

Acquisition of Voice Onset Time in Toddlers at High and Low Risk for Autism Spectrum Disorder

Karen Chenausky  and Helen Tager-Flusberg

Although language delay is common in autism spectrum disorder (ASD), research is equivocal on whether speech development is affected. We used acoustic methods to investigate the existence of sub-perceptual differences in the speech of toddlers who developed ASD. Development of the distinction between b and p was prospectively tracked in 22 toddlers at low risk for ASD (LRC), 22 at high risk for ASD without ASD (HRA-), and 11 at high risk for ASD who were diagnosed with ASD at 36 months (HRA+). Voice onset time (VOT), the main acoustic difference between b and p, was measured from spontaneously produced words at 18, 24, and 36 months. Number of words, number of tokens (instances) of syllable-initial b and p produced, error rates, language scores, and motor ability were also assessed. All groups' mean language scores were within the average range or slightly higher. No between-group differences were found in number of words, b's, p's, or errors produced; or in mean or standard deviation of VOT. Binary logistic regression showed that only diagnostic status, not language score, motor ability, number of words, number of b's and p's, or number of errors significantly predicted whether a toddler produced acoustically distinct b and p populations at 36 months. HRA+ toddlers were significantly less likely to produce acoustically distinct b's and p's at 36 months, which may indicate that the HRA+ group may be using different strategies to produce this distinction. *Autism Res* 2017, 0: 000–000. © 2017 International Society for Autism Research, Wiley Periodicals, Inc.

Keywords: autism; speech development; phonological development; stop consonants; broader autism phenotype; voice onset time

Acquisition of Voice Onset Time in Toddlers at High and Low Risk for Autism

Autism Spectrum Disorder (ASD) is characterized by persistent deficits in social communication and by restricted interests and repetitive movements, which cause clinically significant impairment in functioning [American Psychiatric Association, 2013]. Yet the domains over which ASD is defined (motor skill, social ability, language) are the same as those required for an individual to communicate using spoken language. It is therefore reasonable to expect speech to be affected in ASD. Since speech data can be collected non-invasively, it is worth prospectively investigating speech development in toddlers at risk for ASD. Differences may help us understand how linguistic and motor constraints influence the ability of children with autism spectrum disorder to produce phonological contrasts.

Research on whether speech is impaired in ASD has been equivocal. While some research demonstrates that speech is relatively spared in ASD [Boucher, 1976;

Kjelgaard & Tager-Flusberg, 2001; McCann, Peppé, Gibbon, O'Hare, & Rutherford, 2007; Tager-Flusberg, 1981], others report elevated rates in speakers with verbal ASD of: speech delay [Bartak, Rutter, & Cox, 1975; Bartolucci & Pierce, 1977; Bartolucci, Pierce, Streiner, & Eppel, 1976; Cleland, Gibbon, Peppé, O'Hare, & Rutherford, 2010; McCleery, Tully, Slevc, & Schreibman, 2006; Schoen, Paul, & Chawarska, 2011; Velleman et al., 2009], speech errors [Cleland et al., 2010; Rapin, Dunn, Allen, Stevens, & Fein, 2009], persistent speech disorder [Cleland et al., 2010; Shriberg et al., 2001], and motor speech disorders [Velleman et al., 2009].

Other factors complicate what conclusions can be drawn from the literature regarding whether speech is affected in ASD. First, previous studies have not controlled for language ability, though language delay is known to be associated with speech delay [Pennington & Bishop, 2009]. Second, the literature on infant siblings of children with ASD, who are themselves at higher risk of developing ASD, shows that language and motor development are frequently delayed not only in

From the Music and Neuroimaging Lab, Neurology Department, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, Massachusetts (K.C.); Department of Psychological and Brain Sciences, Center for Autism Research Excellence at Boston University, 100 Cummington Mall, Boston, Massachusetts (H.T.-F)

Received March 27, 2016; accepted for publication February 06, 2017

Address for correspondence and reprints: Karen Chenausky; Music and Neuroimaging Lab, Neurology Department, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, Massachusetts. E-mail: kvchenau@bidmc.harvard.edu

Published online 00 Month 2017 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/aur.1775

© 2017 International Society for Autism Research, Wiley Periodicals, Inc.

those siblings who develop ASD, but also in the high-risk siblings who do not [Gamliel, Yirmiya, Jaffe, Manor, & Sigman, 2009; Iverson & Wozniak, 2007; Landa, Gross, Stuart, & Bauman, 2012]. Thus, it is unclear whether the speech findings in toddlers and children with ASD are associated with an ASD diagnosis specifically or whether they are part of a shared vulnerability for language delay common to high-risk families.

