2015

Speech development in toddlers at high and low risk for autism

Chenausky, Karen Virginia

http://hdl.handle.net/2144/13665

Boston University
Dissertation

SPEECH DEVELOPMENT IN TODDLERS
AT HIGH AND LOW RISK FOR AUTISM

by

KAREN V. CHENAUSKY

S.B., Massachusetts Institute of Technology, 1989
M.A., University of Connecticut, 1992
M.S., Boston University, 2005

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2015
Approved by

First Reader
Melanie L. Matthies, Ph.D., CCC-A
Associate Professor of Speech, Language, and Hearing Sciences
Senior Associate Dean
Sargent College of Health and Rehabilitation Sciences

Second Reader
Helen Tager-Flusberg, Ph.D.
Professor of Psychological & Brain Sciences
Professor of Anatomy & Neurobiology

Third Reader
Kristine Strand, Ed.D., CCC-SLP
Clinical Associate Professor of Speech, Language, and Hearing Sciences

Fourth Reader
Suzanne Boyce, Ph.D.
Professor
Department of Communication Sciences and Disorders
College of Allied Health Sciences
University of Cincinnati
DEDICATION

I would like to dedicate this work to the families, including my own,

who have put so much of their own effort into this research.
ACKNOWLEDGMENTS

Melanie Matthies has been the midwife of this dissertation, especially helping me with quantitative analyses. Helen Tager-Flusberg’s data were the basis of this work, and her comments shaped it. Kristine Strand provided the original kernel of an idea, which has blossomed into a healthy shrub. And Suzanne Boyce was never worried about my finishing.
SPEECH DEVELOPMENT IN TODDLERS
AT HIGH AND LOW RISK FOR AUTISM

KAREN V. CHENAUSKY

Boston University Sargent College of Health and Rehabilitation Sciences, 2015

Major Professor: Melanie L. Matthies, Ph.D., CCC-A, Associate Professor of Speech, Language, and Hearing Sciences, Senior Associate Dean, Sargent College of Health and Rehabilitation Sciences

ABSTRACT

Speech development in autism spectrum disorder (ASD) has rarely been studied, yet residual speech sound errors are over 30 times more common in children with ASD than in the general population. Two main theories could explain this. The Social Feedback Loop proposes that toddlers with ASD vocalize less often and so have fewer opportunities to benefit from adult feedback. Thus, fewer vocalizations and perceptible differences in those vocalizations should be found in toddlers with ASD. The Speech Attunement Framework proposes that while toddlers with ASD “tune in” to their native languages well enough to acquire language normally, they lack the ability to “tune up” their articulation to the same level of precision as typically-developing children. Thus, differences in the vocalizations of toddlers with ASD may be perceptible or not, and should exist irrespective of differences in vocalization rate.

This study longitudinally examines vocalization rate, consonant inventory size, and voice-onset time (VOT) in syllable-initial bilabial stops using 30-minute speech samples from toddlers in three groups: those at low risk for ASD (LRC),
those at high risk for ASD with ASD themselves (HRA+), and those at high risk for ASD without ASD (HRA-). Transient delays in consonant development were found in both HRA- and HRA+, but only HRA+ toddlers vocalized less often. Further, the relationship between vocalization rate and consonant inventory was significantly different from LRC only for HRA-. VOT development was similar across groups, except that fewer HRA+ 36-month-olds produced distinct /b/ and /p/ populations, as measured by t-test and by Cohen’s d ≥ 0.8 between mean VOTs in the two populations of stops.

Results support the Speech Attunement Framework. Consonant acquisition delays are not related to differences in vocalization rate and are not found only in toddlers who develop ASD. The finding of sub-perceptual acoustic differences in stop production in toddlers who develop ASD, with no accompanying differences in production rate, also supports the Speech Attunement Framework. This suggests that the Social Feedback Loop is not diminished in ASD by lower vocalization rate, but that toddlers with ASD may have diminished ability to monitor their own speech.
PREFACE

Several years ago, I worked in a high school program for teens with high-functioning autism (we now refer to this group as having ASD with intact language). These young people ranged from academically average to a lot smarter than you or me. One particular student – let’s call him Mark – inspired this work.

Mark was one of the sweetest kids you could ever hope to meet. He was smart, especially at computers. He once taught me to use a program for a smart blackboard that he’d only seen another adult use a couple of times. He loved cooking and baking, and we gained a lot of weight eating the cookies and cupcakes he brought for everyone’s birthdays. His speech wasn’t the clearest, though he didn’t have any actual residual speech errors like some of the other students in the program did.

But Mark was incredibly friendly, funny, and outgoing. At least once a month he made me laugh so hard my eyes watered. He never made fun of his peers, except when it was clear that we were all gently teasing. Of course, being autistic, his social overtures sometimes revolved around asking people if they would open their mouths so that he could look at their teeth. But, all in all, Mark was a kid you wouldn’t mind hanging out with for an afternoon at a museum – not someone you felt you had to babysit or constantly be on the watch for.

About a year before he was to graduate, I attended a transition planning meeting with him, his parents, the other teachers, and the program director. His mother, a very polished corporate trainer, asked the director to provide articulation
therapy for him. After a few initial refusals, the director finally asked, exasperated, “Why now? Why work on artic therapy when that’s something for first-graders?”

“I’m just afraid,” his mother said, “that when he goes to apply for a job, people won’t think he’s as smart as he is.”

This moment crystallized several issues for me at once. Mark’s mother was referring to the fact that, like a lot of his peers with ASD and intact language, Mark’s speech sounded “funny.” She was also acknowledging that this impression of oddness is extremely socially penalizing. It can interfere with getting friends or a job, as she rightly pointed out. And her request brought home to me the fact that, rather than being a cosmetic procedure of sorts, articulation therapy for young adults with ASD may actually make the difference between social success and being shunned, or between living independently and being dependent on parents or siblings for life.

Mark got his speech therapy. But how did we get to the point where a smart, capable young adult with ASD needed articulation therapy? That question is what spurred the research you’re going to read about. In order to understand how these differences in speech patterns arise, and (one hopes) address them therapeutically when they arise instead of many years after the fact, we need first to understand how spoken language develops in individuals with ASD, and how it may be different from spoken language development in individuals who do not have ASD.
TABLE OF CONTENTS

DEDICATION ........................................................................................................................................ iv
ACKNOWLEDGMENTS ....................................................................................................................... v
ABSTRACT ................................................................................................................................................ vi
PREFACE .................................................................................................................................................. viii
TABLE OF CONTENTS ............................................................................................................................ x
LIST OF TABLES .......................................................................................................................................... xiii
LIST OF FIGURES ....................................................................................................................................... xiv
LIST OF ABBREVIATIONS ........................................................................................................................ xv

CHAPTER ONE: Autism Spectrum Disorder and the Broader Autism Phenotype .......... 1
   Section One: Definition of Autism Spectrum Disorder ................................................................. 1
   Section Two: Sibling Studies ............................................................................................................ 2
   Section Three: The BU/BCH Infant Sibling Project ...................................................................... 4

CHAPTER TWO: Speech Development in ASD ............................................................................... 9
   Section One: Speech vs. Phonological Development ................................................................. 9
   Section Two: Typical Speech Development ............................................................................... 11
   Section Three: What is Known About Speech Development in ASD ..................................... 13
   Section Four: Theories of Speech Development in ASD ........................................................ 16
      Subsection One: Social Feedback Loop .................................................................................... 16
      Subsection Two: Speech Attunement Framework ................................................................. 18
      Subsection Three: Comparison of Predictions ...................................................................... 19
   Section Five: Hypotheses ............................................................................................................. 20
CHAPTER THREE: Perceptual Study: Vocalization Rate and Consonant Development............................................................................................................. 24

Section One: Participants ......................................................................................................................................................................................... 24

Section Two: Methods......................................................................................................................................................................................... 25

Subsection One: Standardized Assessments ................................................................................................................................. 25

Subsection Two: Equipment ............................................................................................................................................................................ 25

Subsection Three: Coding .................................................................................................................................................................................. 26

Subsection Four: Coding Reliability .............................................................................................................................................................. 27

Subsection Five: Perceptual Analyses ......................................................................................................................................................... 28

Section Three: Results ..................................................................................................................................................................................... 28

Subsection One: Vocalization Rate .............................................................................................................................................................. 28

Subsection Two: Phonological Complexity .................................................................................................................................................. 31

Subsection Three: Modeling the Relationship Between Speechlike Vocalization Rate and Number of Middle-8 Consonants ....................... 34

Section Four: Discussion: Vocalization Rate and Consonant Development........ 38

CHAPTER FOUR: Acoustic Study: The Voiced/Voiceless Distinction .................. 43

Section One: Voice Onset Time Development and Analysis........................................ 43

Subsection One: Acoustic vs. Perceptual Analyses .............................................................................................................................. 43

Subsection Two: Phonetic Contrast ................................................................................................................................................................. 44

Subsection Three: Stop Consonant Production .................................................................................................................................................. 45

Subsection Four: Acquisition of the Voicing Contrast in Bilabial Stops ............ 49

Subsection Five: Assessment of Production Accuracy ........................................................................................................................................... 51

Section Two: Participants .................................................................................................................................................................................. 52

Section Three: Methods ..................................................................................................................................................................................... 53
<table>
<thead>
<tr>
<th>Subsection One: Procedures</th>
<th>Subsection Two: Analyses</th>
<th>Subsection Three: Measurement Reliability</th>
<th>Subsection Four: Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section Four: Results</th>
<th>Subsection One: Language Ability</th>
<th>Subsection Two: Token Counts and Error Rates</th>
<th>Subsection Three: Voice Onset Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subsection Four: Production Accuracy</th>
<th>Subsection Five: Controlling for Differences in Language Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section Five: Discussion: Voice Onset Time Development</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER FIVE: Overall Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section One: Implications for Theory</th>
<th>Section Two: Clinical Implications</th>
<th>Section Three: Limitations and Future Work</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BIBLIOGRAPHY</th>
<th>CURRICULUM VITAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. Early-, Middle-, and Late-8 Consonants................................................................. 12
Table 2. Mean Language Scores by Age and Group (Perceptual Study) ......................... 25
Table 3. % Canonical Babbles by Age and Group................................................................. 32
Table 4. Consonant Inventories by Age and Group......................................................... 34
Table 5. Predicted vs. Actual 12-Month M8 Inventories by Group............................... 37
Table 6. Predicted vs. Actual 18-Month M8 Inventories by Group............................... 37
Table 7. Mean Language Scores by Age and Group (Acoustic Study) ......................... 59
Table 8. Total Errors by Age and Group....................................................................... 62
Table 9. Number of Toddlers with Distinct (p < 0.05) /b/ and /p/ VOT Populations by Age and Group....................................................................................... 67
Table 10. Number of Toddlers with Large Cohen’s d (d > 0.8) Separating /b/ and /p/ VOT Populations by Age and Group.............................................................. 68
LIST OF FIGURES

Figure 1. Total Vocalization Rate by Age and Group........................................29

Figure 2. Speechlike Vocalization Rate by Age and Group..........................30

Figure 3. Nonspeechlike Vocalization Rate by Age and Group....................31

Figure 4. Best-Fit Curves for 12-Month Speechlike Vocalization Rate vs. M8

  Inventory Size, by group........................................................................35

Figure 5. Best-Fit Curves for 18-Month Speechlike Vocalization Rate vs. M8

  Inventory Size, by group........................................................................36

Figure 6. VOT Measurements for /b/ in "a big one"......................................55

Figure 7. VOT Measurements for /b/ in "baby"............................................56

Figure 8. VOT Measurements for /b/ in "but"...............................................56

Figure 9. VOT Measurements for /p/ ............................................................57

Figure 10. Mean Number of /b/ and /p/ Tokens by Age and Group............61

Figure 11. Mean /b/ VOT by Age and Group.............................................63

Figure 12. Mean /p/ VOT by Age and Group.............................................63

Figure 13. Mean Within-Child Standard Deviation of VOT for /b/ and /p/ by Age

  and Group................................................................................................66
LIST OF ABBREVIATIONS

ADOS........................................................................................................... Autism Diagnostic Observation Scales
ANOVA........................................................................................................ Analysis of Variance
AOSI ............................................................................................................... Autism Observation Scale for Infants
ASD............................................................................................................... Autism Spectrum Disorder
DD............................................................................................................... Developmentally Delayed
EEG............................................................................................................. Electro-Encephalogram
EL............................................................................................................... Expressive Language
HR............................................................................................................... High Risk
HRA+ ......................................................................................................... High Risk for ASD, Positive Diagnosis
HRA- ......................................................................................................... High Risk for ASD, Negative Diagnosis
LR............................................................................................................... Low Risk
LRC........................................................................................................... Low Risk for ASD (Negative Diagnosis)
MSEL........................................................................................................... Mullen Scales of Early Learning
PDD-NOS................................................................................................. Pervasive Developmental Disorder – Not Otherwise Specified
RL............................................................................................................... Receptive Language
TD............................................................................................................... Typically Developing
TDA........................................................................................................... Typically-Developing, Age-Matched
TDL........................................................................................................... Typically-Developing, Language-Matched
VOT........................................................................................................... Voice Onset Time
WCSD....................................................................................................... Within-Child Standard Deviation
CHAPTER ONE: Autism Spectrum Disorder and the Broader Autism Phenotype

Section One: Definition of Autism Spectrum Disorder

Autism spectrum disorder (ASD) is characterized by persistent deficits in social communication and by restricted, repetitive behaviors or interests that cause clinically significant functional impairment (American Psychiatric Association & American Psychiatric Association, 2013). The symptoms must be present from early childhood on, though they may only manifest themselves when social communication demands exceed an individual's capacity. Expressive language is also frequently delayed in ASD (Gamliel et al., 2009; Gernsbacher et al., 2006), but findings on phonological development, the topic of this work, are mixed. Some researchers (e.g., Kjelgaard and Tager-Flusberg, 2001) have found that phonological ability in ASD is in the normal range, even when language and cognitive ability are below normal. Other researchers (e.g., Rapin et al., 2009) have found evidence of subgroups of children with ASD with impaired phonology.

Speech development in ASD has rarely been studied, but it is important both scientifically and clinically. From a scientific point of view, the skills required to use speech communicatively overlap with the main areas of impairment in ASD. To use spoken language, one must have the social motivation and ability to communicate. One must also possess some capability for language, to give social and semantic meaning to speech. Finally, a degree of motor skill is required in order to perform the movements that comprise speech. These three domains – social communication,
language, and motor function – are the same domains in which the symptoms of ASD are defined.

Add to this the fact that spoken language ability forms its own spectrum within the autism spectrum, and the clinical reasons for studying speech development in ASD become clear. Approximately 9% of individuals with ASD have phonology within the average range and intact language (Rapin et al., 2009), but even in this group there is a much higher prevalence of residual speech errors in children above the age of eight – approximately 30% of teens and adults with ASD show residual speech errors, compared to only 1-2% in the general population (Shriberg et al., 2001). Another 42% of children with ASD have below-average phonological skills, and almost 25% have no language at all (Rapin et al., 2009). Understanding how speech development is affected in ASD can thus inform clinical practice for almost all individuals with ASD, and has the potential to positively impact the lives of the most severely affected individuals. The present studies will address the question of whether toddlers who go on to receive diagnoses of ASD develop speech differently than their peers who do not develop ASD, and whether any differences can be ascribed to ASD itself.

