monosyllabic words: *ball*, *cave*, *done*, and *tile* (the targets), and *bluff*, *cad*, *cling*, *depth*, *dime*, *gnash*, *greet*, *jaw*, *jolt*, *knife*, *lash*, *park*, *priest*, *romp*, and *reek* (the distractors). Two synthetic talkers (e.g., computerized voices) produced one token of each word. Talkers were developed from standard parameters of the DECtalk synthesizer. The two talkers were identical except for one feature: Talker A had an average F0 of 150Hz, while Talker B had an average F0 of 160Hz (that is, Talker B had a slightly higher-pitched voice).

Procedure. Participants completed a timed target-monitoring task, as shown in Figure 1. Subjects were presented with a written form of one of four possible target words and were instructed to press the space bar as quickly as they could whenever they heard the word printed on the screen. Participants then heard a continuous stream of 16 words played at a rate of one

every 583ms; the target word appeared in a random location in the series four times in each run of 16 items. Filler items were chosen randomly from the set of fifteen distractor words.

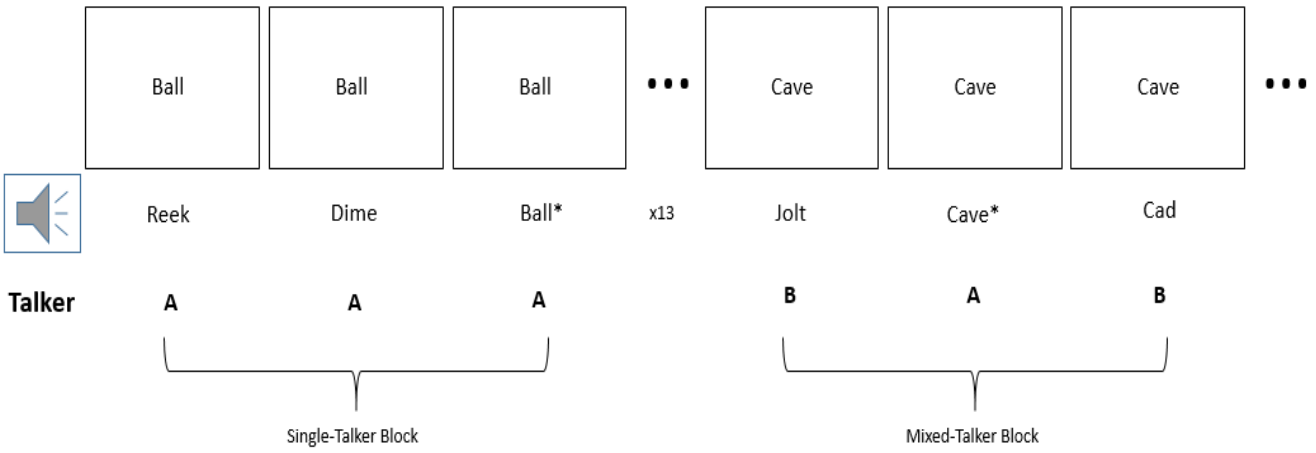


Figure 1. Visual representation of experimental procedure. Participants monitored a continuous stream of speech for a target word, which was displayed on a computer screen. After 16 trials, the target word switched. These 16-trial blocks also varied on the number of talkers. While monitoring for some target words, participants heard only one talker, while in other blocks they heard two talkers. There were a total of 16 blocks (8 single-talker, 8 mixed-talker), for a total of 256 trials. While changes in talker blocks corresponded to changes in target words, changes from single- to mixed-talker blocks were not explicitly cued.

These 16-trial runs were divided into two conditions. In the blocked-talker condition, participants heard the 16 target and distractor words spoken by only one of the two synthetized talkers. In the mixed-talker condition, equal numbers of targets and distractors were produced by each of the two talkers; the order of talkers was randomized. The order of these blocks was random, and blocks flowed seamlessly together, such that participants had no explicit cues to changes in the number of talkers. Participants completed 256 total trials: 128 trials in blocked-talker condition and 128 trials in the mixed-talker condition. Each of the four target words served

as the target on thirty-two total trials: sixteen in the blocked-talker condition and sixteen in the mixed-talker condition.

Additionally, we manipulated participants' expectations of each talker by slightly altering task instructions. Participants were assigned randomly to one of two between-subjects conditions. In the Single Talker expectation group, participants (*n*=10 with ASD, 9 with TD) were told that they would hear only one synthetic talker whose voice went up and down in order to sound "more natural." Participants then heard a 36-second monologue with pitch variation between 150 and 160Hz. In the Two Talkers expectation group, participants (*n*=5 with ASD, 7 with TD) were told that we had produced two synthetic talkers by changing the pitch. Participants then listened to a 40-second dialogue between the two voices. The text of these instructions is included in Appendix A.

Data Analysis

Response time and accuracy were used to measure performance. Response times less than 250ms after stimulus onset were treated as responses to the previous item (i.e., 583ms+RT applied to the previous trial). Approximately 0.1% of trials (*n*=79) met this criteria. Assuming a moderate effect size (f=0.25) and moderate correlation between repeated measures (r=.50), a sample of 72 would be necessary to detect a three-way interaction (list the factors in interaction) with sufficient power (1-β=.80). As the current study was underpowered with respect to this three-way interaction (sample size was determined based on other components of a larger study), analyses utilized two ANCOVAs to assess the interaction of expectation conditions and acoustic variability for each diagnostic group independently. NVIQ was included as a covariate. Post-hoc correlations were planned to assess the relationship between sensory atypicalities and the effect of acoustic variability.