Subphonemic Variation in Typical Development: The Voiced/Voiceless Distinction

Acquisition of certain phonetic contrasts is more complex and nuanced than simple correct/incorrect judgments indicate. One area in which this is true is the development of the voiced/voiceless distinction in stop consonants. Stop consonants (/b, d, g; p, t, k/) are so named because airflow through the vocal tract is completely stopped for a brief period (<100 ms) during their production. The former three are referred to as “voiced” and the latter three as “voiceless” stops. Here, we focus on /b/ and /p/, the earliest-developing voiced–voiceless pair of stops.

To produce the distinction between /b/ and /p/; that is, to convey to a listener the difference between words such as “big” and “pig,” a high degree of motor coordination of laryngeal vibration and between laryngeal vibration and opening of labial closure is necessary [Koenig, 2000]. For American English /b/ in adult speech, laryngeal vibration resumes within 20 ms of labial opening; for /p/, it resumes 60 ms or more after labial opening [Koenig, 2000; Stevens, 1998]. This interval is called *voice onset time*, or VOT, and is the main acoustic difference between voiced and voiceless stops [Lisker & Abramson 1964]. VOT measurements can thus reliably quantify how consistently a speaker is able to meet the simultaneous linguistic and motor goals of producing two separate categories of labial stops (voiced and voiceless).

Development of the /b/-/p/ distinction is protracted and includes a period during which children produce a covert distinction—that is, they produce acoustically, but not perceptually, distinct /b/s and /p/s. Because sub-perceptual differences in speech sounds are a part of normal speech development, another limitation of previous research is that, with a few exceptions [e.g., Diehl & Paul, 2012, 2013; Diehl, Watson, Bennetto, McDonough, & Gunlogson, 2009], perceptual methods have been used for assessing the presence of articulatory differences in toddlers and children with or at risk for ASD. While perceptual methods are ecologically valid, one of their limitations is that that acoustic differences between tokens (exemplars) of the same phonetic

category are very difficult to identify perceptually [Kent, 1996], so they cannot reveal covert distinctions.

The stages in acquisition of the /b/-/p/ contrast in syllable-initial stops in English were first documented in a longitudinal study by Macken and Barton [1980] of 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 /b/-/p/ contrast. In Stage 1, the “no contrast” stage, there are no perceptual or acoustic differences in the intended voiced and voiceless stops that children produce. Toddlers’ productions of “bat” and “pat” not only both sound like “bat” to adults during this stage, but there is no acoustic difference between the two consonants.

At the beginning of Stage 2, the “covert contrast” stage, the mean VOT for intended /b/ is unchanged, but the mean VOT for intended /p/ lengthens—though it remains within the adult “voiced” category and there is still considerable overlap between the two populations of stops. In Stage 2 stage, though “bat” and “pat” still sound the same to adults, there is now a measurable acoustic difference between intended /b/ and /p/.

Finally, in Stage 3, toddlers produce intended /p/s with VOTs in a range that sounds like /p/ to adults. Koenig [2000] showed that, by the time children are five years of age, the means and standard deviations for their /b/ and /p/ populations were similar to those of adults (mean 12.6 ms, S.D. 6.1 for /b/; mean 216 ms, S.D. 52.6 for /p/). Thus, as children develop in their ability to produce the voicing distinction, there is less overlap between VOT populations for voiced and voiceless stops.

Lowenstein and Nitttrouer [2008] largely replicated the findings of Macken and Barton [1980] with a group of seven toddlers, taped at 2-month intervals between the ages of 14 and 31 months. In addition, Lowenstein and Nitttrouer found that the variability in VOT for toddlers’ /b/s was lower than for /p/s and decreased slightly as children matured, which the authors interpreted as reflecting improved production accuracy. The variability in VOT for /p/, by contrast, was significantly higher than for /b/ and was interpreted as lower production accuracy. These results were taken by the authors to indicate that, while the target VOTs for /p/ became more adult-like, toddlers’ skill in producing those target values did not improve, over the course of the study. Research on children up to age 5 has not found sex differences in development of the /b/-/p/ distinction [Whiteside, Henry, & Dobbin, 2004].