**Section Two: Sibling Studies**

In order to study speech development, we need to track children’s vocal productions from the earliest ages until at least kindergarten. But this poses a problem with ASD, because it is not possible to reliably diagnose ASD until 18 or 24 months at the
earliest (Ozonoff, et al. 2011; Chawarska et al., 2014). Thus, if we aim to track speech development in children who are later diagnosed with ASD, we need to start before they show definitive signs of the disorder. Knowing that 1 in 68 children will be diagnosed as being on the autism spectrum (Centers for Disease Control, 2010), it is theoretically possible but practically infeasible to recruit enough infants to eventually find a group of just 30 children with ASD. Over 2,600 infants would have to be followed to find that group of 30.

However, we also know that ASD is one of the most heritable of developmental disorders, if not the most heritable of them (Zwaigenbaum et al., 2013). The recurrence rate of ASD diagnosis in younger children who have an older sibling with ASD is almost 20% (Ozonoff et al., 2011; Hallmayer et al., 2011): one-fifth of younger siblings from these families will also have ASD. This fact makes the search for children with ASD more feasible, because only 150 high-risk families will be required to find a group of 30 infants who go on to develop ASD. This is one of the two great strengths of the infant sibling study paradigm.

The other important advantage of infant sibling studies is that, in addition to revealing much about the early development of children with ASD, they have also led to the identification of ‘endophenotypes’, or collections of more subtle traits of markers that are related to an elevated genetic risk for ASD. While a fifth of younger siblings develop a full-blown diagnosis themselves, and while approximately half develop typically, the remaining children are more likely to show sub-threshold levels of ASD symptoms (i.e., are part of the broader autism phenotype) or have
learning, communication, or attention deficits (Ozonoff, et al., 2014). The elevated prevalence of such developmental differences in siblings of children with ASD has led to a view of ASD that, rather than being qualitatively different from typical development, it is the result of an accumulation of traits or markers that, together, give rise to a diagnosis.

In the work discussed here, a group of toddlers at high risk for developing ASD (by virtue of having an older sibling with the diagnosis) will be compared to two other groups of toddlers: a group of high-risk siblings who do not develop ASD and a group of low-risk siblings (those who have no first-degree relatives with ASD). Prospectively tracking high-risk siblings who do and do not develop ASD yields greater insight into both the developmental vulnerabilities and the protective factors or skills that these children possess. Identification of vulnerabilities is crucial to designing effective, targeted therapeutic interventions that minimize the downstream effects of a diagnosis. It is equally important, however, to identify the strengths that high-risk children bring to the table.

**Section Three: The BU/BCH Infant Sibling Project**

The work that will be discussed here is part of a larger ongoing study of infant siblings of children with ASD (the Infant Sibling Project, or ISP) conducted jointly by Boston University and Boston Children’s Hospital. Interested families were contacted by the study coordinator, who conducted a detailed telephone eligibility interview. Exclusionary criteria included a gestational age less than 36 weeks; an
extended stay in a neonatal intensive care unit; maternal drug, alcohol, or steroid use during pregnancy; family history of genetic disorders associated with ASD; or a primary language other than English. Infants were enrolled into the high risk for ASD group if they had an older sibling with a diagnosis of Autism, Asperger Syndrome, or Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS). Diagnostic information for the proband was obtained from parent report during the screening process and confirmed for probands at least 48 months of age by a score over 14 on the Social Communication Questionnaire (SCQ, Rutter et al., 2003), meeting diagnostic criteria on the Autism Diagnostic Observation Schedule—Revised (ADOS; Lord et al., 2000), or through diagnosis by an expert clinician in the community. Infants were enrolled into the low-risk control group if they had at least one older sibling who was typically developing, confirmed by a score under 12 on the SCQ for those over 48 months of age; and no first-degree relatives diagnosed with an ASD or another neurodevelopmental disorder. Informed consent was obtained from parents prior to participation.

Previous work from the ISP has revealed much about both the strengths and challenges of these participants. For example, Wagner et al. (2013) used eye-tracking methods to show that there were age-related, but not group-related, changes in visual attention in infants from 6 to 12 months of age. While 6- and 9-month-olds looked equally often at the eyes and mouth of their mothers or a unfamiliar speaker, 12-month-olds looked a little more at mouths and a little less at eyes. However, there was no age at which the infants looked longer at mouths than
at eyes (i.e., the differences in looking location and duration were not significant). Differences between risk groups were not assessed in this study, as stable diagnoses were not yet available. However, the results pertain to visual attention development in general, and no bimodal distribution in attentional variables was identified that would indicate between-group differences. Luyster et al. (2014) examined neural markers of social attention in the infant sibs up to 12 months of age, finding that four EEG responses of high-risk infants to social stimuli were generally indistinguishable from those of low-risk infants. Together, these studies suggest that group-level differences in social attention do not appear before the age of one year.

Seery et al. (2013) looked at event-related EEG potentials (ERPs) in response to repeated speech stimuli in 9-month-old high- and low-risk siblings. Unlike the low-risk infants, the high-risk infants showed a faster response to repetitions of the syllable /da/ over the right hemisphere than the left. For the high-risk, but not the low-risk siblings, the ERP amplitude to the repeated stimulus was positively related to later language ability. Also, when Seery et al. (2014) examined ERP response to a non-native deviant stimulus (the syllable /gda/, with a dentalized alveolar stop), the high-risk siblings showed the same perceptual narrowing (Werker & Tees, 2002) to phonemes not in the ambient language as did the low-risk siblings. The results were interpreted as being consistent with other findings suggesting atypical lateralization for language in children with ASD in the first year of life, but also suggesting that perceptual narrowing took place at the same age in high-risk as in low-risk siblings.
Other work by Talbott et al. (2015a) showed that 12-month-old infants later diagnosed with ASD produced about half the number of communicative gestures as did infants of the same age who did not receive diagnoses of ASD. Mothers of the high-risk infants not diagnosed with ASD produced significantly more of their own gestures than mothers of the low-risk infants. And for all of the infants who did not receive ASD diagnoses, maternal gesture use was associated with infant language scores at 18 months, even after controlling for child gesture use. Later work by Talbott et al. (2015b) found mothers of high-risk infants to be significantly more likely at 6, 9, and 12 months of age to report concerns relating to the core symptoms of ASD than were mothers of low-risk infants. However, at no age did more mothers of high-risk infants with ASD report concerns about their children’s development than did mothers of high-risk infants without ASD.

Finally, work by Talbott et al. (in prep.) concerned maternal verbal responses to their infants’ vocalizations. Mothers’ vocalization rates did not differ across groups, and neither did mothers differ in the proportion of infant vocalizations that they responded to. All three groups of mothers (mothers of high-risk infants later diagnosed with ASD, mothers of high-risk infants not diagnosed with ASD, and mothers of low-risk infants) responded more frequently with feedback known to promote language to infants’ vocalizations that contained consonants, as compared to the vocalizations that contained only vowels. Thus, Talbott et al. conclude, feedback from mothers of high-risk infants is no different in kind or frequency than that of mothers of low-risk infants, and less or lower-quality maternal feedback is
not a factor responsible for poorer language outcomes in high-risk infants. Quite the contrary: the high-quality, high-frequency contingent feedback that mothers of high-risk infants provide was considered a protective factor in the language development of these children.

Taken together, previous findings from the infant sibs indicate that all three groups of infants performed similarly on many developmentally-related tasks and experienced similar developmental environments. Before one year of age, there were no group differences in social attention. Perceptual narrowing, in favor of phonemes of the ambient language, took place at the same time across groups. And all three groups received the same amount and quality of maternal feedback to their vocalizations that contained CV syllables. Where groups differed was in the rate of ASD-related concerns that the mothers reported, which was higher for the mothers of high-risk infants that for the mothers of low-risk infants (though the number of concerns did not differ by eventual diagnosis within the high-risk group). Relatedly, high-risk infants where shown to gesture less frequently than low-risk infants, and they showed atypical neural lateralization to speech stimuli. Thus, the high-risk siblings, even those who eventually received diagnoses of ASD, show considerable strengths and experienced the same enriching environments as did their low-risk peers. What differences did show up were in their rate of communicative gesturing and in language-specific brain adaptations to language. It remains to be seen how these siblings differ in their development of spoken language.
CHAPTER TWO: Speech Development in ASD

Section One: Speech vs. Phonological Development

An important preliminary at this point is a discussion of terms related to spoken language development. For a child to develop spoken language normally, inputs and capabilities at a variety of levels are required (Bleile, 1998). First of all, she must be exposed to a language, along with a minimum level of social interaction from a caregiver. In order to develop abstract phonological representations of the sounds of language – to realize that the speech stream can be thought of as being comprised of units such as phonemes, syllables, and metrical feet – this child’s auditory processing must be intact.

Typical phonological development also requires that a child practice making speech sounds, in order to learn the relationship between articulation and acoustics (also called the “feedforward mode” in some frameworks, such as that of Tourville and Guenther (2011)). With practice and maturation of speech motor skills, this child develops a gestural score for the words of her native language. Experience with using speech sounds to convey meaning contributes to the development of increasingly numerous and sophisticated representations of words. Thus, the phonological representation of a word emerges from all of the child’s experiences with its phonetic characteristics and with its semantic, syntactic, and phonological relationships to other words in the lexicon.

The focus of the work discussed here will be on the transition from pre-speech babbling to first words, in toddlers whose diagnostic status (relative to ASD)
was ascertained at 36 months of age. In particular, the work will aim to replicate previous findings about consonant development and production between the ages of one and three years. Both perceptual and acoustic methods of evaluating speech will be used. Some of the analyses will take into account what consonants appear in the phonemic inventories of the participants, while others will focus on phonemes as produced correctly in words. Consonants that appear in pre-speech babble are assumed to be speechlike but, in the words of Rvachew and Brousseau-Lapré (2012), there is no expectation that they be communicative in function (p. 155).

When consonants are being used in words and being assessed as correctly produced or not, they do communicate meaning and therefore reflect a child’s developing phonological knowledge and ability in addition to demonstrating speech motor performance. Thus, although speech and phonological development are different and involve separate levels of abstraction, they will be considered part of the same process here. The term ‘speech development’ will be used as a general term that includes both motor speech development and cognitive phonological development. The usage is intended to reflect the idea that speech and phonological skills develop in tandem with each other in typical development, with pressures from the one driving development in the other at different times in a child’s life. The overall aim of this work is to identify how toddlers who develop ASD perform both similarly to and differently from toddlers who do not develop ASD on a series of milestones marking the journey from pre-speech babble to spoken production of language.
Section Two: Typical Speech Development

In order to understand how speech development might be affected in ASD, it is important to prospectively track the speech development of high-risk infants who do and do not develop ASD and to compare it to typical development.

The course of typical speech development has been described as occurring in five stages (Eilers and Oller, 1994; Oller, Eilers, and Basinger, 2001; Oller, 2000; Rvachew and Brousseau-Lapré, 2012). The ages at which the different stages are described as appearing are approximate and given in ranges, reflecting variability in the typical population. In the Phonation stage (0-2 mos.), infants produce their first intentional vocalizations, called protovowels. Protovowels are centralized, often highly nasalized, vocalic sounds that are produced with the vocal tract more or less at rest. Next, during the Primitive Articulation stage (1-4 mos.), infants begin to produce their first syllables, which contain protovowel nuclei and glidelike onsets or offsets. By the Expansion stage (3-8 mos.), infants are producing fully adult-like vowels along with a variety of other sounds, such as squeals, growls, yells, whispers, and raspberries. In the Canonical Syllable stage (5-10 mos.), toddlers develop the ability to produce well-formed singleton canonical syllables, which have rapid, adultlike consonantal onsets and offsets; and reduplicated sequences of canonical syllables. Finally, in the Variegated stage (9-18 mos.), toddlers produce a mixture of babbling, word approximations, and meaningful speech.

The phonemes of a child’s native language are acquired gradually as well, and the ages at which specific consonants are said to be acquired differ greatly between
children. Two methods of describing consonant mastery are typically employed in studies of phonological development (Sander, 1972). One method is to perceptually identify the appearance of the sound form of a consonant in babble, where the target production is not known. This criterion reflects the growing motor ability of a child, reflected in his or her ability to articulate sounds that resemble adult consonants. The other method involves perceptually judging whether a child has correctly produced a consonant in a word, where the target is known. This is a more stringent criterion, reflecting the ability of a child to intentionally produce a phoneme that conveys meaning to adults, and to distinguish that phoneme from others that differ from it in one or more phonetic features. Based on the latter criterion, the consonants of English have been divided into three groups based on trends of their acquisition in both typically developing children and in children with speech sound disorders (Shriberg, 1993). These groups are referred to as the Early-8, Middle-8, and Late-8 consonants, and they are listed in Table 1.

<table>
<thead>
<tr>
<th>Early-8</th>
<th>m, b, j, n, w, d, p, h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Middle-8</td>
<td>t, η, k, g, f, v, tf, dʒ</td>
</tr>
<tr>
<td>Late-8</td>
<td>ʃ, θ, ð, s, z, l, ʒ, r</td>
</tr>
</tbody>
</table>

Table 1. Early, Middle, and Late 8 Consonants (Shriberg, 1993)

The Early-Middle-Late-8 classification has also been used to describe the order in which consonants appear in children’s babble inventories (Paul et al., 2011; Schoen et al., 2011), the logic being that consonants being produced correctly in words earlier also occur earlier in babble.
The transition from babble to words is a gradual one, during which children produce babbles and meaningful speech together. Practice babbling increases the skill with which children are able to produce speech sounds, and the larger a child’s repertoire of speech sounds in babble, the earlier she will learn to talk and the larger a productive vocabulary she will acquire (Rvachew & Brousseau-Lapré, 2012; Stoel-Gammon, 1998). Children tend to acquire words earlier that contain the sounds they produce in babble, avoiding words with phonemes that are difficult for them; and they acquire words later that contain sounds they rarely produce in babble. Thus, early in speech development it is the phonological patterns of their native languages that drive children’s word acquisition. For example, a child who produces a high proportion of anterior stops in babble will tend to pick up words with /b, p, t, d/ earlier than words with /g/ and /k/. Later, however, the relationship between phonology and lexicon becomes more bidirectional and semantic pressures begin to drive phonological development. A child who produces both voiced and unvoiced stops in babble will learn that they convey different meanings when she tries to produce words like “big” and “pig” to get what she wants.

Section Three: What is Known About Speech Development in ASD
Several research groups have examined speech development in ASD. Not all have also looked at receptive and expressive language at the same time, though these are
known to be related to phonological development (Tyler, 2010). Studies where language was not assessed have found lower vocalization rate and delays in the acquisition of complex syllable shapes in children with ASD. For example, Patten et al. (2014) report on a prospective study of 23 toddlers later diagnosed with ASD and 14 not diagnosed with ASD, examining vocalizations at 9-12 and at 15-18 months of age. The ASD group’s vocalization rate was significantly lower than that of the non-ASD group at both ages. Members of the ASD group also produced a significantly lower proportion of canonical syllables at each age and were less likely to be classified as being in the canonical babbling stage (defined as producing at least 15% canonical babbles) at each age than the non-ASD group. Warlaumont et al. (2014) examined 12-hour audio recordings from 77 children with ASD and 106 typically-developing children between the ages of 16 and 48 months, finding a significantly lower vocalization rate in the ASD group as well as a lower rate of speechlike vocalization.