Results

Initial Data Examination. Data were first examined to ensure that they met necessary assumptions of statistical analyses. Accuracy of responses was measured using d', a measure of response sensitivity which balances correct hits and false alarms. The d' scores were approximately normal with a slight rightward skew in the ASD group, as seen in Figure 2a. Visual inspection of the data revealed relatively high accuracy and small variability in both groups; all participants appeared to be performing significantly above chance. Figure 2b shows the distribution of reaction time (RT) across all conditions for each group. A Q-Q Plot for RT residuals showed that data were relatively normal; when broken down by diagnostic group, the distribution of data did not significantly differ from a normal distribution (ASD: Kolmogorov-Smirnov=0.19, p=0.13; TD: Kolmogorov-Smirnov=0.18, p=0.20).

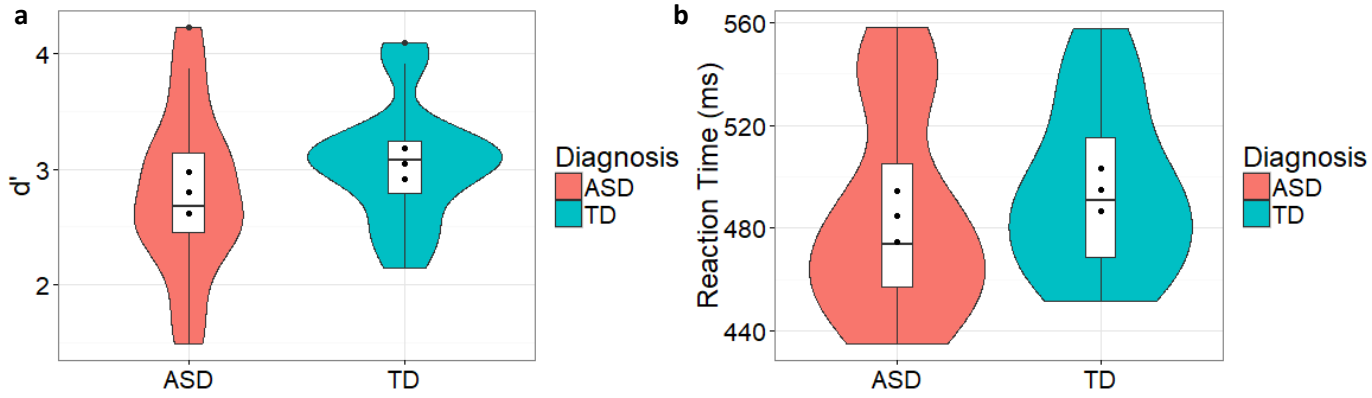


Figure 2: Figure 2a (left) shows the distribution of d' scores for the ASD and TD groups; figure 2b (right) shows the distribution of reaction time for each group. Dots within the box plot indicate the mean ± 1 standard deviation.

As indicated above, the following analyses utilized RT as the primary dependent variable. First, however, it was important to ascertain whether significant differences in RT were due to overall group differences in the task. For instance, if the ASD group performed less accurately overall, any effects of RT may be the result of overall speech processing deficits rather than

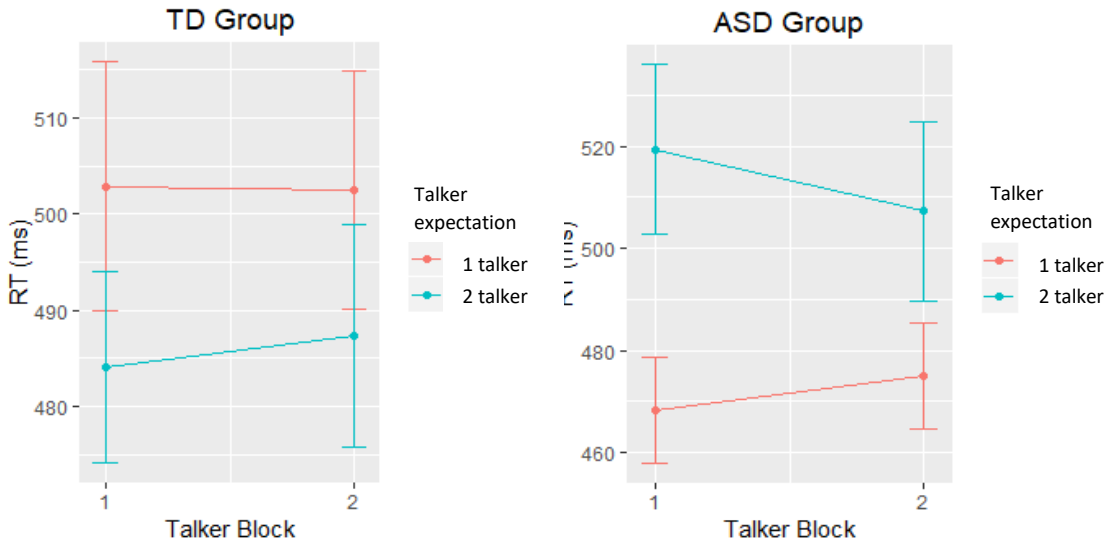
specific effects of the task manipulation. To test this possibility, an *a priori* t-test was conducted to compare accuracy scores between the TD and ASD groups. Results indicated that d' scores did not differ for the two groups, $t(29)=1.01$, $p=0.28$.

Expectation by Variability within Diagnosis. Given the limited sample size and low predicted power, two separate repeated-measures linear models were used to assess the two-way interaction of expectation and talkers within each diagnostic group. These analyses included age and non-verbal IQ as covariates. Non-verbal IQ was included rather than full-scale or verbal IQ because utilizing these measures as covariates may eliminate meaningful information which varies by diagnostic group and may influence task performance (Dennis et al., 2009)

Participants with ASD did not show a significant main effect of the number of talkers on RT, $F(1,12)=2.13$, $p=0.17$, $\eta_p^2=0.14$; they were equally fast to respond to target words within mixed- and single-talker blocks. There was a marginally significant main effect of expectations on RT amongst participants with ASD such that participants who expected a single talker were faster than those who expected two talkers, $F(1,12)=3.5$, $p=0.06$, $\eta_p^2=.38$. Furthermore, there was an interaction between acoustic variability and expectations in the ASD group, as shown in Figure 3a, $F(1,12)=7.81$, $p=0.01$, $\eta_p^2=0.38$.

TD participant RT was not significantly influenced by experimental manipulations, as seen in Figure 3b. There was no main effect on RT for number of talkers, $F(1,12)=0.01$, $p=0.92$, $\eta_p^2<0.01$, nor of expectations, $F(1,12)=1.63$, $p=0.56$, $\eta_p^2=0.12$. There was also no interaction between the number of talkers and expectations, $F(1,12)=0.35$, $p=0.56$, $\eta_p^2=0.03$.

Figure 3



Panel 3a (left) shows the interaction of acoustic variability and listener expectation within the ASD group; participants responded more quickly when their expectations aligned with stimulus features. Participants with TD (3b, right) did not show this effect.

Expectations x Variability across Diagnosis. Given that results in the TD group failed to replicate Magnuson & Nusbaum (2007), we explored whether this may be due to lack of power. (It is worth noting that while the ASD group did demonstrate an expectation by acoustic variability interaction, it differed in direction from the originally reported data.) In order to increase the functional sample size, we re-analyzed the data as described above collapsing across diagnostic groups. Once again, there was no main effect of acoustic variability, $F(1,28)=0.14$, $p=0.71$, or expectation condition, $F(1,28)=0.51$, $p=0.48$. Similarly, there was no interaction between acoustic variability and talker expectations, $F(1,28)=1.46$, $p=0.24$.

Sensory Differences and Acoustic Variability. Given our hypothesis that sensory processing differences might influence talker normalization effects in ASD, we next explored the role of sensory differences in accounting for this effect. Sensory processing differences were operationalized as the (overall) score on the Short Sensory Profile. Note that on the Short

Sensory Profile, higher scores indicate more *typical* sensory processing. We first obtained a measure of the degree of sensitivity to acoustic variability, calculated as the RT difference for blocks of single and multiple talkers (Multiple Talkers – Single Talker). We then ran a partial correlation of scores on the Short Sensory Profile and RT difference, controlling for average RT across conditions. This variable was included as a control, because an individual's overall RT influences the degree to which they can be influenced by the experimental manipulation. For example, a difference in RT of 500ms represents a 50% change if the overall RT average is 1000ms, but only a 25% change if the overall average RT is 2000ms. For this analysis, one participant with ASD was excluded due to missing Sensory Profile scores.

There was no significant correlation between ASD group Sensory Profile Score and RT difference as a result of acoustic variability, r^2= 0.08, p=0.33. Similarly, there was no correlation in the TD group, r^2<0.01, p=0.94. In order to assess whether these results were influenced by gross differences in hearing, we compared the performance of individuals who failed at least one item on our hearing screening to participants who passed all items, collapsing across diagnostic group. Individuals who failed at least one item did not differ in their performance on the task in terms of either overall accuracy (measured by d'), $t(29)=1.10$, p=0.22, or overall RT, $t(29)=1.82$, p=0.08.

Executive Functioning and Expectations

Another potential explanation for the significant effect of talker block in the ASD group is differences in executive functioning. A number of researchers have suggested that executive functioning may be a core deficit which underlies social, behavioral, and sensory abnormalities in ASD (Craig et al., 2016; Hill, 2004). It is possible that executive functioning challenges made it more difficult to switch from talker to talker. Given that some have hypothesized that

executive functioning plays a foundational role in ASD, any correlation between executive functioning and language impairment may be mediated by overall ASD severity.

In order to test this hypothesis, we conducted a mediation analysis to examine the relationships among BRIEF scores, SCQ total scores, and RT difference as a result of acoustic variability, including overall average RT as a covariate. This analysis used SCQ rather than ADOS scores because SCQ data was available for both the ASD and TD groups. We used 5000 bootstrapped samples to calculated bias-corrected bootstrap confidence intervals.

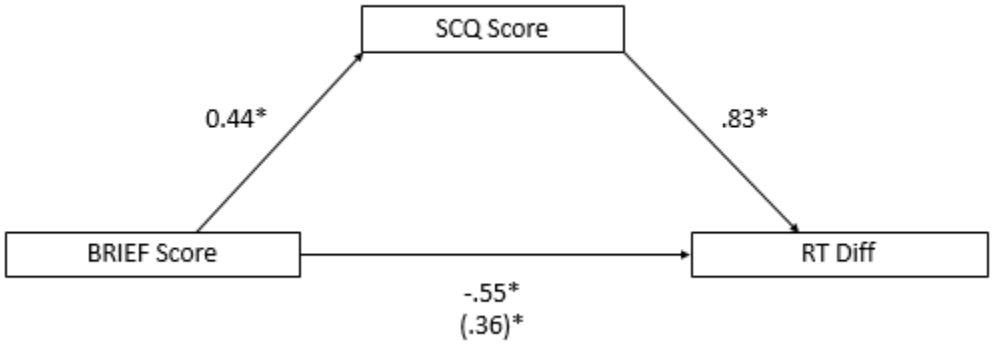


Figure 5: Results of analysis of the relationship between BRIEF scores and RT differences (as a results of acoustic variance), showing a significant partial mediation by autism symptom severity.

The results of this analysis are shown in Figure 5. There was a significant direct effect of BRIEF scores on RT difference, c'= -0.55, *95% CI* [-1.01, -0.08]. This relationship was partially mediated by autism symptom severity as measured by the SCQ, *ab*=0.37, *95% CI* [0.15, 0.78]. The results of this analysis indicate that the correlation between executive functioning and talker normalization effects is due in part to the relationship between executive functioning deficits and ASD severity.