However, other groups have examined language and speech development together in ASD. For example, Paul et al. (2011) looked cross-sectionally at toddlers at high and low risk for ASD at 6, 9, and 12 months of age. The high-risk (HR) toddlers in this study produced fewer speechlike vocalizations and more nonspeech vocalizations than the low-risk (LR) toddlers. At 9 months, the HR group produced significantly fewer canonical syllables and fewer consonants of all types than the LR group and showed lower expressive and receptive language scores than the LR group at 6, 12, and 24 months.
Another cross-sectional study by Schoen et al. (2011) studied 30 toddlers with ASD, 11 typically developing, age-matched (TDA) controls, and 23 typically developing, language-matched (TDL) controls, aged 18 to 36 months. Toddlers with ASD produced the same number of total vocalizations as both other groups and significantly more nonspeech vocalizations than the TDA group. The participants with ASD who produced word approximations or words also produced significantly fewer of them than their TDA peers, but percent consonants correct in those words was the same for both groups.

Finally, Plumb & Wetherby (2013) compared 50 toddlers with ASD, 25 with non-ASD developmental delay (DD), and 50 typically-developing (TD) toddlers between the ages of 18 and 24 months. There were no differences between language scores for the ASD and DD groups, but the ASD group's verbal developmental quotient on the Mullen Early Learning Scales (Mullen, 1985) was significantly lower than that of the TD group (because the developmental quotient takes into account age, and the ASD group was approximately 3 months older). Compared to the TD group, toddlers with ASD produced a lower percentage of speechlike vocalizations and a higher percentage of nonspeechlike vocalizations. There were no differences between the ASD and DD groups on these measures, however, and the speechlike vocalizations did not differ in syllabic complexity between the ASD and TD groups.
Section Four: Theories of Speech Development in ASD

In summary, several findings have emerged regarding speech development in toddlers with or at risk for ASD, even when language is within normal limits. High-risk toddlers produce fewer speechlike vocalizations per unit time, have a lower proportion of canonical syllables, and show at least transient delays in consonant acquisition. Two mechanisms have been proposed to explain these findings.

Subsection One: Social Feedback Loop

The first theory is called the social feedback loop, and it comes from a research tradition documenting the impact of environmental input, especially that of caregiver feedback during interactions, on children’s spoken language development. Socially contingent verbal feedback from adults has been shown to affect infants’ phoneme perception (Kuhl et al., 2003) and production (Goldstein & Schwade, 2008). When infant vocalizations contain consonants, caregivers spontaneously provide positively reinforcing, language-promoting responses more often than when the vocalizations do not contain consonants (Gros-Louis et al., 2006; Talbott et al. (in prep)). Further, infants who receive adult feedback that is contingent on their own babbles restructure their babbles to include phonological patterns from their caregivers’ speech. Infants who receive non-contingent feedback do not show this effect (Goldstein & Schwade, 2008).

Thus, adults’ contingent responses to infants’ early vocalizations form part of a social feedback loop (Warlaumont et al., 2014) that is active in speech acquisition.
In this loop, a speechlike vocalization from an infant is more likely than a nonspeechlike vocalization to receive a contingent, speechlike, language-promoting response from the adults in that infant’s environment. The adult’s speechlike response is in turn more likely to receive a contingent speechlike vocalization from the infant, and so on, in a virtuous cycle.

Warlaumont et al. (2014) propose that the social feedback loop can be disordered in ASD in three ways. First, because of their lower vocalization rate, infants and toddlers with ASD simply have fewer opportunities to engage in these language-promoting exchanges. Second, adults’ responses to infants with ASD may be less contingent on the infants’ vocalizations being speechlike, possibly because of a less social quality to those vocalizations, or because of the presence of vocal or prosodic abnormalities. Third, children with ASD may be less able to make use of adult feedback because of auditory perceptual difficulties or social impairments. However, the findings of Talbott et al. (in prep) demonstrate that mothers of high-risk children do contingently reinforce their infants’ speechlike vocalizations at the same rates as do mothers of low-risk infants. Related work by Yoder and Stone (2006) shows that interventions to specifically train parents to respond positively to their high-risk infants’ attempts at communicative vocalizations do in fact promote spoken language acquisition. Thus, it is unlikely that the social feedback loop is diminished by parents providing less or lower-quality feedback to their high-risk children.
Subsection Two: Speech Attunement Framework

A second framework in which to understand the findings of delay in speech acquisition in toddlers with ASD is called the speech attunement framework. It comes from a large body of work concerning the acquisition of speech in both normal and disordered populations and explaining the factors that make an individual child with speech delay more or less likely to respond to treatment. This framework posits that phonological learning takes place via a combination of an individual child’s capability for learning and his or her focus on the process (Kwiatkowski & Shriberg, 1993).

In the speech attunement framework, capability includes both linguistic abilities (such as phonological comprehension or production) and risk factors (structural or functional constraints like motor speech disorders, hearing impairment, or very low socioeconomic status). Focus includes motivational events that reinforce a child’s desire to change his or her speech (such as external reinforcement or not being understood by the adults in one’s environment) and the amount of effort a child puts into making those changes. In the speech attunement framework, capability defines the potential for stimulability for change and self-monitoring, but some level of focus is also necessary to make use of one’s capacity, whatever its level.

More recently, the speech attunement framework has been applied to ASD (Shriberg et al., 2001; Paul et al., 2007; Paul et al., 2011). Under this view, while the capability of children with ASD with normal language to learn their native languages
is (by definition) intact, their focus on the nuances of speech production in their ambient language environment is diminished. The first claim is bolstered by evidence that at least some individuals with ASD have average auditory (Čeponienè et al., 2003; Lepistö et al., 2006; see also the review in Haessen et al., 2011) and speech-perceptual (Bonnel et al., 2010; Mottron et al., 2006) abilities and average or above-average language ability (Kjelgaard and Tager-Flusberg, 2001). These allow them to “tune in” appropriately to the acoustic-perceptual features of speech. However, because of the challenges in social reciprocity experienced by individuals with ASD, they are less socially motivated to “tune up” the precision of their articulation or prosody to level of their non-ASD peers (Shriberg et al., 2001). Failure to ‘tune up’ results in the phonological delays documented in Paul et al. (2011) and others, as well as in a thirty-fold higher prevalence of residual speech sound disorders in adolescents with ASD (Shriberg et al., 2001).

Subsection Three: Comparison of Predictions

In this section, we compare and contrast the predictions made by the social feedback loop and the speech attunement framework about speech development in toddlers who develop ASD. This will lay the groundwork for the hypotheses to be tested, which are discussed in Section Four.

First of all, both theories predict some level of speech impairment or delay in toddlers with ASD with average language ability. The social feedback loop predicts
that this should arise as a consequence of reduced vocalization rate. The speech attunement framework predicts that delays or impairments arise because of lower “tuning up” ability, making no predictions regarding vocalization rate.

To the extent that speech delays in ASD are related to diminished ability to benefit from adult feedback, the social feedback loop predicts that any differences in the developing speech of toddlers with ASD will be large enough to be perceptually detectable. The speech attunement framework, however, predicts that both perceptually evident and sub-perceptual differences will be found in the speech of toddlers with ASD. Further, the speech attunement framework predicts that these differences will be found even when there is semantic pressure to exploit phonetic differences to convey meaning.

Neither theory predicts that differences will be found in the speech of high-risk siblings who do not develop ASD.

**Section Five: Hypotheses**

The predictions outlined above will be tested in this study. Using perceptual methods, we will attempt to replicate the findings of reduced speechlike vocalization rate, lower canonical syllable production rate, and delays in consonant acquisition in toddlers at high risk for ASD between the ages of 12 and 24 months; and to compare this group to a group of low-risk toddlers. The replication portion of the perceptual study will test the prediction of both theories that between-group differences in speech production exist.
In terms of consonant development in the second year of life, we will concentrate on the acquisition of the Middle-8 (M8) consonants. Paul et al. (2011) found that frequency of M8 consonant production at 6 months of age in babble predicted diagnostic outcome in high-risk siblings. Before 24 months, fewer than 50% of children articulate these consonants correctly in at least two of three word positions (Sander, 1972), so by the middle of the second year, typically-developing toddlers are beginning to use the M8 consonants in words. Thus, production of the M8 consonants will be the focus of the perceptual study, which focuses on toddlers between 12 and 24 months of age.

In the perceptual study, we will also mathematically model the relationship between vocalization rate and differences in consonant production between the three groups of toddlers. Taking the relationship of these variables in the low-risk group as a reference, we will use it to predict the number of consonants in the inventories of the other two groups of toddlers based on their vocalization rates. This procedure allows us to determine whether the relationship between vocalization rate and consonant production is the same across groups. It will test the social feedback loop's prediction that lower vocalization rate is at least partially responsible for the group differences in speech development seen in previous work.

Then, using both perceptual and acoustic methods, we will test the prediction of the speech attunement framework that sub-perceptual differences in consonant production exist in the speech of toddlers who develop ASD, even when language is normal, when there is semantic pressure to correctly produce a meaning-bearing
phonetic difference, when there are no perceptible errors, and when there are no differences in production rate. In particular, we will examine production of the /b/-/p/ distinction in words between the ages of 18 and 36 months, which is when the distinction between voiced and voiceless stop consonants is consolidated in typical development (Mackey and Barton, 1980; Lowenstein and Nittroouer, 2008). Both /b/ and /p/ are in the Early-8 group, meaning that by age 3, at least 90% of typically-developing children produce them correctly in two of three word positions (Sander, 1972). Even speech-delayed three-year-olds produce /b/ and /p/ with over 85% accuracy in words (Shriberg, 1993), so that these two stops should be within the repertoires of all toddlers in the study. The acoustic correlate of the /b/-/p/ distinction that we will use is voice-onset time (VOT), discussed in more detail in Chapter Four. Seminal work by Lisker and Abramson (1964) showed VOT to be the main acoustic correlate of the voiced/voiceless distinction in stops in adult speech.

As mentioned, three groups of toddlers will be compared. Including not only high-risk toddlers who develop ASD (HRA+ toddlers) and low-risk toddlers who do not develop ASD (LRC toddlers) but also high-risk toddlers who do not develop ASD (HRA- toddlers) allows us to begin to isolate the effect of a diagnosis of ASD from the effects of shared genetic risk for speech delay that previous sibling studies have shown to be common in the high-risk population (Chawarska et al., 2014). It is important to note that, while there is considerable overlap in the groups of LRC, HRA-, and HRA+ toddlers who were included in each of the two present studies, the
groups are not identical. This is a result of the fact that not all toddlers came to the lab for all visits: for various reasons, a toddler might miss one or another visit. To perform longitudinal analyses, rather than a series of cross-sectional analyses, toddlers were chosen for the perceptual study who participated in the 12-, 18-, and 24-month visits. For the acoustic study, toddlers who participated in the 18-, 24-, and 36-month visits were chosen. Toddlers who participated in all four visits between 12 and 36 months thus appear in both the perceptual and the acoustic studies.

Once the results of the perceptual and acoustic studies have been reported, they will be discussed in the context of the social feedback loop and speech attunement frameworks. Finally, clinical implications of the results and future avenues of research will be discussed.
CHAPTER THREE: Perceptual Study:
Vocalization Rate and Consonant Development

Section One: Participants

As mentioned, data for the 46 toddlers included in the perceptual study were obtained as part of the ISP, a larger ongoing study of infants at risk for ASD conducted jointly by Boston University and Boston Children's Hospital. The inclusion and exclusion criteria for the ISP were detailed in Section Three of Chapter Two.

To be included in the perceptual study, toddlers needed to participate in the behavioral measures for at least 30 minutes at the 12-, 18-, and 24-month visits. Since vocalization rate was one of the measures used in this study, it was decided that toddlers who cried during a significant portion of the behavioral assessment would not be vocalizing at their usual rate, so 4 toddlers who "fussed out" (i.e., were not able to fully participate in the assessment because of excessive crying or tantrums) were excluded. None of the excluded toddlers received a diagnosis of ASD at 36 months. Of the 46 toddlers who met inclusion criteria for the perceptual study, 18 were at low risk for ASD (low-risk controls, LRC); the other 28 were at high risk for ASD (HRA). Of the 28 HRA toddlers, 10 received diagnoses of ASD at 36 months (HRA+); 18 HRA toddlers did not (HRA-).
Section Two: Methods

Subsection One: Standardized Assessments

Expressive Language (EL) and Receptive Language (RL) scores from the Mullen Scales of Early Development (MSEL; Mullen, 1995) were collected at 12, 18, and 24 months. EL and RL scores for the three groups of participants in the perceptual study are shown in Table 2. At the 12-month visit, the Autism Observation Schedule for Infants (AOSI; Bryson, Zwaigenbaum, McDermott, Rombough, et al., 2007) was administered. At 18 and 24 months, the ADOS was administered. Choice of ADOS module (1, 2, or Toddler) was determined by the child’s level of language at the visit, as per ADOS guidelines. 36-month ADOS score, along with best clinical judgment, was used as diagnostic outcome.

<table>
<thead>
<tr>
<th></th>
<th>Receptive Language</th>
<th></th>
<th>Expressive Language</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12 mo.</td>
<td>18 mo.</td>
<td>24 mo.</td>
<td>12 mo.</td>
</tr>
<tr>
<td>HRA+</td>
<td>40.1</td>
<td>37.1</td>
<td>49.0</td>
<td>41.7</td>
</tr>
<tr>
<td></td>
<td>(11.0)</td>
<td>(12.4)</td>
<td>(12.2)</td>
<td>(5.7)</td>
</tr>
<tr>
<td>HRA-</td>
<td>43.9</td>
<td>50.6</td>
<td>56.4</td>
<td>45.5</td>
</tr>
<tr>
<td></td>
<td>(7.2)</td>
<td>(15.3)</td>
<td>(7.8)</td>
<td>(8.8)</td>
</tr>
<tr>
<td>LRC</td>
<td>45.1</td>
<td>52.2</td>
<td>55.4</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td>(9.1)</td>
<td>(13.6)</td>
<td>(8.5)</td>
<td>(7.8)</td>
</tr>
</tbody>
</table>

Table 2. Mean (s.d.) language scores by age and group (perceptual study). T-score descriptive categories: 20-30 = very low, 31-39 = below average, 40-60 = average, 61-69 = above average, 70-80 = very high.

Subsection Two: Equipment

ADOS or AOSI sessions were recorded in rooms equipped with two Sony SNC-RZ30N cameras and two SHURE SM57 microphones. Audio was extracted from the
video files using AoA Audio Extractor (AoAMedia, 2013), downsamplied to 16 kHz with Audacity (Audacity, 2013), and visualized and played back with Wavesurfer (Sjölander and Beskow, 2011). All three are freely available audio processing software packages.

Subsection Three: Coding

All utterances by the toddler in each 30-minute audio file were classified into one of two categories: Speechlike or Nonspeechlike (defined below). An utterance was defined as including all vocalizations within the same breath group (Oller and Lynch, 1992; Lynch et al., 1995). Utterances that were obscured by another noise or speaker were noted in the corpus, but were not coded or counted.