Discussion

The present experiment investigated the extent to which individuals with ASD utilize expectations about talkers to actively process the speech signal. Utilizing Magnuson & Nusbaum's (2007) talker normalization paradigm, adolescents with ASD were asked to monitor a continuous stream of speech which was either acoustically constant or variable. Half of participants were told that this acoustic variability represented multiple talkers, while others were told that it merely represented one individual whose voice "went up and down" Given pragmatic language deficits and broad social deficits in ASD, we predicted an interaction between diagnosis and talker normalization effects, such that participants with ASD would show a weaker influence of their own expectations about the number of talkers to inform speech processing (i.e., expecting multiple talkers would *not* moderate the effect of acoustic variability on RT). Due to the statistical power of our sample, this hypothesis could not be directly tested. Furthermore, we anticipated that given sensory processing differences in ASD, variability in the speech signal would cause an effect on RT. We hypothesized that the TD sample would demonstrate a slowing of RT only when instructed to expect multiple talkers; when expecting only one talker, we hypothesized that the small amount of acoustic variability in the speech signal would not lead to a cognitive cost.

TD participants did not show any effect of acoustic variability on RT, even when instructions directed them to expect multiple talkers. This finding does not replicate previous research using this model. This lack of effect does not reflect lack of engagement with the task broadly; accuracy was high across all conditions, suggesting that participants were actively engaged in the task. Participants may have ignored or not believed the expectation manipulation and treated acoustic variability equally across conditions.

Our data did reveal an interaction between acoustic variability in the speech stream and expectations regarding the number of talkers in the ASD group, though this interaction was different than previously identified by Magnuson & Nusbaum (2007). Participants with ASD were faster to respond to target words when their expectations matched their sensory experience; that is, when participants with ASD expected to hear two talkers, they were faster when acoustic variability was present, and vice versa. Participants with ASD were also marginally slower overall when they expected multiple talkers.

Differences in sensory processing did not meaningfully relate to the degree of slowing as a result of acoustic variability across talkers as predicted. However, executive functioning did significantly predict the degree to which individuals were impaired by acoustic variability. Participants' scores on the BRIEF were significantly related to their ability to deal with acoustic variability in the speech signal. This relationship was mediated by overall autism symptom severity, suggesting that executive functioning significantly contributes to difficulty dealing with acoustic variability in ASD.

A great deal of literature has examined deficits in executive functioning in ASD. People with ASD demonstrate executive functioning deficits not only in so-called "hot" executive functioning tasks (e.g., emotion regulation, social cognition), but also "cold" executive functions such as working memory, planning, and cognitive flexibility (Zimmerman, Ownsworth, Donovan, Roberts, & Gullo, 2016). Furthermore, the degree of impairment on measures of these "cold" executive functions is related to social impairments within school-aged children with ASD (Freeman, Locke, Rotherman-Fuller, & Mandell, 2017). While less research has been conducted on executive functioning deficits in adults with ASD, there appears to be a high

degree of variability in executive functioning, even within samples with relatively average IQ (Brady et al., 2017).

In contrast to our data, Landa & Goldberg (2005) suggested that language skills and executive functioning were independent in ASD. In this study, the authors used the Clinical Examination of Language Fundamentals to assess expressive grammar skills. In contrast, our study examined more subtle language processing deficits. Participants in our sample did not have gross language impairments; rather, executive functioning was related to slowed processing, a more subtle measure which may have implications for fluid social cognition and interaction.

Executive dysfunction may be the result of weakened long-distance functional connectivity in ASD. Courchesne et al. (2007) suggest that brain development in ASD follows a different trajectory that may have cascading effects on cognition. At birth, individuals with ASD have on average smaller head circumference (Mason-Brothers et al., 1990), a measure which is highly correlated with brain size at birth (Bartholomeusz, Courchesne, & Karns, 2002). Over the first year of life, rapid growth takes place leading to larger-than-expected head circumference; however the next several years are marked by rapid *deceleration* of growth (Dawson et al., 2007; Hazlett et al., 2005).

Alterations in neuronal development may influence not only local brain volumes, but long-distance connectivity between regions. During the first year of life, white matter in the brain develops expansively and there is massive organization of long-distance white matter tracts (Dubois et al., 2014; Homae et al., 2010). In ASD the development of connectivity between regions is also disrupted (Belmonte et al., 2004). Courchesne et al. (2007) suggests that in ASD excess neuron *numbers* disrupt the development of brain circuits, resulting in an atypical pattern

of heightened local functional connectivity and weakened or noisier long-distance functional connectivity.

The results of these differences in connectivity may be profound. Language acquisition and processing is founded upon long-distance connectivity between frontal and temporal regions, and the development of these connections over the first two years of life may underlie the massive expansion in linguistic abilities (Bates et al., 1992). Indeed, typically developing one- to two-year olds produce significant activation of frontal, occipital, and cerebellar regions in response to forward compared to backward speech; this pattern of activation is diminished in three-year-olds, suggesting that this diffuse activation may facilitate the vocabulary burst that occurs over this period (Redcay, Haist, & Courchesne, 2008).

The atypical pattern of brain growth in ASD may "leave some neurons under-innervated and alter the afferent signals to these higher-order cortical regions" (Courchesne et al., 2007). Such a pattern may account for early delays in language development in ASD. However, while these neural differences may be the most important in the *development* of language, the top-down modulation of sensory regions is crucial in the *perception* of language. Changes in patterns of long-distance functional connectivity may represent not only the disruption of feed-forward processing of language, but also alterations in the extent to which individuals with ASD may be able to rely on knowledge or domain-general cognitive processes to regulate lower-order areas. Indeed, a number of studies have demonstrated a relationship between functional connective and executive functioning in ASD (Chan et al., 2009; Gilbert, Bird, Brindley, Frith, & Burgess, 2008; Han & Chan, 2017).

The TD group demonstrated none of the above effects, including previously well-documented costs of acoustic variability on speech processing. Taken on its own, the

discrepancy between diagnostic groups within our dataset does provide some evidence for greater interference of sensory sensitivity in ASD for speech processing. However, the absence of this expectation effect which has been produced in a previous study forces us to question why our results may differ from past studies, despite the fact that the present study utilizes stimuli and manipulations identical to previously published work (Magnuson & Nusbaum, 2007).

One obvious difference between the present study and previous work is the population sampled. Previous work examining talker normalization has sampled adult populations. The different results in this study may be the result of changes in speech processing over development. It is somewhat unusual however that adolescents with ASD, who broadly demonstrated delayed language development, *would* demonstrate a processing cost as a result of acoustic variability if this were the case.

Alternatively, over the past decade, synthesized talkers have become increasingly realistic. Computer talkers are also much more common in daily life; they are frequently included in applications on smart phones and smart home devices and are fairly sophisticated in their emulation of human speech (e.g., Apple's Siri). Today's participants may have more experience with this synthesized speech, leading them to question the apparent motive behind the acoustic variability. Informally, several participants in the present study remarked that the computer speech used was unsophisticated.

Limitations

Because the present study failed to replicate the results of previous research in the TD sample, we should be cautious about drawing conclusions about null results in the ASD sample. If previous findings had been replicated in the TD sample, then the pattern of results in the ASD sample would suggest that adolescents with ASD may have impaired top-down modulation of

speech processing, at least insofar as they fail to utilize social information (i.e., expectations about talkers) to modulate acoustic processing. However, the failure to demonstrate this effect in the TD sample makes it unclear whether or not the null results in the ASD sample are the result of true failure to utilize this information or represent a failure of the experimental manipulation. It is unwise to draw conclusions on the basis of null results.

The sample size of the present study also left us underpowered to examine some effects. Achieving enough power to consistently find a three-way interaction requires a large sample size. However, the sample in Magnuson & Nusbaum's (2007) study (Experiment Four) included only eight participants per expectation condition, and they found a moderate interaction effect size. Despite the additional grouping variable of diagnostic group, sample sizes in the current study were approximately equal to those in the original study. It is also noteworthy that due to the randomization procedures in the current study, there was an uneven distribution of participants across expectation conditions. Furthermore, the manipulation of expectations used in this paradigm necessitated a between-subjects design, inflating the number of participants needed to achieve appropriate power. Given the large number of characterizing and other experimental tasks completed by participants over the course of their enrollment in the study, it was logistically difficult to enroll a larger sample size. Of course, this small sample size may have limited our ability to detect small to medium effect sizes present in the population. This issue also led us to run separate analyses for diagnostic groups. By running multiple analyses, we run the risk of inflating the likelihood of Type I error, and limit the direct comparison of results across the two diagnostic groups.

It is important to note several potential limitations of the Short Sensory Profile for the current study. This measure of sensory processing differences was obtained via parent report

rather than direct measurement. Given that sensory sensitivities are a symptom of ASD, parents of adolescents with ASD may be more likely to notice and report these differences than parents of TD adolescents. More reliable results may have been obtained by directly measuring participants' auditory discrimination abilities. Furthermore, this measure assesses differences across sensory modalities rather than within the auditory domain. The breadth of the measure may therefore cloud the relationship between speech processing and the acoustic domain specifically.

Clinical Implications. The present work builds on a growing body of literature investigating ways that executive dysfunction in ASD may lead to or exacerbate other deficits. Despite the above limitations, participants with ASD did demonstrate small but significant costs associated with the expectation of acoustic variability of speech. While this pattern of results has also been true of TD adults, increased processing costs due to acoustic variability may be especially important in understanding language deficits in ASD.

A number of studies have suggested that individuals with ASD have heightened sensitivity to acoustic differences (Liss, Saulnier, Fein, & Kinsbourne, 2006; Ouimet, Foster, Tryfon, & Hyde, 2012). While these sensory differences typically lead to enhanced performance in low-level identification or discrimination tasks, they may interfere with efficiently processing information in more complex ways, such as integrating sensory features across multiple domains or integrating sensory features with expectations (Hubert, Mottron, Dawson, Soulie, & Burack, 2006).

These findings may have important implications for speech perception. Eigsti & Fein (2013) demonstrated that perceptual discrimination skills in adolescence were negatively correlated with acquisition of language milestones earlier in life. This relationship may be due to

generalization across early language tokens; if an individual is more aware of acoustic differences and/or less likely to group tokens together across talkers or situations, the difficulty of solving the many-to-many mapping problem of language and forming a stable acoustic-phonemic map must increase drastically (Bortfeld & Morgan, 2010).

This previous work has focused on the ways in which acoustic processing differences may influence language acquisition at a young age. However, the talker normalization research presented here suggest that subtle language processing deficits may continue into adolescence. The present experiment presents an artificial situation that could not happen in the real world, which may underestimate the impact of acoustic variability. For example, in the present experiment, acoustic variability was generated by manipulating F0 while leaving other aspects of the speech signal unchanged; introducing greater acoustic variability may make the impact of acoustic processing differences more apparent.

Could training paradigms be used to facilitate this process? Many social skills focused treatments focus on training individuals with ASD to attend to aspects of interactions that they may ignore, though these programs generally do not extend beyond pragmatics insofar as they relate to language. Some research has demonstrated that direct education can result in positive changes in speech *production* in ASD (Mayo, 2014), but no research has demonstrated if this strategy could also alter speech comprehension, especially in older children with ASD. Such a program may be problematic, because the cognitive cost associated with the conscious utilization of these cues may be higher than the cost associated from ignoring the cues altogether. It is also possible that the mere exposure to more speech inherent in these programs may also lead to more typical processing of acoustic variability.