Speechlike utterances included protovowels, babbles, word approximations, and words. Protovowels were centralized (schwa-like) vocalizations produced with the vocal tract at rest. All other speechlike utterances possessed articulatory movements sufficient to produce the percept of a transcribable English vowel or consonant. They included full, non-centralized vocalic nuclei; multi-vowel or multiphthongal sequences; sequences of vowels separated by glottal stops or [h]; canonical babbles (CV or VC sequences with clear, adult-like consonantal or glide formant transitions); and word approximations or recognizable words. Whispering was counted as speechlike as long as the utterance was intelligible and contained at least one canonical syllable, indicating significant articulatory competence on the part of the child. Whispered vowels were treated as nonspeechlike unless they
sounded like a word (e.g. “uh-oh”).

*Non-speechlike* utterances were vocalizations that did not constitute speech because of abnormal phonation amplitude, source, type, or pitch. Generally, screamed or yelled words were counted as speechlike, despite their intensity or pitch; however, if intensity or pitch made an utterance unintelligible, it was counted as nonspeechlike. Trills and raspberries were produced with a vibratory source like the tongue or lips, not the vocal folds. Ingressive utterances were produced on the inhale, rather than the exhale. Grunts were short sounds of mental or physical effort, bordered at the beginning (and sometimes the end too) by glottal stops. Finally, laugh and cry were vocal but non-language utterances.

In addition to being coded as such, Speechlike utterances were also broadly transcribed. The number of syllables, as well as the number of canonical syllables, in each Speechlike utterance was counted. Consonant inventories for each toddler were assembled from the transcriptions and divided into the Early-, Middle-, and Late-8 categories (Shriberg, 1993; Sander, 1972) for ease of comparison to previous work (e.g., Paul et al., 2011).

*Subsection Four: Coding Reliability*

An additional coder, blind to subject status, was trained to use the coding scheme using video from participants whose data were excluded from the main study. She then independently scored 14 randomly-selected audio files (10% of the total). Inter-rater reliability for identifying utterance codes was 80.1%. Cohen’s $\kappa$ (Cohen,
To assess intra-rater reliability, 12 randomly-selected audio files (9% of the total) were re-coded by the first coder at the end of the study. Average percent agreement between codes on the two dates, approximately 18 months apart, was 80.0%. Cohen’s κ equaled .670, p < .0005.

Subsection Five: Perceptual Analyses

Three categories of analysis were performed: (1) Analyses concerning *vocalization rate* included the total number of utterances, number of Speechlike utterances, and number of Nonspeechlike utterances in 30 minutes. (2) Analyses examining *phonological complexity* included the percentage of canonical syllables; the total number of consonant types; and the number of Early-, Middle-, and Late-8 consonants in each toddlers’ inventory. (3) *Mathematical modeling* of the relationship between vocalization rate and indices of phonological development was performed to investigate the extent to which between-group differences were associated with ASD. All statistical analyses were performed using the SPSS statistical software package (IBM Corp., 2013).

Section Three: Results

Subsection One: Vocalization Rate

Total Utterances in 30 minutes was used as an overall measure of vocalization rate to give some sense of how much speech each toddler generated during the
assessment sessions. A repeated-measures ANOVA on Total Utterances was performed with age (12, 18, 24 months) as a within-subject factor and group (HRA+, HRA-, LRC) as a between-subjects factor. There was a significant main effect of age, $F(2,86) = 18.662$, $p < .0005$, $\eta_p^2 = .303$. Bonferroni-corrected post-hoc analyses showed that toddlers produced significantly more utterances at 24 than 18 months ($p = .002$). There was a significant effect of group as well $F(2,43) = 3.25$, $p = .048$, $\eta_p^2 = .131$. Bonferroni-corrected post-hoc analyses showed that HRA+ produced significantly fewer Total Utterances than HRA-, $p = .05$. There was no significant interaction between age and group on Total Utterances, $F(4,86) = .442$, $p = .778$. Figure 1 shows Total Utterances by age and group.

![Figure 1: Total Vocalization Rate (total utterances in 30’ samples) by Age and Group. Error bars indicate ± one standard error of the mean.](image-url)
A repeated measures ANOVA was also performed on the number of Speechlike utterances in the 30-minute samples with age as a within-subjects factor and group as a between-subjects factor. There was a significant main effect of age, F(2,86) = 16.47, p < .0005, \( \eta^2 = .282 \). Bonferroni-corrected post-hoc analyses showed that participants produced significantly more Speechlike utterances at 18 than 12 months (\( p = .004 \)) and at 24 than 12 months (\( p < .0005 \)). There was also a significant main effect of group, F(2,43) = 4.23, \( p = .021, \eta^2 = .168 \). Bonferroni-corrected post-hoc analyses showed that the HRA+ group produced significantly fewer Speechlike utterances than both the HRA- (\( p = .027 \)) and LRC (\( p = .05 \)) groups. There was no difference in the number of Speechlike utterances produced between the HRA- and LRC groups (\( p = 1.0 \)). There was no significant interaction between age and group (F(4,86) = .87, \( p = .483, \eta^2 = .040 \). Figure 2 shows Speechlike utterance production rate by age and group.

![Speechlike Vocalization Rate by Age and Group](image.png)

**Figure 2.** Speechlike Vocalization Rate by Age and Group. Error bars indicate ± one standard error of the mean.
Finally, a repeated-measures ANOVA on the number of Nonspeechlike utterances in 30 minutes with age as a within-subjects factor and group as a between-subjects factor revealed no main effect of age, $F(2,86) = 0.58$, $p = .587$, $\eta_p^2 = .558$. There was also no main effect of group, $F(2,43) = .08$, $p = .925$, $\eta_p^2 = .004$, and no age x group interaction, $F(4,86) = .83$, $p = .512$, $\eta_p^2 = .037$.

Figure 3. Nonspeechlike Vocalization Rate by Age and Group. Error bars indicate ± one standard error of the mean.

Subsection Two: Phonological Complexity

Comparing the percent canonical syllables produced by each group required, first, an arcsine transformation of the percentage variable to make it normally distributed. A repeated measures ANOVA on the arcsine-transformed %Canonical Syllables revealed a significant main effect of age only, $F(2,84) = 21.79$, $p < .0005$, $\eta_p^2 = .342$. Bonferroni-corrected post-hoc analyses showed that all pairwise age
comparisons were significant at $p \leq .0009$, meaning that all participants produced a significantly greater %Canonical Syllables with age. There was no main effect of group on %Canonical Syllables ($F(2,42) = 1.15, p = .325, \eta^2_p = .052$) and no significant age x group interaction $F(4,84) = .72, p = .584, \eta^2_p = .033$. Table 3 shows the canonical babble data.

<table>
<thead>
<tr>
<th></th>
<th>12 mo.</th>
<th>18 mo.</th>
<th>24 mo.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRA+</td>
<td>50 (25.3)</td>
<td>66 (25.6)</td>
<td>81 (6.2)</td>
</tr>
<tr>
<td>HRA-</td>
<td>62 (18.2)</td>
<td>73 (14.5)</td>
<td>79 (13.1)</td>
</tr>
<tr>
<td>LRC</td>
<td>67 (11.6)</td>
<td>73 (16.7)</td>
<td>85 (12.7)</td>
</tr>
</tbody>
</table>

Table 3. Canonical Babble Ratios by Age and Group. Mean (s.d.).

A repeated measures ANOVA was performed on the Number of Consonant Types in each toddler's inventory with age as a within-subject factor and group as a between-subjects factor. Results showed a significant main effect of age, $F(2,86) = 56.415, p < .0005, \eta^2_p = .567$. Bonferroni-corrected post-hoc analyses revealed that all pairwise comparisons were significant at $p < .0005$, indicating that all children produced more consonant types with increasing age. There was no main effect of group, $F(2,43) = 2.00, p = .147, \eta^2_p = .052$. There was a significant age x group interaction $F(4,86) = 2.563, p = .044, \eta^2_p = .107$, meaning that the rate of increase in Number of Consonant Types differed across groups.

Next, consonant inventories were broken into the Early-, Middle-, and Late-8 categories (listed in Table 2). A repeated measures ANOVA showed a significant
main effect of age on number of E8 consonants, $F(2,86) = 16.97, p < .0005$, $\eta_p^2 = .283$. Post-hoc, Bonferroni-corrected analyses showed a statistically significant increase in E8 consonants across groups from 12 months to 18 months, but not from 18 to 24 months (as expected, given that they appear early in development). There was no main effect of group ($F(2,43) = 1.03, p = .367, \eta_p^2 = .046$) and no significant age x group interaction on number of E8 consonants ($F(4,86) = 1.343, p = .261, \eta_p^2 = .059$).

For number of M8 consonants, a repeated measures ANOVA also showed a significant main effect of age, $F(2,86) = 66.860, p < .0005, \eta_p^2 = .609$. Post-hoc, Bonferroni-corrected analyses showed all pairwise age comparisons to be significant at $p \leq .001$, indicating that all three groups of toddlers added more M8 consonants to their repertoires with increasing age. In addition, there was a significant effect of group for the number of M8 consonants, $F(2,43) = 3.65, p = .034, \eta_p^2 = .145$. Overall, HRA+ children had significantly fewer M8 consonants in their repertoires than LRC children ($p = .03$). The age x group interaction was also significant, $F(4,86) = 4.38, p = .003, \eta_p^2 = .169$, indicating different developmental trajectories between groups. At 12 months, it was HRA- toddlers who produced the lowest number of M8 consonants; but at 18 months, HRA- produced the largest number of M8 consonants. At 24 months, there was no difference in the number of M8 consonants between the HRA- and LRC groups.

Finally, for number of L8 consonants, a repeated measures ANOVA showed a significant main effect of age, $F(2,86) = 34.89, p < .0005, \eta_p^2 = .448$. Post-hoc,
Bonferroni-corrected analyses showed that all pairwise age comparisons were significant at $p \leq .003$, indicating that all three groups of toddlers produced more of the L8 consonants with increasing age. There was no effect of group ($F(2,43) = 1.21$, $p = .308$, $\eta_p^2 = .053$) and no significant age x group interaction ($F(4,86) = 1.46$, $p = .222$, $\eta_p^2 = .063$) on number of L8 consonants in toddlers’ repertoires.

<table>
<thead>
<tr>
<th></th>
<th>Early-8</th>
<th></th>
<th>Middle-8</th>
<th></th>
<th>Late-8</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12 mo.</td>
<td>18 mo.</td>
<td>24 mo.</td>
<td>12 mo.</td>
<td>18 mo.</td>
<td>24 mo.</td>
</tr>
<tr>
<td>HRA+</td>
<td>5.2</td>
<td>6.4</td>
<td>6.4</td>
<td>2.2</td>
<td>2.7*</td>
<td>5.1</td>
</tr>
<tr>
<td>(1.9)</td>
<td>(2.2)</td>
<td>(2.4)</td>
<td>(1.4)</td>
<td>(2.1)</td>
<td>(2.1)</td>
<td></td>
</tr>
<tr>
<td>HRA-</td>
<td>5.3</td>
<td>6.7</td>
<td>7.3</td>
<td>1.8**</td>
<td>4.6</td>
<td>6.3</td>
</tr>
<tr>
<td>(1.4)</td>
<td>(0.8)</td>
<td>(0.5)</td>
<td>(1.4)</td>
<td>(1.3)</td>
<td>(1.1)</td>
<td></td>
</tr>
<tr>
<td>LRC</td>
<td>6.0</td>
<td>6.6</td>
<td>6.9</td>
<td>3.4</td>
<td>3.9</td>
<td>6.1</td>
</tr>
<tr>
<td>(1.3)</td>
<td>(0.9)</td>
<td>(0.6)</td>
<td>(1.6)</td>
<td>(1.8)</td>
<td>(1.4)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Consonant Inventories by Age, Group. Mean (s.d.). * $p = .019$ re HRA-, ** $p = .005$ re LRC

*Subsection Three: Modeling the Relationship Between Speechlike Vocalization Rate and Number of Middle-8 Consonants*

A mathematical model is a description of a system using numerical relationships. In this case, the relationship between Speechlike vocalization rate and number of M8 consonants for the LRC group was modeled using a logarithmic equation. This equation was then used to assess whether the number of M8 consonants that the HRA+ and HRA- groups produced was statistically similar to or different from the
number the LRC group produced, based on vocalization rate.

For ages at which group differences in inventory size appeared (12 and 18 months), a variety of curves was first fit to the plot of each group’s Speechlike vocalizations vs number of M8 consonants data using SPSS. The mean absolute error for each type of curve (e.g., sigmoid, log, linear) was calculated for each group at each age. A one-way ANOVA was performed to compare the mean absolute error between curve types, to determine which curve produced the least error. In all cases, the best-fit curve was a log curve. Figures 4 and 5 show the best-fit curves for each group at 12 and 18 months, respectively.

Figure 4. Best-Fit Curves for 12-month Speechlike Vocalization Rate vs. M8 Inventory Size, by Group.
Figure 5: Best-Fit Curves for 18-Month Speechlike Vocalization Rate vs. M8 Inventory Size, by Group.

Then, to determine whether the log curves best describing the relationship between each group's Speechlike vocalization rate and number of M8 consonants were significantly different from each other, a prediction procedure was used. Taking the LRC group as the reference, the number of Speechlike vocalizations for each other group in turn was entered into the equation describing the LRC curve, yielding a number of predicted consonants. A t-test was performed comparing the mean number of Predicted consonants to the mean number of Actual consonants for each group. For the 12-month data, the mean number of M8 consonants predicted for the HRA+ group was 3.04 (s.d. = 0.50). A t-test revealed that this number was not significantly different from the actual mean number of M8 consonants for HRA+ (2.20, s.d. = 1.33), p = .1021. The mean number of M8 consonants predicted for the
HRA- group at 12 months, however, was 3.20 (s.d. = 1.49); this was significantly different from the actual mean number of M8 consonants for HRA- (1.65, s.d. = 1.23), $p = 0.003$. For the 18-month data, the mean number of M8 consonants predicted for the HRA+ group by the LRC group’s 18-month equation was 2.80 (s.d. = 0.69) and the actual number of M8 consonants was 2.70 (s.d. = 1.95). These were not significantly different ($p = .8660$). For the HRA- group, the mean number of predicted consonants was 3.32 (s.d. = 0.44) and the actual number of consonants was 4.59 (s.d. = 1.33), $p = .001$. Tables 5 and 6 show the predicted and actual numbers of consonants for 12 and 18 months, respectively.

<table>
<thead>
<tr>
<th></th>
<th>predicted # M8 consonants</th>
<th>actual # M8 consonants</th>
<th>mean absolute error</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRA+</td>
<td>3.04 (0.50)</td>
<td>2.20 (1.33)</td>
<td>1.28</td>
</tr>
<tr>
<td>HRA-</td>
<td>3.20 (1.49)</td>
<td>1.65 (1.23)*</td>
<td>1.60</td>
</tr>
<tr>
<td>LRC</td>
<td>3.38 (1.19)</td>
<td>3.39 (1.53)</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Table 5. Predicted vs. Actual 12-Month Middle 8 Inventories, by Group. Mean (s.d.). *$p=.003$.

<table>
<thead>
<tr>
<th></th>
<th>predicted # M8 consonants</th>
<th>actual # M8 consonants</th>
<th>mean absolute error</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRA+</td>
<td>2.80 (0.69)</td>
<td>2.70 (1.95)</td>
<td>1.33</td>
</tr>
<tr>
<td>HRA-</td>
<td>3.32 (0.44)</td>
<td>4.59 (1.33)*</td>
<td>1.40</td>
</tr>
<tr>
<td>LRC</td>
<td>3.14 (0.05)</td>
<td>3.94 (1.78)</td>
<td>1.50</td>
</tr>
</tbody>
</table>

Table 6. Predicted vs. Actual 18-Month Middle 8 Inventories, by Group. Mean (s.d.). *$p=.001$. 


**Section Four: Discussion: Vocalization Rate and Consonant Development**

In general, the results show few differences in speech development between the HRA+ and HRA- groups, as compared to LRC. All three groups vocalized more, produced more speechlike vocalizations, produced a larger % canonical babbles, and acquired more consonants with age. This is not surprising, given that all groups’ language scores were within the average range (within 1 standard deviation of the mean) on the MSEL, the one exception being the HRA+ group’s Receptive Language score at 18 months. However, the fact that this group’s Expressive Language score at the same date was within the average range suggests that the Receptive Language score was not representative of this group’s true abilities. The finding of lower Receptive than Expressive Language scores for children with ASD is common (Paul et al., 2011) and generally ascribed to differing task demands on the two tests and to low social motivation in children with ASD. Overall, the current results are consistent with those of Paul et al. (2011) and Schoen et al. (2011) in that they demonstrate that the children who go on to receive diagnoses of ASD are developing phonologically on par with their general language abilities.

Significant between-group differences were found in two main areas: rate of production of Speechlike vocalizations and size of M8 consonant inventory. The HRA+ group had a Speechlike vocalization rate significantly lower than both the HRA- and the LRC groups. The HRA- group showed a significantly smaller number of M8 consonants at 12 months than LRC. The HRA+ group showed a significantly smaller number of M8 consonants than the HRA- group at 18 months. These results
are similar to those of Paul et al. (2011), whose HR group had fewer M8 consonants in their babble repertoires than the LR group at 9 months.

The validity of Paul’s assumption that the Early-Middle-Late distinction applies to consonants in babble as well as consonants in words is supported by the current results. The mean change from 12 to 24 months, averaged over all groups in this study, in number of E8 consonants was +1.3 (from 5.5 to 6.8); for M8, +3.3 (from 2.5 to 5.8); and for L8, +2.5 (from 2.0 to 4.6). During the second year of life, then, the M8 consonant inventory size increased the most. This is consistent with the idea that all groups had more or less mastered the E8 consonants in words, were in the process of mastering the M8 consonants in words, and had not yet begun to master the L8 consonants in words. Looking at the same phenomenon from another angle, the proportion of variance in consonant inventory size explained by age ($\eta^2_p$) was .283 for E8, .609 for M8, and .448 for L8. Age thus explained the most variance in the M8 inventory size between 12 and 24 months, compared to E8 and L8.

The homogeneity of the participants with respect to consonant development throws the differences between groups into sharp relief. The questions investigated were whether the HRA+ group had a lower vocalization rate than the others, whether they showed delays in acquisition of canonical babble or consonant inventories, and whether any delays could be reasonably ascribed to a lower vocalization rate. When vocalization rate was limited just to Speechlike ones, the HRA+ group did have a significantly lower rate than both other groups. The HRA+ group also showed transient delays in M8 consonant acquisition (significantly fewer
at 18 months), but no differences in canonical babble production.

The results also showed that consonant production is related to (speechlike) vocalization rate for all participants. There was a positive concurrent relationship between rate of speechlike vocalization and size of M8 consonant inventory at 12 and 18 months for all three groups. This is not surprising, because producing more speechlike vocalizations per unit time increases the chances of showing all of the consonants in one’s inventory. Also, the HRA+ group did not have a significantly different M8 inventory size at 12 months, despite a lower speechlike vocalization rate. At the same time, the HRA- group did show a significantly smaller M8 inventory at 12 months, despite a vocalization rate equal to that of the LRC group. There is thus a double dissociation between Speechlike vocalization rate and M8 inventory delays.

In order to understand whether the relationship between Speechlike vocalization rate and M8 consonant inventory size differed between groups, log curves were found that best fit the relationship between the two variables for each group. Then, using the LRC group’s equation as a reference, a predicted inventory size was found for the HRA+ group and compared to the actual inventory size. The difference between predicted and actual M8 inventory size at 12 months was not significant for the HRA+ group. This suggests that the relationship between speechlike vocalization rate and M8 consonant inventory size is the same for typically-developing toddlers and those who develop ASD.

The converse situation was found at 18 months. At this age, the HRA+ group
showed a significantly smaller M8 inventory size than the HRA- group, along with a lower Speechlike vocalization rate than the other two groups. However, the number of M8 consonants predicted from the HRA+ group’s Speechlike vocalization rate at this age was not significantly different from the actual number. Again, this result supports the idea that there is a relationship between Speechlike vocalization rate and consonant inventory size – but that it is not different in ASD.

The results quantifying the HRA- group’s M8 consonant inventories at 12 and 18 months shed further light on the relationship between vocalization rate and consonant acquisition. At no age was the Speechlike vocalization rate for this group significantly different from the LRC group’s. However, at 12 months, the HRA-group was predicted to have, on average, 3.2 M8 consonants. Instead, they had an average of only half that number, a significant deficit. This finding strongly suggests that there is some other factor above and beyond vocalization rate that accounts for the HRA- group’s smaller in inventory size at 12 months. The 18-month results further support the idea of a factor unrelated to vocalization rate accounting for the differences in HRA- inventory size. At this age, the mean predicted number of consonants for HRA- was 3.32, compared to a mean actual number of consonants of 4.59. Some factor other than vocalization rate or ASD must be responsible for the significantly larger M8 inventory in HRA- compared to LRC at 18 months.

The speech attunement framework provides such a factor. ‘Tuning up’ is predicted to be intact in the HRA- group, and is plausibly responsible for their ability to overcome early delays and achieve an M8 consonant inventory size.
indistinguishable from LRC at 18 and 24 months. The present results do not, however, provide direct evidence for a lack of ability to ‘tune up’ in the HRA+ group. Still, the lack of ‘tuning up’ is not always predicted to be great enough to result in perceptually-detectable speech delays or errors. Thus, an investigation into speech production using finer-grained methods is needed.
CHAPTER FOUR: Acoustic Study: The Voiced/Voiceless Distinction

Section One: Voice Onset Time Development and Analysis

Subsection One: Acoustic vs. Perceptual Analyses

Studies that have looked at phonological development in toddlers with ASD have almost exclusively employed perceptual judgments. Perceptual measures of phonological performance are useful and ecologically valid, but simple ratings of “correct”, “incorrect”, or “distorted” for consonants (for example) do not tell the whole story of how toddlers learn to pronounce their native languages. This limitation arises because adult perception of consonants is categorical, meaning that acoustic differences between tokens within the same consonant category are difficult or impossible for adults to hear.

When some within-category acoustic differences between tokens are large enough, adults may perceive the phonemes as distorted, but other within-category acoustic differences may go unnoticed. Thus, perceptual judgments are sufficient for situations where acoustic differences between phoneme tokens are large enough to be detected by ear, but not for situations where even consistent acoustic differences between tokens are below the reporting threshold. In this vein, Shriberg et al. (2001) called perceptual measures “generally adequate for summative levels of description but inadequate in some domains for the level of precision needed in descriptive-explanatory work.” (p. 1111) Acoustic measures are required, therefore, to investigate the subphonemic aspects of development of phonetic
contrasts, especially in a group of toddlers hypothesized to show subtle deficits in the ability to 'tune up' to typical levels of articulation.

Subsection Two: Phonetic Contrast

Like the choice of analysis method, the choice of phonetic contrast to be used to investigate intact or impaired ability to ‘tune up’ bears careful consideration. First of all, the contrast should be meaning-bearing, not just allophonic. An example of a meaning-bearing contrast is the difference in place of articulation of stop consonants. “Dot” and “got”, for example, are two different words because their initial consonants differ in place of articulation (alveolar vs velar). An allophonic difference, on the other hand, refers to the way that a phoneme is articulated in a particular context in a particular dialect. For example, in American English, [t] is produced as an unvoiced, aspirated alveolar stop in syllable-initial position (e.g., “top”); but in final position it is often produced as a glottal stop (e.g., “pot”). It is not incorrect to aspirate the /t/ at the end of a word like “pot”, but not customary in American English, either. A non-meaning-bearing allophonic difference in production between speakers with and without ASD would support the speech attunement framework, but a stronger prediction is that, even in cases where there is semantic pressure to produce a phonetic contrast accurately, language-normal toddlers who later receive diagnoses of ASD will show reduced accuracy at a sub-perceptual level. One potential place where this may show up is the contrast
between voiced and unvoiced minimal pair phonemes.

The choice of phonetic contrast used to test this theory must also fulfil other criteria. It should be an early-developing contrast, so that both members of the minimal pair will be present in the early words of toddlers and so that most toddlers can be assumed to have mastered the contrast before the age of 36 months. The typical developmental progression of the contrast should also be well characterized. Finally, the perceptual difference between members of the minimal pair should be closely linked to an acoustic difference that is relatively easy to measure.

The voiced/voiceless distinction in syllable-initial bilabial stops fulfils all of these criteria. Sander (1972) reports that approximately 93% of typical 2-year-olds produced /b/ correctly in syllable-initial position, and 87% produced initial /p/ correctly. Even the speech-delayed 3-6-year-olds in Shriberg’s (1993) work produced /b/ with over 90% accuracy in words and /p/ at an average of 82% correct in words. Macken and Barton (1980) and Lowenstein and Nittrouer (2008) have detailed the acquisition process for bilabial stops (among others), and Lisker and Abramson’s (1964) work identifies VOT as the main acoustic correlate of the perception of voicing for stops.

Subsection Three: Stop Consonant Production

To produce a stop consonant, several glottal and supraglottal events must occur in sequence and with precise coordination. Since this work concerns /b/ and /p/, the
following discussion will focus on bilabial stops. First, airflow through the vocal tract is stopped by a closure at the lips. In adults, this is accomplished by raising both the mandible and the lower lip (Green et al., 2000); the upper lip contributes slightly, but not substantially, to this process. The rate of closure of the bilabial constriction has its own timecourse, with a maximum just before the actual closure, and is in the range of 40 to 100 cm$^2$/second (Stevens, 1998). Because airflow continues through the glottis during and after the labial closure, air pressure builds in the vocal tract, causing the vocal tract walls to bulge outward slightly and subjecting the vocal folds to an abducting force. Glottal vibration is inhibited if the closure lasts longer than approximately 50 msec.

Following release of the bilabial closure, rate of increase of the constriction is quite high (100 cm$^2$/sec), and the correspondingly rapid increase in airflow gives rise to a wideband transient of sound energy, called a “release burst”. Following the burst, rapid airflow through the widening constriction causes turbulent airflow, giving rise to frication noise at the lips. Simultaneously, and in response to the decreasing abducting force on the vocal folds, the glottis begins to narrow. When the glottal opening becomes narrower than the labial opening, frication noise from the lips is replaced by aspiration noise at the glottis. Soon after this, as the vocal folds continue to adduct, glottal vibration resumes. The amount of time between labial release and glottal vibration varies depending on the voicing status of the bilabial consonant, and ranges from 0 to 20 msec for American English voiced stops (/b, d, g/) to 60 msec or more for unvoiced aspirated stops (/p, t, k/) (Stevens,
Several acoustic features combine to create the percept of a voiced or voiceless stop (Lieberman and Blumstein, 1988): formant transitions, burst amplitude, aspiration amplitude are three. However, the main acoustic correlate of the percept of voicing in stop consonants is called voice onset time (VOT), and was identified by Lisker and Abramson (1964). VOT is defined as the interval of time between oral release of a stop consonant (generally indicated by the time of the burst) and the onset of voicing for the following vowel.

The child’s ability to produce the oral movements for stop consonants develops over time and has been documented for children aged one, two, and six years by Green et al. (2000). At 12 months, the jaw is the largest contributor to oral closure for bilabial stops and nasals. At this age, closures are accomplished by the use of rapid, ballistic movements that exhibit high velocity, acceleration, and deceleration. By the age of 24 months, children’s lip and jaw movements become more closely coupled in both space and time. A smaller increase in spatiotemporal coupling is seen between upper and lower lip movements. In fact, the contribution of the upper lip to labial closure increases from one to two years of age, before decreasing to near-adult levels by age six. During the same time interval, the relative contribution of the jaw to labial closure also decreases slightly, while the lower lip’s contribution increases.

In work that complements that of Green et al. (2000), Koenig (2000) found that distributional measures of VOT in /p/ correlated with distributional measures of voicing in /h/ in children and adults. That is, the children with the longest
intervals between peak airflow for /h/ and the onset of voicing for the following vowel also produced the longest VOTs for /p/. Koenig interpreted this to mean that learning to produce unvoiced, aspirated stop consonants correctly in words involves learning the ability to control voicing itself.