However, intervention programs may be able to improve executive functioning skills. Given the relationship between executive functioning and atypical talker normalization patterns in the current study, rectifying these executive dysfunctions early in life could alleviate some of the long-term irregularities in speech processing. The best way to intervene in this area remains unclear, as the mechanism underlying executive functioning differences in ASD remains a matter of dispute.

Future Directions. One of the largest gaps in the present study is the lack of talker normalization effects in the TD sample. Before future research is conducted utilizing clinical populations, it may be important to validate a talker normalization paradigm in a typically developing adolescent population. Given the developmental trajectory of language acquisition, it is unlikely that the null result in the present study is merely the result of lack of linguistic skill. Rather, adolescents may need a richer manipulation in order to be influenced.

It may be beneficial therefore to forgo the experimental control offered by utilizing synthetic speech in favor of utilizing real talkers. Such a design would also afford the opportunity to study how alternating acoustic features along social important characteristics (e.g., gender) may differentially influence individuals with ASD. While expectations about the *number* of talkers may be more difficult to manipulate in such a paradigm, using voices of ambiguous genders without visual cues could offer an appropriate substitute. It is also important to consider that such a design would necessarily sacrifice control of the stability of the stimuli. By using natural talkers, it would be incredibly difficult to achieve a consistent, small difference between talkers necessary to elicit the expectation effect on talker normalization.

Finally, it is important to continue to investigate the nature of sensory processing differences in ASD and their relation to speech process over the course of language

development. In order to develop appropriate interventions for these issues, it is crucial to understand the degree to which these differences are caused by bottom-up or top-down differences in sensory processing. It remains for future research to determine whether differences in sensory processing relate to language acquisition and speech processing early in life.

This study investigated the role that sensory differences and talker expectations play in speech processing in ASD. Participants were told to expect speech from either one or two talkers, then monitored a continuous stream of speech under two conditions of acoustic variability. While the expectation manipulation did not elicit difference in RT, participants with ASD were significantly slower to react to target words embedded in acoustically variable speech. Furthermore, the effect of this variability was correlated with the degree of parent-reported sensory atypicality, independent of broader ASD symptom severity. These findings suggest that sensory processing differences in ASD may play a significant role in language acquisition and speech processing deficits characteristic of the disorder, above and beyond social dysfunction.

Appendix A- Expectation Instructions

One-Talker Instructions and Dialogue:

In this game, you will be listening to computer speech. Sometimes, computer speech sounds like a monotone. We want it to sound more natural, so we've changed the pitch for some words. Listen to how the computer voice will sound:

I have a ton of homework tonight. I'm not sure if I'm going to make it to practice. But if I don't make it to tonight's practice, then I won't be able to play in the game on Saturday. I don't want to miss the first game of the season, but I know that if I don't do my Spanish project, I may not get a passing grade. Why did I wait until the last minute to do the project? I knew that I would be benched for the rest of the season if I got a failing grade. Well, I guess I'll just have to miss practice to get the project done and wait until next week's game to play. And I should really try harder to get my grades up. My team needs me on the field.

[Procedural Instructions]

Remember, you will hear that one computer voice in this game. Sometimes the pitch will go up and down, but it is always the same voice.

Two-Talker Dialogue and Instructions:

In this game, you will listen to computer speech. We have changed the pitch of the computer voice so that it sounds like two people. Now we will play a recording of a dialogue between the two people as an example:

Bill: Joe, I have a ton of homework tonight. I'm not sure if I'm going to make it to practice.
Joe: But Bill if you don't make it to tonight's practice, then you won't be able to play in the game on Saturday.
B: I don't want to miss the first game of the season Joe, but I know that if I don't do my Spanish project, I may not get a passing grade.
J: Bill, why did you wait until the last minute to do the project? You knew that you'd be benched for the rest of the season if you got a failing grade.
B: Well Joe, I guess I'll just have to miss practice to get the project done and wait until next week's game to play.
J: Yea Bill, and you should really try harder to get your grades up. Your team needs you on the field.

[Procedural Instructions]

In some parts of the game, you will hear words from only one voice. In other parts, you will hear words from two voices.

References

Anderson, J. M., Gilmore, R., Roper, S., Crosson, B., & Bauer, R. M. (1999). Conduction Aphasia and the Arcuate Fasciculus : A Reexamination of the Wernicke – Geschwind Model. *Brain and Language*, *12*(151), 1–12.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?*. *Cogniton*, *21*, 37–46.

Bartholomeusz, H. H., Courchesne, E., & Karns, C. M. (2002). Relationship Between Head Circumference and Brain Volume in Healthy Normal Toddlers , Children , and Adults. *Neuropediatrics*, *33*, 239–241.

Bates, E., Thal, D., Finlay, B., Clancy, B., Origins, P. D.-, Bates, E., … Clancy, B. (1992). Early Language Development and its Neural Correlates. In I. Rapin & S. Segalowitz (Eds.), *Handbook of Neuropsychology* (Vol. 6). Amsterdam: Elsevier.

Belmonte, M. K., Allen, G., Beckel-Mitchener, A., Boulanger, L. M., Carper, R. A., & Webb, S. J. (2004). Autism and Abnormal Development of Brain Connectivity. *The Journal of Neuroscience*, *24*(42), 9228–9231. http://doi.org/10.1523/JNEUROSCI.3340-04.2004

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A. (2003). Enhanced Pitch Sensitivity in Individuals with Autism : A Signal Detection Analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226–235.