Thus, at the same time as the child’s ability to coordinate upper lip, lower lip, and jaw movements for the production of bilabial stops develops, so must her ability to produce accurate laryngeal movements, and to smoothly mesh them with the preceding three. While for voiced stops it is sufficient to release the oral closure and initiate voicing at the same time (Kent’s (1983) “everything moves at once” principle of children’s speech), this is not the case for voiceless stops. According to Koenig (2000), voiceless stops develop later than voiced stops in children’s babbles and words precisely because they require that the oral articulatory gestures be separated from the laryngeal ones. Longitudinal studies from English and other languages confirm that acquiring the voiced/voiceless distinction in stops follows a developmental course of differentiation of gestures, refinement of them, and finally a stage of gestural stabilization (Green et al., 2000; Rvachew and Brousseau-Lapré, 2012). During the stabilization period, production variability decreases with increasing mastery. Because of the precise timing it requires between oral release and glottal abduction, production of voiceless aspirated stops has been regarded as marking a milestone in articulatory development (Koenig, 2000).
Subsection Four: Acquisition of the Voicing Contrast in Bilabial Stops

The stages in acquisition of the voicing contrast in word-initial stops in English were first documented in a longitudinal study by Macken and Barton (1980). Every two weeks, they recorded spontaneous productions of words beginning with voiced and voiceless stops by four toddlers, beginning at 16-18 months of age and ending at 19-23 months. These researchers identified three general stages in acquisition of the voicing contrast. In Stage 1, the “no contrast” stage, there is no measurable acoustic difference in the waveforms of intended voiced and voiceless stops that children produce. Both are produced with short-lag VOTs, meaning that both types of stops are produced with VOT values below 20 ms, consistent with adult values corresponding to voiced stops. The two populations of stops do not differ significantly in terms of VOT in Stage 1. At the beginning of Stage 2, the “covert contrast” stage, the mean VOT for intended voiced stops is still close to 0 ms, but the mean VOT for intended voiceless stops lengthens -- though it remains within the adult “voiced” category, so the two still do not sound different to adults. Later in Stage 2, intended voiceless stops begin to be produced with VOTs in a range that is perceptually ambiguous to adults (40-55 ms). Stage 3 also has two subparts. In the first part of Stage 3, the “overshoot” stage, toddlers produce intended voiceless stops with VOTs in a range that finally sounds unvoiced to adults, but now the mean VOT for their voiceless stops is much longer than the adult mean, around 100 ms. Finally, in the second half of Stage 3, toddlers “retract” their voiceless stop VOT values back to a value closer to, though still longer than, the mean adult value.
Other groups have also examined VOT changes in typical development. Lowenstein and Nittrouer (2008) largely replicated the findings of Macken and Barton (1980) with a group of seven toddlers, taped at two-month intervals between the ages of 14 and 31 months. They showed that toddlers’ mean voiced VOTs started at 11 ms and stabilized at 16 ms by 21-22 months. Mean voiceless VOTs, on the other hand, were consistent at approximately 40 ms until the age of 23-24 months, when they increased to over 60 ms. No evidence for retraction was found in this study, though it had been observed for older ages in a previous study (Nittrouer, 1993). In addition, Lowenstein and Nittrouer found that the variability in VOT for the toddlers’ voiced stops was lower than for unvoiced stops, and it decreased slightly throughout the study, suggesting improved production accuracy. The variability in VOT for voiceless stops, by contrast, was significantly higher than for voiced stops, suggesting less accuracy in production. More specifically, Koenig (2000) found that the coefficient of variation (a normalized version of the standard deviation equal to the standard deviation divided by the mean) was larger for children’s unvoiced VOTs than for adults’, indicating that the increased variability in unvoiced VOT is not simply due to the larger mean. Lowenstein and Nittrouer (2008) take their results to indicate that, while toddlers’ target VOTs for voiceless stops became more adultlike, their skill in producing those target values did not improve over the course of the study.
Subsection Five: Assessment of Production Accuracy

Precision of timing between articulatory gestures is what gives rise in adult speech to the structure of phonemes. In comparison to the speech of adults, acoustic measures of children’s speech show more variability in interarticulator timing (Nittrouer, 1993; Lowenstein and Nittrouer, 2008; Macken and Barton, 1980), and this increased variability is generally assumed to reflect lower production accuracy. However, increased variability in production is not always pathological; in some cases, too little variability indicates a problem. For example, McGowan et al. (2008) examined the spontaneous babbles of 10 one-year-old toddlers with normal hearing and 10 with hearing loss. Though the hearing-impaired toddlers were all using hearing prostheses, they showed consistently less variability in one measure of vowel production (second formant frequency, or F2) than toddlers with normal hearing. Specifically, the range of F2 was significantly smaller for the hearing-impaired toddlers than for the normal-hearing group, and the relative deviation was also smaller, meaning that the front-back tongue movements for the hearing-impaired group were less likely to be distant from the mid-position than for the children with normal hearing (there were no differences in tongue height between groups). Thus, variability does not necessarily always index production accuracy.

Furthermore, when investigating the development of a phonetic contrast, variability alone fails to capture some crucial details. This limitation arises because a phonetic contrast involves being able to produce two similar, but distinct, populations of phoneme tokens. As toddlers learn to produce these distinct
populations, we might expect a temporary increase in variability in the dimension that separates the two populations, and the subsequent creation of a bimodal distribution with respect to the acoustic contrast between the phonemes. This, in fact, is what Macken and Barton (1980) and Lowenstein and Nittouer (2008) found as toddlers developed the ability to produce unvoiced stops in words. Looking at overall VOT, both groups found not only an increase in mean VOT, but an increase in the standard deviation. When stops were separated by voicing, the mean VOT for voiced stops remained constant with age, as did its standard deviation. For unvoiced stops, both the mean and standard deviation of VOT increased for a time, then decreased slightly to reach adult levels in late childhood. Thus, an increase in standard deviation of VOT, at least for unvoiced stops, is part of the normal developmental progression and may not distinguish toddlers with ASD from those without ASD. Instead of variability, we need a metric that will assess how consistently toddlers are able to produce a phonetic distinction, not just a certain phoneme type. One potential metric is the $t$ statistic, which is used to determine whether two groups have different means. Another is Cohen’s $d$ (Cohen, 1992), which measures the standardized difference between two means.

**Section Two: Participants**

Data for the 55 toddlers included in the acoustic study were obtained as part of the larger ISP conducted jointly by Boston University and Boston Children’s Hospital, whose details were discussed in Chapter Three, Section One. No hearing deficits
were present in the participants for this study, as determined from neonatal hearing screening and from evoked-response potentials collected for other parts of the larger study. Furthermore, all toddlers in the acoustic study were monolingual speakers of English. Of the 55 toddlers, 11 were HRA+, 22 were HRA-, and 22 were LRC.

**Section Three: Methods**

*Subsection One: Procedures*

Audio from the first 30 minutes of the ADOS, this time from the 18-, 24-, and 36-month visits, was used as a spontaneous speech sample. Words in the 30-minute samples containing syllable-initial bilabial stops were extracted, yielding a total sample of 5,389 stops. Both words (i.e., utterances that closely matched the adult form, such as [bʌbo] for ‘bubble’) and word approximations (i.e., utterances that less closely matched the adult form, such as [bʊn] for ‘balloon’) were included in the sample. Each utterance was glossed according to the intended word and broadly phonetically transcribed. Words were glossed based on the conversational context; that is, the parent’s or examiner’s repetition after the child’s utterance, or the toy or activity the child was referring to.

*Subsection Two: Analyses*

Language ability was again assessed using the Receptive and Expressive Language subtests of the Mullen Scales of Early Learning (MSEL; Mullen, 1995).
Several counts were made of the words in each session for each participant. First, the number of /b/ and /p/ tokens in words in the 30-minute samples was counted. Words were counted as containing a /b/ or a /p/ depending on the intended word; that is, what the stop would be in the adult version of the word. Thus, the word ‘pop’ counted as a /p/ even if it was pronounced [bap]. Also included in the count of /b/ and /p/ tokens were words that were intelligible but obscured by noise from a toy or another speaker, as were intelligible but whispered words.

Some words’ VOTs could not be measured because of a speech error, such as stops with no oral closure or stops that were fricated, nasalized, or voiced all the way through. Utterances for which the intended word could not be discerned (for example, [bæp] or [gæbwʊd]) were also counted as errors (“word unknown”). The total number of errors was tallied for each child.

VOT was measured for each syllable-initial bilabial stop in identifiable, unobscured, non-whispered words without one of the error types listed above. Stops in clusters were not measured. Markers were placed at (1) the beginning of regular, periodic glottal vibrations marking the voice onset and (2) the broadband, aperiodic burst marking the oral release. Both waveforms and spectrograms were used to select and confirm marker placement. VOT was then calculated as the time interval between the two markers. Several examples follow.

The first example (Figure 6) shows a spectrogram (bottom window) and waveform (top window) of the utterance “a big one”. The left cursor is placed on the
release burst of the /b/, which is visible as a wideband spectral element that is very brief in time (approximate time 18:38.074). The right cursor is placed at the onset of voicing for the following vowel, which is indicated by the first of a series of evenly-spaced positive peaks in the waveform (approximate time 18:38.106). The difference in time between the two cursor marks represents the VOT for this token, .032 seconds.

As mentioned, some tokens contained stops with a VOT of zero because the stops were voiced all the way through. Figure 7 shows an example of this. The word is “baby”, and the burst for the first /b/ can be seen just before 18:44.4. The right cursor marks the location of the burst for the second /b/. However, whereas there is a small gap between the first burst (located between 18:44.3 and 18:44.4) and the onset of voicing for the first /b/ (indicated by the wide dark bar below 2 kHz, beginning just after the burst), there is no such gap for the second /b/. The voicing is continuous, yielding a VOT of zero.

**Figure 6: VOT measurement for /b/ in “a big one”**
Still other stops were prevoiced, meaning that the onset of voicing preceded, rather than followed, the release burst and that the VOT was negative. Figure 8 shows an example of a prevoiced stop.

In Figure 8, the left cursor is placed at the onset of voicing, which is indicated by the voicing bar that begins at 27:35.531. The right cursor is placed at the wideband transient indicating the release burst at 27:35.459. Since VOT is defined as voicing
onset time minus burst time, and since the burst is later than the voicing onset, the VOT (-0.072 seconds) is negative in this case.

Finally, Figure 9 shows an example of an unvoiced stop, /p/ in “open it”. Again, the left cursor is placed at the location of the release burst at 24:41.592 and the right cursor at the onset of voicing (as determined by the onset of periodicity in the time waveform, not shown) at 24:41.697. The difference between these times is 0.105 seconds, considerably longer than that for the /b/ in Figure 6.

![Figure 9: VOT measurement for /p/ in “open it”.](image)

**Subsection Three: Measurement Reliability**

An additional coder, blind to subject status, was trained to measure VOT. She then independently measured 11 randomly-selected audio files (10% of the total). Results showed a mean VOT of 18.6 msec (SD 36.0) for /b/ for the second judge, compared to a mean of 20.2 msec (SD 32.3) for the first judge, a difference of 8.1%. A t-test revealed no significant difference between the two sets of measurements
The second judge’s mean VOT for /p/ was 60.4 msec (SD 63.0) as compared to 55.6 (SD 38.4) for the first judge, a difference of 8.0%, \( p = 0.5829 \).

**Subsection Four: Analyses**

The mean VOT for voiced and voiceless stops was calculated for each toddler at each age. The within-child standard deviation (WCSD) for VOT was also calculated for voiced and unvoiced stops, for toddlers who produced at least three tokens of each category at each age. Note that some toddlers produced no /b/ or /p/ tokens at certain testing sessions. When appropriate, cross-sectional group analyses were used in full factorial ANOVAs. Otherwise, repeated-measures ANOVAs including only the toddlers who had the relevant data at all three ages were used. As in the perceptual study, statistical analyses were performed using the SPSS statistical software package (IBM Corp., 2013).

Two methods were used to assess the degree of separation of each toddler’s VOT populations for /b/ and /p/. To assess how distinct the VOT populations were for each toddler producing at least three /b/ and three /p/ tokens, a two-tailed t-test was used, with \( \alpha < .05 \) indicating that the populations were statistically separate. The number of toddlers in each group at each age whose voiced and voiceless VOT populations were statistically distinct was then counted and compared using a chi-square analysis.

Also, Cohen’s \( d \) (Cohen, 1992) was calculated for the difference in voiced and
unvoiced stop means. This measure assesses how separate the means of the two VOT populations are, without assuming that the populations are statistically distinct. Then, the number of toddlers in each group whose mean voiced and voiceless VOT populations were associated with $d > .8$ was counted at each age and compared, again using a chi-square analysis.

**Section Four: Results**

*Subsection One: Language Ability*

First, MSEL language scores were examined to verify that all groups were at least in the average range for language level. These are detailed in Table 7.

<table>
<thead>
<tr>
<th></th>
<th>Receptive Language</th>
<th></th>
<th>Expressive Language</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18 mo.</td>
<td>24 mo.</td>
<td>36 mo.</td>
<td>18 mo.</td>
</tr>
<tr>
<td><strong>HRA+ (n=9)</strong></td>
<td>41.1 (17.3)</td>
<td>47.9 (16.6)</td>
<td>48.8 (16.6)</td>
<td>44.7 (11.7)</td>
</tr>
<tr>
<td><strong>HRA- (n=15)</strong></td>
<td>51.3 (13.5)</td>
<td>54.1 (5.8)</td>
<td>53.9 (7.9)</td>
<td>50.5 (10.6)</td>
</tr>
<tr>
<td><strong>LRC (n=19)</strong></td>
<td>59.6 (13.0)</td>
<td>63.1 (6.9)</td>
<td>61.2 (8.3)</td>
<td>51.0 (6.7)</td>
</tr>
</tbody>
</table>

*Table 7. Mean (s.d.) language scores by age and group (acoustic study).*

T-score descriptive categories: 20-30 = very low, 31-39 = below average, 40-60 = average, 61-69 = above average, 70-80 = very high.

It will be noted that, in this sample of toddlers from the ISP, the mean RL and EL scores at each age were not always within the average range. Specifically, the LRC group in this study had slightly above-average RL scores at 24 and 36 months, and
slightly above-average El scores at 36 months. These between-group differences will be taken into account and discussed in Subsection 5 below.

Subsection Two: Token Counts and Error Rates

At 18 months, all but four toddlers produced at least one word with /b/ or one with /p/. This decreased to all but one toddler at 24 months, and at 36 months all toddlers produced at least one word with /b/ or /p/. A 3-way ANOVA with number of stop tokens as the dependent variable and age, group, and voicing status (/b/ or /p/) as between-group factors showed a main effect of age (F(2,274) = 11.94, p < .0005, ηp² = .080) and of voicing (F(1,274) = 50.7, p < .0005, ηp² = .156). There was no main effect of group (F(2,274) = 2.5, p = .084, ηp² = .018). All three groups produced statistically similar numbers of stops, produced more stops with increasing age, and produced more /b/ tokens than /p/ tokens. There were no significant 2- or 3-way interactions. Figure 10 shows the mean number of /b/ and /p/ tokens produced by each group at each age.
A repeated measures ANOVA was performed on the total number of errors in the 30-minute samples, with group as a between-subjects factor. It showed no main effect of age ($F(1,37 = .37, p = .547, \eta_{p}^2 = .028$). There was no main effect of group ($F(2,37 = 2.29, p = .116, \eta_{p}^2 = .038$) and no age x group interaction ($F(2,37) = 1.12, p = .336, \eta_{p}^2 = .006$). Table 8 shows the mean total number of errors by age and group.
Table 8. Total errors by age and group. Mean (s.d.).

<table>
<thead>
<tr>
<th></th>
<th>18 mo.</th>
<th>24 mo.</th>
<th>36 mo.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRA+</td>
<td>2.9 (2.1)</td>
<td>6.0 (10.3)</td>
<td>6.0 (6.9)</td>
</tr>
<tr>
<td>HRA-</td>
<td>6.6 (9.1)</td>
<td>8.6 (7.6)</td>
<td>7.1 (7.8)</td>
</tr>
<tr>
<td>LRC</td>
<td>5.1 (7.3)</td>
<td>7.0 (5.4)</td>
<td>5.6 (7.1)</td>
</tr>
</tbody>
</table>

Subsection Three: Voice Onset Time

A repeated measures ANOVA was performed on mean /b/ VOT with group as a between-subjects factor. There was no main effect of age on /b/ VOT (F(2,68) = .64, p = .53, ηp² = .018). Consistent with previous work, mean /b/ VOT remained unchanged over the course of the study. There was no main effect of group (F(4,68) = .59, p = .670, ηp² = .039). There was no age x group interaction (F(4,68) = .59, p = .67, ηp² = .034). Figure 11 shows mean /b/ VOT by age and group.
Figure 11: Mean /b/ VOT by Age and Group. Error bars indicate ± one standard error of the mean.

Figure 12. Mean /p/ VOT by Age and Group. Error bars indicate ± one standard error of the mean.
A repeated measures ANOVA was also performed on mean /p/ VOT with group as a between-subjects factor. As expected, there was a main effect of age (F(2,40) = 11.4, p < .0005, η²p = .364), with Bonferroni-corrected post-hoc analyses showing that mean VOT for /p/ increased significantly from 12 to 24 months (p = .012) and from 24 to 36 months (p < .0005). There was no main effect of group (F(2,20) = .303, p = .742, η²p = .029) and no age x group interaction (F(4,40) = 1.79, p = .150, η²p = .152). Figure 12 shows mean /p/ VOT by age and group.

**Subsection Four: Production Accuracy**

Three potential measures of production accuracy were investigated: within-child standard deviation (WCSD), t-test, and Cohen’s d. First, a 3-way ANOVA was performed on WCSD of VOT for /b/ and /p/ to examine the effects of age, group, and voicing. Here, the three ages were treated cross-sectionally and groups included participants who were not assessed at all ages. No main effects of age, group, or voicing were found. No age x group interaction was found, and no age x group x voicing interaction was found. However, an age x voicing interaction was found (F(2, 238) = 4.04, p = .019, η²p = .033. This reflects the fact that, consistent with previous findings, the SD of /b/ remained relatively constant with age, while the SD of /p/ increased with age.

A repeated measures ANOVA was also performed on the WCSD of /b/ VOT with group as a between-subjects factor. The repeated-measures ANOVA treats the three ages longitudinally; however, because some toddlers were not assessed at all
three ages, the overall number of subjects was lower than for the 3-way ANOVA. There was no main effect of age (F(2,68) = .27, p = .15, \( \eta^2_p = .008 \)), indicating that toddlers’ production accuracy for /b/ did not change over the course of the study. There was also no main effect of group (F(2,34) = 3.17, p = .055, \( \eta^2_p = .157 \)), indicating that no group of toddlers produced voiced stops with greater variability than any other. Finally, there was no age x group interaction (F(4,68) = .31, p = .870, \( \eta^2_p = .018 \)).

A repeated measures ANOVA was also performed on the WCSD of /p/ VOT with group as a between-subjects factor. Again, there was no main effect of age (F(2,40) = 2.28, p = .115, \( \eta^2_p = .103 \)), indicating that toddlers’ production accuracy for /p/ did not change significantly over the course of the study. There was no effect of group (F(2,20) = .303, p = .742, \( \eta^2_p = .029 \)), and no age x group interaction (F(4,40) = .26, p = .901, \( \eta^2_p = .025 \)). Figure 13 shows the WCSDs of /b/ and /p/ by age and group.
Figure 13. Mean within-child standard deviation of VOT for /b/ and /p/ by age and group. Error bars indicate ± one standard error of the mean.

Second, for toddlers producing more than three tokens of /b/ and of /p/ at each age, a two-tailed t-test was performed to determine whether the two VOT populations were statistically distinct. The inclusion only of toddlers with at least three tokens per category was used to ensure that the standard deviation would be meaningful in each case. A chi-square test for independence was performed between group and number of toddlers with \( p < .05 \), indicating those toddlers with distinct /b/ and /p/ VOT populations. There was a statistically significant difference between group and the number of toddlers who produced VOT populations with \( p < .05 \) at 36 months, \( \chi^2(2) = 7.31, p = .024 \). Table 9 shows the proportion of toddlers at each age and in each group with \( p < .05 \) for voiced and voiceless stop populations.
Third, Cohen's d was calculated for the difference in mean VOT between the two stop populations. Cohen's d in this case indicates the degree of separation of the means of each population. It is calculated as the difference between the means of the two populations divided by the pooled standard deviation, so toddlers with just one token per category (resulting in a standard deviation of 0) or just two tokens per category (resulting in a standard deviation that may not be representative of a larger population) were included in this analysis. A chi-square test for independence was then performed between group and number of toddlers with d > .8, indicating those toddlers with a large mean difference between /b/ and /p/. There was a statistically significant difference between group and the number of toddlers who produced VOT populations with d > .8 at 36 months, $\chi^2(2) = 6.86$, $p = .032$. Table 10 shows the proportion of toddlers at each age and in each group with d > .8 for the mean difference between voiced and voiceless stop populations.

Table 9. Number of Toddlers with Distinct (p < .05) /b/ and /p/ Populations, by Age and Group. *$\chi^2(2) = 7.31$, p = .024.

<table>
<thead>
<tr>
<th></th>
<th>18 mo.</th>
<th>24 mo.</th>
<th>36 mo.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRA+</td>
<td>0/4 (0%)</td>
<td>2/8 (25%)</td>
<td>3/11 (27%)</td>
</tr>
<tr>
<td>HRA-</td>
<td>1/10 (10%)</td>
<td>4/13 (31%)</td>
<td>13/17 (77%)</td>
</tr>
<tr>
<td>LRC</td>
<td>1/4 (25%)</td>
<td>2/14 (14%)</td>
<td>13/19 (68%)</td>
</tr>
</tbody>
</table>
the t-test analysis, supporting the inclusion of toddlers with just one or two /b/ or /p/ tokens per age.

<table>
<thead>
<tr>
<th></th>
<th>18 mo.</th>
<th>24 mo.</th>
<th>36 mo.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRA+</td>
<td>3/11 (27%)</td>
<td>7/9 (78%)</td>
<td>6/11 (55%)</td>
</tr>
<tr>
<td>HRA-</td>
<td>6/12 (50%)</td>
<td>9/16 (56%)</td>
<td>14/18 (78%)</td>
</tr>
<tr>
<td>LRC</td>
<td>2/7 (29%)</td>
<td>12/15 (80%)</td>
<td>18/19 (95%)</td>
</tr>
</tbody>
</table>

Table 10. Number of Toddlers with Large Cohen’s d (d > .8) Separating /b/ and /p/ Populations, by Age and Group. *χ²(2) = 6.86, p = .032.

Subsection Five: Controlling for Differences in Language Scores

In examining the results of the chi-square analysis on distinct VOT populations, the question arises how they might be affected by the differences in language scores seen in Table 7. Specifically, at 24 and 36 months, the LRC group’s mean RL score was slightly above the average range (T-scores of 63.1 and 61.2, respectively; where “average” is defined as ranging from 40 to 60). At 36 months, the LRC group’s mean EL score was 62.7. It is plausible that above-average language scores (especially expressive language) might be associated with superior articulatory performance and thus give rise to between-group differences that are unrelated to ASD diagnosis.

To address this question, a logistic regression was performed to ascertain the effects of RL and EL scores on the likelihood that a participant produced distinct VOT populations at 30 months. RL and EL scores were entered as independent variables; the dependent variable was whether an individual produced distinct VOT populations. The logistic regression model was not statistically significant,
\[ \chi^2(4) = 2.396, p = 0.663. \]
Thus, including the independent variables of RL and EL did not cause the model to predict whether an individual subject’s VOT populations were distinct any better than not including them. Neither independent variable was a significant predictor (for RL, \( p = .786 \); for EL, \( p = .772 \)). Therefore, we conclude that neither RL score nor EL score was significantly related to whether a child produced distinct VOT populations, and that the slightly above-average RL and EL scores of the LRC group at 36 months did not contribute to the between-group differences in distinctiveness of voiced and unvoiced VOT populations.

**Section Five: Discussion: Voice Onset Time Development**

The aim of the acoustic study was to determine whether HRA+ toddlers who have language scores in the normal range and who show no delay in acquisition of the voiced/voiceless distinction in syllable-initial bilabial consonants nonetheless show evidence of a reduced ability to ‘tune up’ their productions to the same degree as HRA- or LRC toddlers.

Language scores show that all three groups scored in at least the average range on tests of both Receptive and Expressive Language. The LRC group showed slightly above-average mean scores at 24 months (RL) and at 36 months (RL and EL). Therefore, group differences in consonant production cannot be ascribed to higher-than-average language ability on the part of the LRC groups, as indicated by a nonsignificant logistic regression model.
In terms of stop production, all three groups produced similar numbers of /b/ and /p/ tokens at all three age. All three groups followed the typical development trajectory described by Mackey and Barton (1980) and Lowenstein and Nittrouer (2008) in that they produced more /b/ tokens than /p/ tokens at all ages, and more tokens of both with increasing age. Thus, group differences cannot be ascribed to a lower production rate. Error rates in the 30-minute samples were also similar between groups. Therefore, HRA+ toddlers did not make more speech errors than HRA- or LRC toddlers.

Consistent with previous results, mean VOT for /b/ remained unchanged from 18 to 36 months in all groups. Overall mean /b/ VOT values of 18.8 ms at 18 months, 17.4 ms at 24 months, and 17.3 ms at 36 months are slightly higher than those of Lowenstein and Nittrouer (2008)’s values of 11 ms at 15-16 months and 16 ms at 21-22 months. Mean VOT for /p/ increased over the course of the study, as expected; groups did not differ on this factor either. The present /p/ VOT values of 22.1 ms at 18 months, 22.4 ms at 24 months, and 54.0 ms at 36 months differ slightly from those of Lowenstein and Nittrouer (2008)’s values of roughly 40 ms through 21-22 months and over 60 ms at 23-24 months, but show the same age-related progression.

Taken together, the bilabial stop production rates, error rates, and VOT values for voiced and voiceless stops were similar across groups in this study and show that all groups followed the normal course of development. This is consistent with the speech attunement framework’s prediction that the ability to ‘tune in’ to
the native language is intact in ASD, at least for participants in the current study whose language scores are within normal limits.

The results for WCSD of VOT for /b/ and /p/ differ slightly from previous findings, but are still consistent with an intact ‘tuning-in’ ability. Unlike in Lowenstein and Nittrouer (2008), no statistically significant difference was found between the SD for /b/ and that for /p/. However, the values found in this study for the SD of /b/ tokens (21.8 ms at 18 months, 23.6 ms at 24 months, and 19.9 ms at 36 months) are higher than previous values (which range between 10 and 15 ms). They are consistent with findings from Bailey and Haggard (1980) and Simon and Fourcin (1978), however, of an age-related production trend toward lower WCSD of VOT for voiced stops. The values reported here for WCSD of /p/ VOT (19.2 ms at 18 months, 28.1 ms at 24 months, and 33.8 ms at 36 months) are close to those measured by Lowenstein & Nittrouer (2008), which ranged from 22 to approximately 38 ms. A possible source for the differences between the present and previous studies is that the toddlers in Lowenstein and Nittrouer (2008) were recorded during a relatively unstructured play session, sitting in a high chair; while those in the present study were free to move about the room (as required for the ADOS). The increased task demands of the ADOS, compared to those of the play interaction described in Lowenstein and Nittrouer, could have affected speech production accuracy in the same way that spontaneous production affects speech as compared to imitated production. Relatedly, though neither study specifically used imitated tokens, there may have been differences in the way that stimuli were
elicited. Children may have repeated the names of toys or actions more in Lowenstein and Nittrouer (2008) than here. These methodological differences could plausibly have resulted in the greater variability seen in the present study.

The major difference found between the HRA+, HRA-, and LRC groups was that significantly fewer HRA+ toddlers produced distinct VOT populations for /b/ and /p/ at 36 months, as defined by a t-test and by a value for Cohen's d > .8 between the mean VOTs for voiced and voiceless stops. The comparison between HRA+ and HRA- toddlers shows that the reduced distinctiveness of VOT populations in the HRA+ group cannot be due to lower stop production rates, higher error rates, or greater variability in /b/ or /p/ production. All of these were the same for these groups. Neither were they affected by slightly above-average language scores in the control group. Instead, a larger proportion of HRA+ toddlers was simply less able to produce a phonetic distinction that communicates meaning to other listeners.

The findings of this study are consistent with the strongest prediction of the speech attunement framework. They show an association between ASD diagnosis and phoneme category distinctiveness in HRA+ toddlers whose language is within the normal range, even when there is semantic pressure to produce the phonetic contrast correctly. The fact that more HRA+ toddlers showed reduced distinctiveness between /b/ and /p/ populations than toddlers in the other groups did not compromise their ability to communicate, but does suggest a lower ability to monitor one’s articulation. Self-monitoring is a skill which plausibly involves some degree of perspective-taking, a skill known to be impaired in ASD. Whatever the
underlying cause of a reduced ability to ‘tune up’, it appears to be present in even
the language-normal group of toddlers with ASD reported on here – but not in their
high-risk peers without ASD.
CHAPTER FIVE: Overall Conclusions

Section One: Implications for Theory

To recap, the social feedback loop posits that ASD diminishes the amount and quality of adult feedback on speechlike vocalizations that children with ASD receive in three ways. First, a lower vocalization rate would afford fewer overall opportunities for receiving feedback. Second, adults’ responses to the speechlike vocalizations of children with ASD have been shown (Warlaumont et al., 2014) to be less contingent on those utterances being speechlike, so the feedback that children with ASD receive could be less clear. Third, children with ASD may not be as able to make use of the feedback they receive from adults in general, either because of auditory processing difficulties or because of reduced ability to self-monitor. However, other work has shown that high-risk infants do receive as much high-quality, language-promoting feedback to their speechlike vocalizations as do low-risk infants (Talbott, et al. (2015a), (2015b), (in prep)); and Haessen et al. (2010) provide evidence of at least normal auditory perceptual skills in children with ASD. Thus, disruptions in the social feedback loop would have to come either from lower vocalization rate or from reduced self-monitoring ability during speech on the part of children with ASD.

On the other hand, the speech attunement framework posits that children who have ASD and no other speech, language, or psychological issues have intact ability to ‘tune in’ to their native languages, so that they perform within normal limits on standardized tests. What children with ASD lack, in this model, is the
ability to ‘tune up’ to a more precise standard of speech production that is characteristic of typically developing children. This could arise as a consequence of lower self-monitoring ability or lower motivation to precisely tune up speech production as finely as children without ASD.

Lower vocalization rate, or lower rate of speechlike vocalizations, is consistent with the social feedback loop framework, has been reported previously, and was found in the present work. The social feedback loop also predicts a positive relationship between speechlike vocalization rate and phonological development. The present study found a positive relationship between rate of speechlike vocalizations and (broadly) size of consonant inventory. However, while the social feedback loop predicts that phonological development in ASD is impeded by a lower speechlike vocalization rate, this prediction was not borne out by the present results. HRA+ toddlers in this sample had a lower speechlike vocalization rate at 12, 18, and 24 months; but did not have smaller consonant inventories at all ages and were no different in terms of canonical syllable production at any age. Although the HRA- toddlers’ vocalization rate was never different from LRC, the former group did show smaller consonant inventories at one age. Lastly, the predicted number of M8 consonants for HRA+ was never significantly different from the actual number. Thus, neither ASD itself nor lower vocalization rate were found to be associated with smaller consonant inventories at any age. In fact, for the HRA- group, it is clear that some factor other than ASD or vocalization rate must be responsible for this group’s different consonant acquisition trajectory.
The speech attunement framework, on the other hand, does not make any predictions about vocalization rate, though lower vocalization rate has been found in ASD as well as in other developmental disorders (Plumb and Wetherby, 2013). Instead, the speech attunement framework addresses the ability of a child with ASD to modify his or her own speech production to precisely match the ambient model. Whether this ability is diminished in ASD by reduced motivation to work toward precise imitation of speech (beyond making oneself understood) or by reduced ability to self-monitor is an open question. Similar speech development trajectories for children with and without ASD whose language is in the normal range are therefore consistent with the ‘tuning in’ part of the speech attunement framework.

The observed differences in speech development between HR and LR toddlers are consistent with a reduced ability to self-monitor and/or reduced motivation to ‘tune in’ in HRA+, as well as with intact ability to self-monitor and intact motivation to ‘tune in’ in HRA-. Thus, the use of both HRA+ and HRA- groups to investigate how speech development is affected by ASD is an important addition to the literature. Comparing high-risk toddlers who do develop ASD to those who do not allows a fractionation of the protective factors that HRA- children possess but that HRA+ children may not.