Bortfeld, H., & Morgan, J. L. (2010). Is early word-form processing stress-full? How natural variability supports recognition. *Cognitive Psychology*, *60*(4), 241–266. http://doi.org/10.1016/j.cogpsych.2010.01.002.Is

Brady, D. I., Saklofske, D. H., Schwean, V. L., Montgomery, J. M., Thorne, K. J., & Mccrimmon, A. W. (2017). Executive Functions in Young Adults With Autism Spectrum Disorder. *Focus on Autism and Other Developmental Disabilities*, *32*(1), 31–43. http://doi.org/10.1177/1088357615609306

Capps, L., Kehres, J., & Sigman, M. (1998). Conversational abilities mong children with autism and children with developmental delays. *Autism1*, *2*(4), 325–344.

Caron, M. J., Mottron, L., Berthiaume, C., & Dawson, M. (2006). Cognitive mechanisms, specificity and neural underpinnings of visuospatial peaks in autism. *Brain*, *129*(7), 1789–1802. http://doi.org/10.1093/brain/awl072

Chan, A. S., Cheung, M., Han, Y. M. Y., Sze, S. L., Leung, W. W., Sum, H., & Yee, C. (2009). Executive function deficits and neural discordance in children with Autism Spectrum Disorders. *Clinical Neurophysiology*, *120*(6), 1107–1115. http://doi.org/10.1016/j.clinph.2009.04.002

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*. http://doi.org/10.1016/j.tics.2012.02.007

Courchesne, E., Pierce, K., Schumann, C. M., Redcay, E., Buckwalter, J. A., Kennedy, D. P., & Morgan, J. (2007). Review Mapping Early Brain Development in Autism. *Neuron Review*, *56*, 399–413. http://doi.org/10.1016/j.neuron.2007.10.016

Craig, F., Margari, F., Legrottaglie, A. R., Palumbi, R., de Giambattista, C., & Margari, L. (2016). A review of executive function deficits in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Neuropsychiatric Disease and Treatment*, *12*, 1191–1202.

Dawson, G., Munson, J., Webb, S. J., Nalty, T., Abbott, R., & Toth, K. (2007). Rate of Head Growth Decelerates and Symptoms Worsen in the Second Year of Life in Autism.

Biological Psychology, *61*(4), 458–464. http://doi.org/10.1016/j.biopsych.2006.07.016.Rate
Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic Loci and Transitional Cues for Consonants. *The Journal of the Acoustical Society of America*, *27*(4), 769–773. http://doi.org/10.1121/1.1908024
Dennis, M., Francis, D. J., Cirino, P. T., Schachar, R., Barnes, M. A., & Fletcher, J. M. (2009). Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal of the International Neuropsychological Society*, *15*, 331–343. http://doi.org/10.1017/S1355617709090481
Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop-consonant recognition : Release bursts and formant transitions as functionally equivalent , context-dependent cues. *Perception & Psychophysics*, *22*(2), 109–122.
Dubois, J., Dehaene-Lambertz, G., Kulikova, S., Poupon, C., Huppi, P. S., & Hertz-Pannier, L. (2014). THE EARLY DEVELOPMENT OF BRAIN WHITE MATTER : A REVIEW OF IMAGING STUDIES IN FETUSES , NEWBORNS AND INFANTS. *Neuroscience*, *276*, 48–71. http://doi.org/10.1016/j.neuroscience.2013.12.044
Eigsti, I., & Bennetto, L. (2009). Grammaticality judgments in autism: Deviance or delay. *Journal of Child Language*, *36*(5), 999–1021. http://doi.org/10.1017/S0305000909009362
Eigsti, I. M., De Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, *5*(2), 681–691. http://doi.org/10.1016/j.rasd.2010.09.001
Eigsti, I. M., & Fein, D. A. (2013). More is less: Pitch discrimination and language delays in children with optimal outcomes from autism. *Autism Research*, *6*(6), 605–613. http://doi.org/10.1002/aur.1324
Freeman, L. M., Locke, J., Rotherman-Fuller, E., & Mandell, D. (2017). Brief Report : Examining Executive and Social Functioning in Elementary-Aged Children with Autism impairments in social functioning and communication. *Journal of Autism and Developmental Disorders*, *47*(6), 1890–1895. http://doi.org/10.1007/s10803-017-3079-3
Frith, U., & Happe, F. (1994). Autism: beyond "theory of mind." *Cognition*, *50*(1–3), 115–132. http://doi.org/10.1016/0010-0277(94)90024-8
Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377.
Galbraith, G. C., & Arroyo, C. (1993). Selective attention responses and brainstem. *Biological Psychology*, *37*, 3–22.
Gamliel, I., Yirmiya, N., Jaffe, D. H., Manor, O., & Sigman, M. (2009). Developmental Trajectories in Siblings of Children with Autism : Cognition and Language from 4 Months to 7 Years. *Journal of Autism and Developmental Disorders*, *39*, 1131–1144. http://doi.org/10.1007/s10803-009-0727-2
Giard, M.-H., Collet, L., Bouchet, P., & Pernier, J. (1994). Auditory selective attention in the human cochlea. *Brain Research*, *633*, 353–356.
Gilbert, S. J., Bird, G., Brindley, R., Frith, C. D., & Burgess, P. W. (2008). Atypical recruitment of medial prefrontal cortex in autism spectrum disorders : An fMRI study of two executive function tasks. *Neuropsychologia*, *46*(9), 2281–2291. http://doi.org/10.1016/j.neuropsychologia.2008.03.025
Gioia, G., Guy, S., Isquith, P., & Kenworthy, L. (1996). *Behavior rating inventory for executive function*. Psychological assessment resources.
Han, Y. M. Y., & Chan, A. S. (2017). Disordered cortical connectivity underlies the executive