The use of both perceptual and acoustic methods of speech analysis also deepens our understanding of how spoken language development is affected in ASD. While the social feedback loop predicts that differences in speech production between children with ASD and those without will always rise to the level of
perceptibility by adults, the speech attunement framework allows for the presence of imperceptible but measurable differences. These were found in the production of syllable-initial bilabial stops -- in some of the HRA+ toddlers in this study. The present VOT results constitute evidence for the speech attunement framework as it applies to ASD. All groups of toddlers in the studies discussed here had language within the normal range, so impaired language was ruled out as a primary cause of any between-group differences. All of the between-group differences must be ascribed to risk group and diagnostic status. Also, since all children are under semantic pressure to clearly articulate /b/ and /p/ differently, the presence of imperceptible but measurable differences between the children who develop ASD and those who do not can be associated with their diagnostic group.

The social feedback loop framework seems not to be diminished by lower vocalization rate or by less or lower-quality parental feedback. By 24 months, the performance of the HRA+ group was indistinguishable from that of the other groups, suggesting that none of these factors ultimately affected speech development for the HRA+ toddlers. It is still possible, however, that the differences we see in the speech production of children who develop ASD result from impaired ability to make use of the feedback they receive from adults. In this sense, the social feedback loop and speech attunement frameworks may converge.
Section Two: Clinical Implications

The present work suggests that the Middle-8 consonants are, across a range of ages, a sensitive indicator of phonological progress in infants and toddlers. Thus, reduced appearance of M8 consonants in pre-speech babble and in speech may be a subtle indicator of delay. The findings of this study also underscore the vigilance argued for by Paul et al. (2011) in monitoring the early speech behavior of high-risk toddlers in the second year of life. Given the higher prevalence of residual speech errors documented by Shriberg et al. (2001) in teens and young adults with ASD (approximately 33%) than in the general population (approximately 2%) and given the degree to which such residual errors are socially penalizing (Cleland et al., 2010), it is important to provide early, targeted therapy to minimize the downstream effects of such misarticulations.

Because the results of the acoustic study did not reveal an elevated level of perceptually evident speech production differences between HRA+ and HRA- or LRC toddlers, they do not motivate treatment of /b/ and /p/ per se. Instead, they speak to the hypothesis of reduced ‘tuning up’ ability, a skill presumably related to other findings in ASD, such as reduced perspective-taking and reduced interest in social stimuli. To the extent that reduced ‘tuning up’ ability does result in perceptually evident misarticulations, however, it is important to address these issues promptly, so they do not persist into later childhood.

Thus, the current results point to the necessity of articulation therapy for young children with ASD who have misarticulations, to ensure that these speech
errors do not persist. For older children with ASD, however, activities targeting both speech clarity and self-monitoring may be more effective and interesting. Many older children and teens with ASD are interested in drama or broadcasting, and these areas form a useful context in which to frame the concepts of maximizing speech clarity. In their text on speaking for broadcasters, for example, Utterback and Freeman (2005) provide a systematic framework and set of exercises related to engaging, effective, and vocally healthy public speaking. It begins with education on speech breathing, phonation, and articulation to build awareness of the speech mechanism and how it is used. Further chapters cover how to enhance meaning through stress and intonation – important for students with ASD, as stress and intonation are often disordered in this population. Linking prosody to communication of meaning helps students who have difficulty spontaneously producing appropriate prosody not only understand what appropriate prosody is, but what its function is. Other chapters cover sounding conversational, which is important for students who might otherwise develop a robotic-sounding way of speaking. Training programs like that of Utterback and Freeman teach awareness of one’s own spoken communication and build perspective-taking into the process of speaking to inform or entertain listeners. Such programs are potentially a powerful way to teach older students with ASD and articulation issues how to produce speech that is more clear in a way that feels less like speech therapy and more like professional training. For those students whose special interest is in television,
video games, or other forms of electronic entertainment, the link with their topics of interest can also be tremendously motivating.

Finally, as Cleland et al. (2010) state, the presence of speech sound disorders “adds an additional communication and social barrier” for individuals with ASD, and as such demands therapy. For these individuals, reducing or eliminating persistent speech-sound distortions is far more than a cosmetic procedure. For younger children with ASD, it could substantially improve the desire of peers to interact with them, thus offering more opportunities for improving social skills and ameliorating the effects of an ASD diagnosis. For teens and young adults, elimination of persistent speech-sound errors could mean the difference between getting a well-paying job or not, and thus between living independently or being dependent on parents or even siblings for life.

Section Three: Limitations and Future Work

A prominent limitation to the present work is the number of participants. The work described here involved approximately 15 toddlers with ASD. While this is a promising start to investigations in speech production in ASD, this work should be expanded by including more children with or at risk for ASD.

Further, the focus here on children with average language ability is both a strength and a limitation. On the one hand, group differences found in such a high functioning group of high-risk toddlers strongly suggest an ASD-related cause for the differences (rather than one related to language or cognitive level). On the other
hand, limited group differences between high-functioning toddlers who develop ASD and low-risk controls does not rule out the possibility that ASD is only a weak contributor to speech delay or disorder in the current participants. In more affected individuals, ASD may have a larger role to play in speech delays or deficits, possibly because of the lack of protective factors such as high cognitive ability in these populations. Thus, more research, involving more individuals from every range of the autism spectrum, is motivated.

A skill not assessed in this study is auditory perception, one source of feedback involved in self-monitoring of speech production. A relationship between perception and production in stop consonants was found by Bailey and Haggard (1980), who demonstrated that children requiring a long VOT for consistent voiceless identification also produced voiceless stops with long-lag VOTs. It is possible that the HRA+ individuals whose /b/ and /p/ populations were less distinct than for the other children would also show differences in the mean VOT or the slope of the identification function near the boundary between voiced and voiceless stops. The relationship between speech perceptual ability and speech production should be assessed in the ASD population, as Bailey and Haggard (1980) did for typically-developing toddlers. The speech attunement framework predicts that individuals with ASD have auditory-perceptual skills at least in the normal range (this is supported by work cited in Haessen et al. (2010)) but that it is challenges in social reciprocity that make it difficult for them to monitor their own speech and voice production in ongoing interactions. Similarly, the social feedback
loop framework posits that reduced ability to make use of adult feedback may affect phonological development in children who develop ASD. This reduction my come about precisely because of reduced ability to self-monitor.

Self-monitoring speech requires attention to both acoustic and somatosensory feedback. The existence of multiple sources of feedback and the question of how children with ASD make use of them during speech relates to another avenue of future investigation. This avenue involves computational models of speech development and production, such as the DIVA model (Tourville and Guenther, 2011; Terband et al., 2014a; Terband et al. 2014b), and their relationship to childhood speech sound disorders. The DIVA model is a neural network that models the feedforward (articulatory to acoustic) and feedback (acoustic to articulatory) control loops that are posited to operate in speech development and in typical adult speech production. DIVA uses the articulatory-to-acoustic mappings it has learned during its babbling stage to produce approximations of its native language targets, and it uses the acoustic-to-articulatory mappings it has acquired to refine its production attempts. When DIVA can produce the targets without auditory errors, it has reached the mature production stage.

Terband et al. (2014a) compared DIVA-modeled predictions to behavioral measures of the speech of children whose auditory feedback of their own speech had been experimentally perturbed. They investigated the hypothesis that diminished quality of auditory feedback and diminished ability to use it to make online corrections in one’s speech production may be implicated in childhood
speech sound disorders. They found that typically-developing children showed more of a compensatory response to acoustic perturbation of their speech, but children with speech sound disorders showed more of a following response. Terband et al. (2014b) subsequently extended these results, using DIVA to model both a motor processing disorder and a motor plus auditory processing disorder during development, by adding varying amounts of noise into the motor and auditory state representations in the system. A motor processing disorder alone in DIVA resulted in impaired phonological representations, as long as auditory self-monitoring was intact. By contrast, a motor plus auditory processing disorder in the model resulted in impaired articulatory-to-acoustic mappings, which the authors interpreted as being consistent with childhood apraxia of speech (CAS).

In related work, Iuzzini and colleagues (Iuzzini et al., 2013; Iuzzini et al., 2014) investigated the effects of masking auditory feedback in children with CAS and in typically-developing children. They found that when auditory feedback was masked, the children with CAS showed a deterioration of the voicing contrast in their speech, while the typically-developing children did not. They interpreted this result to indicate that children with CAS are over-reliant on using acoustic feedback to produce speech because the articulatory-to-acoustic mapping (DIVA’s “feedforward model”) is impoverished.

These two lines of research play into a long-standing hypothesis that CAS is responsible for the differences in speech found in individuals with ASD compared to typical speakers (Shriberg, 2010). The strong form of the hypothesis states that CAS
accounts for the failure of speech to develop in minimally verbal children with ASD; a weaker form of the hypothesis states that milder versions of CAS are responsible for the speech-related deficits in verbal individuals with ASD. Shriberg et al. (2001) examined a variety of speech, voice, and prosodic findings in older children and teens with ASD and compared them to children with CAS, showing that the acoustic and perceptual features of the two groups were different. This suggests that the weak form of the hypothesis, at least, may not be tenable.

However, the research by Terband and colleagues and by Iuzzini and colleagues raises interesting speculations and provides testable hypotheses regarding the role of CAS in the speech of individuals with ASD. First of all, it should be possible to perform the experiments of Terband et al. (2014a) with children with ASD, providing them with perturbed auditory feedback (shifted formants) to their own speech. A similar experiment was performed by Russo et al. (2008), comparing how children with ASD and typically-developing children responded to pitch-perturbed auditory feedback. Of the 19 TD children, 16 produced compensatory responses and the other 3 produced following responses. In the ASD group, not surprisingly, there was more heterogeneity: 5 produced following responses and 13 produced compensatory responses. And of the 13 children with ASD who compensated for the pitch-altered feedback, 8 had a response magnitude that was lower than that of the TD controls. The other 5 had a response magnitude that was greater than that of the TD controls. These results suggest that acoustic, somatosensory, and motor integration may be disrupted in many ways in children
with ASD, and that there may be subgroups with one (or more) forms of disruption. This, in turn, accords with the view of ASD as quantitatively, rather than qualitatively, different from typical development in that different combinations of risk and protective factors combine to give rise to the variety of behaviors that we see. Because different underlying pathologies demand different treatment techniques, and because so many children with ASD have disordered spoken language or lack it entirely, more research in this area is of the utmost importance both for theoretical and clinical reasons.


CURRICULUM VITAE

Karen Chenausky, M.S., CCC-SLP

100 Cummington Mall, Rm. 152  kchenaus@bu.edu
Boston, MA 02215  617-358-5914 (office)

Education
Massachusetts Institute of Technology  S.B.  1989  Philosophy and Linguistics
University of Connecticut  M.A.  1992  Linguistics
Boston University  M.S.  2005  Speech Pathology
Boston University  Ph.D.  2015  Speech Pathology

Honors
2004  S. Klein Prize for Science and Technical Writing, MIT.
2009  Massachusetts Educator’s License.

Clinical Positions and Employment
1989-92  Univ. of CT, Int'l Teaching Asst’s. Program. Instructor, English as a Foreign Language.
Developed curriculum to teach professional and scientific English communication to graduate students.

2001-2006  Staff, Ely Associates (social communication private practice, Auburndale, MA)
Created group activities for school-age pragmatics groups, worked 1:1 with clients requiring speech/language and pragmatics work, and co-led community outings for children in pragmatics groups.

2004-2007  Speech Pathologist, Trauma and Learning Policy Initiative, Harvard Law School
Co-wrote book chapter about how to perform speech-language evaluations with children who have been exposed to trauma. Performed trauma-sensitive speech and language evaluations for clients receiving legal aid from Massachusetts Advocates for Children.

2005-2006  Research Associate, LADDERS division of Spaulding-Wellesley, Wellesley MA
Administered a battery of developmental tests to siblings of children with autism, aged one to sixteen years of age. Tests included ADOS, Mullen, Vineland, PLS, and others. Counseled parents as to results and made referrals as appropriate.

2006-2007  **Clinical Fellow in Speech Pathology, OTA Watertown, Watertown, MA**  
Provided 1:1 speech/language services to clients from one to four years of age in a sensory-integration milieu. Co-led feeding groups with occupational therapists for school-age children. Mentor: Elsa Abele, M.S., CCC-SLP.

2006-2009  **Consulting Speech Pathologist, YouthCare at MGH for Children, Charlestown, MA**  
Educated camp counselors about language and autism. Supervised counselors. Advised group leaders as to activities, program structure, and specific clients.

2006-2011  **Consulting Speech Pathologist, EDCO Educational Collaborative, Waltham MA**  
Provided speech/language services to behaviorally-disordered and high-functioning autistic high school students. Led social skills groups for autistic students. Co-led community outings and travel training for both groups of students.

2011-2013  **Graduate Intern, Office for Disability Services, Boston University**  
Supported students with learning and psychological disabilities. Arranged letters of accommodation and found items and services that students may need, in compliance with the ADA and with University requirements. Provided coaching in organizational strategies and referred for outside services as needed.

2013-pres.  **Research Assistant, Autism Center for Excellence, Boston University**  
Behavioral testing of typical and autistic, verbal and minimally-verbal children and teens. Administer ADI, Vineland, proprietary questionnaires to parents and ADOS to children. Collaborate on developing appropriate modifications to standardized testing as needed to ensure best performance from participants. Design experiments to investigate speech and language capabilities of minimally-verbal participants. Write evaluation reports for parents as requested.

**Other Experience and Professional Memberships**

- **1993-1996**  MIT Speech and Hearing Sciences Program. Trainee.
- **1998-2003**  Northeastern University. Lecturer, Speech Science; Phonetics.
- **2010-2012**  Speech Technology and Applied Research Corp, Bedford MA. Senior Scientist.
- **2013**  College of Our Lady of the Elms (Chicopee, MA). Adjunct faculty, lecturing on autism.
Selected Peer-reviewed Publications


Presentations


“Speech Development in Children at High and Low Risk for Autism.” Invited Presentation, NIH Autism Center of Excellence Workshop on Analysis and Assessment of Early Vocal Behavior. May 2014, Emory University, Atlanta, GA.

Doctoral Dissertation

“Speech Development in Children at High and Low Risk for Autism.” Aim: Document differences in the pre-speech babbles of children with and without siblings on the autism spectrum by examining the types and proportions of utterances the children produce, by documenting the expansion of their vowel spaces, and by tracking the development of the voiced/voiceless distinction in word-initial bilabial stops.

Research Support


Past Research Support


Past Research Positions

“System to Detect Severe Respiratory Illness by Vocalization Analysis of Cough”, NIAID Phase-1 STTR (R43-HL096246)  P.I.: Suzanne Smith. Responsibilities: Acoustic analysis of recorded coughs, creation of novel measures distinguishing coughs from sick and well individuals.


“Dynamic Aspects of Respiratory Control in Normal and Stuttering Speakers.”, NIDCD Research Project Grant (R01 DC03781-02). P.I.: Margaret Denny. Responsibilities: Syntactic analysis of speech samples; matching with breath-control data.