function deficits in children with autism spectrum disorders. *Research in Developmental Disabilities*, *61*, 19–31. http://doi.org/10.1016/j.ridd.2016.12.010
Happe, F. G. E. (1997). Central coherence and theory of mind in autism : Reading homographs in context. *British Journal of Developmental Psychology*, *15*, 1–12.
Hazlett, H. C., Poe, M., Gerig, G., Smith, R. G., Provenzale, J., Ross, A., … Piven, J. (2005). Magnetic Resonance Imaging and Head Circumference Study of Brain Size in Autism. *Archive of General Psychiatry*, *62*, 1366–1376.
Heald, S. L. M., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, *8*(March), 1–15. http://doi.org/10.3389/fnsys.2014.00035
Heaton, P., Hermelin, B., & Pring, L. (2016). Autism and Pitch Processing : A Precursor for Savant Musical Ability ?, *15*(3), 291–305.
Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences*, *8*(1), 26–32. http://doi.org/10.1016/j.tics.2003.11.003
Homae, F., Watanabe, H., Otobe, T., Nakano, T., Go, T., Konishi, Y., & Taga, G. (2010). Development of Global Cortical Networks in Early Infancy. *The Journal of Neuroscience*, *30*(14), 4877–4882. http://doi.org/10.1523/JNEUROSCI.5618-09.2010
Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry*, *45*(2), 212–229.
Hubert, B., Mottron, L., Dawson, M., Soulie, I., & Burack, J. (2006). Enhanced Perceptual Functioning in Autism : An Update , and Eight Principles of Autistic Perception. *Journal of Autism and Developmental Disorders*, *36*(1). http://doi.org/10.1007/s10803-005-0040-7
Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nat Rev Neurosci*, *7*(December), 942–951. http://doi.org/10.1038/nrn2024
Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *Neuroreport*, *14*(8), 1129–1133. http://doi.org/10.1097/01.wnr.0000074343.81633.2a
Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Jindblom, B. (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science*, *255*(February), 606–608. http://doi.org/10.1126/science.1736364
Landa, R. J., & Goldberg, M. C. (2005). Language , Social , and Executive Functions in High Functioning Autism : A Continuum of Performance. *Journal of Autism and Developmental Disorders*, *35*(5), 557–573. http://doi.org/10.1007/s10803-005-0001-1
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461.
Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised*. *Cognition*, *21*, 1–36.
Liss, M., Saulnier, C., Fein, D., & Kinsbourne, M. (2006). Sensory and attention abnormalities in autistic spectrum disorders. *Autism : The International Journal of Research and Practice*, *10*(2), 155–172. http://doi.org/10.1177/1362361306062021
Lord, C., Rutter, M., DiLavorie, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism diagnostic observation schedule, (ADOS-2) modules 1-4*. Los Angeles, California: Western Psychological Services.
Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 391–409. http://doi.org/10.1037/0096-

1523.33.2.391
Mason-Brothers, A., Ritvo, E. R., Pingree, C., Petersen, P. B., Jenson, W. R., McMahon, W. M., … Ritvo, A. (1990). The UCLA-University of Utah Epidemiologic Survey of Autism: Prenatal, Perinatal, and Postnatal Factors. *Pediatrics*, *86*(4), 514 LP-519. Retrieved from http://pediatrics.aappublications.org/content/86/4/514.abstract
Matchin, W., Groulx, K., & Hickok, G. (2014). Audiovisual Speech Integration Does Not Rely on the Motor System : Evidence from Articulatory Suppression , the McGurk Effect , and fMRI. *Journal of Cognitive Neuroscience*, *26*(3), 606–620. http://doi.org/10.1162/jocn
Mayo, J. (2014). *Production of prosodic cues to clause structure in ASD: Role of intervention and working memory*. University of Connecticut.
McIntosh, D. N., Miller, L. J., Shyu, V., & Dunn, W. (1999). Development and validation of the short sensory profile. In *Sensory profile manual* (pp. 59–73).
Mottron, L., Dawson, M., Souli??res, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*. http://doi.org/10.1007/s10803-005-0040-7
Ouimet, T., Foster, N. E. V, Tryfon, A., & Hyde, K. L. (2012). Auditory-musical processing in autism spectrum disorders: A review of behavioral and brain imaging studies. *Annals of the New York Academy of Sciences*, *1252*(1), 325–331. http://doi.org/10.1111/j.1749-6632.2012.06453.x
Ozonoff, S., & Miller, J. N. (1996). An exploration of right-hemisphere contributions to pragmatic impairments of autism. *Brain and Language*, *52*(3), 411–434.
Redcay, E., Haist, F., & Courchesne, E. (2008). Functional neuroimaging of speech perception during a pivotal period in language acquisition. *Developmental Science*, *11*(2), 237–252. http://doi.org/10.1111/j.1467-7687.2008.00674.x
Roid, G. H. (2003). *Stanford-Binet intelligence scales.* Itasca, IL: Riverside Publishing.
Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008). Deficient brainstem encoding of pitch in children with Autism Spectrum Disorders. *Clinical Neurophysiology*, *119*, 1720–1731. http://doi.org/10.1016/j.clinph.2008.01.108
Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*. Western Psychological Services.
Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2010). Compensation for Coarticulation : Disentangling Auditory and Gestural Theories of Perception of Coarticulatory Effects in Speech. *Journal of Experimental Psychology*, *36*(4), 1005–1015. http://doi.org/10.1037/a0018391
Zimmerman, D. L., Ownsworth, T., Donovan, A. O., Roberts, J., & Gullo, M. J. (2016). Independence of Hot and Cold Executive Function Deficits in High-Functioning Adults with Autism Spectrum Disorder. *Frontiers in Human Neuroscience*, *10*(February), 1–14. http://doi.org/10.3389/fnhum.2016.00024