Taken together, then, previous research suggests that the mean VOT for voiced and voiceless stops, the standard deviation of VOT for those stops, and the degree to which voiced and voiceless stop populations are distinct (acoustically and statistically) all index the development of the distinction between /b/ and /p/. This

developmental progression is related to speech delay in that speech-delayed children produce /p/ with lower accuracy rates than typically developing children. For example, Shriberg [1993] found that speech-delayed children between 3 and 6 years old produced /p/ with approximately 83% accuracy, compared with rates of 90%–92% for typically developing children. In this context, lower accuracy rates mean that what the child produced when intending /p/ was not heard as a /p/ by adults. Without acoustic data, it is not possible to know whether these misarticulated /p/s had VOT values in the /b/ range. Regardless, the finding does suggest that the ability to produce acoustically distinct /b/ and /p/ populations is a skill that is acquired later by children with speech delay. As such, it may indicate that motor and language constraints differently affect the ability of at-risk groups, including toddlers at risk for ASD, to accurately produce phonological distinctions.

Finally, there is research showing that both gross motor [Bedford, Pickles, & Lord, 2015] and fine motor ability [Sauer LeBarton & Iverson, 2013] predict larger expressive language ability in children with or at risk for ASD [but see also Wang, Lekhal, Aaro, Holte, & Schjolberg, 2014]. However, around the time of canonical babbling onset, when children begin to produce consonant-vowel syllables and strings of syllables (between 6 and 10 months in typically developing infants), there is also a significant increase in rates of rhythmic upper limb movement—that is, simultaneous banging and babbling [Iverson & Fagan, 2004; Iverson & Wozniak, 2007]. To understand whether language or motor ability was involved in the development of the /b/-/p/ distinction, we also included measures of expressive language, receptive language, and fine motor (rather than gross motor) scores predictors in our analyses.

Research Questions

Because previous research suggests that speech development is delayed in at least some toddlers at high risk for ASD, and because few other studies have employed acoustic methods to investigate speech production in toddlers at high risk for ASD, we sought to answer two questions:

1. Are perceptual or sub-perceptual production differences present in the speech of toddlers who go on to develop ASD?
2. Are these differences shared by high-risk siblings who do not go on to develop ASD, thus forming part of a separable comorbidity, or are they associated with ASD specifically, suggesting altered speech development in some children with ASD?

Methods

Participants

Our sample included 55 infants: 22 low-risk controls (LRC) with a typically developing older sibling and no family history of ASD (5 male) and 33 toddlers at high risk for ASD (HRA) who had an older sibling with ASD. 11 HRA toddlers (7 male) received diagnoses of ASD at 36 months and are referred to as HRA+ (i.e., high-risk siblings with ASD). The remaining 22 HRA toddlers (13 male) did not develop ASD and are referred to as HRA– (i.e., high-risk siblings without ASD).

Family history of ASD was queried during a pre-enrollment phone screen. Diagnosis of ASD in the HRA probands (and confirmation that the LRC probands did not have ASD) was corroborated via parent report using an age-appropriate screener prior to enrollment: for probands over 4 years old, the Social Communication Questionnaire was used [SCQ; Rutter, Bailey, & Lord, 2003]; for probands under 4 years old, the Pervasive Developmental Disorders Screening Test-II was used [PDDST-II; Seigel, 2004]. After initial screening, participants were enrolled in a longitudinal infant sibling project and invited to participate regularly until 36 months, with data collected through parent report, behavioral, eye-tracking, and neural measures of development.

To be included in this study, infants needed to complete lab visits at 18, 24, and 36 months and to have received an Autism Diagnostic Observation Schedule [ADOS; Lord et al., 2000] assessment from a research-reliable experimenter at 36 months. Diagnoses of ASD were made according to expert community clinicians before the family's enrollment in the study; after enrollment, diagnoses were verified using the ADOS, the Social Communication Questionnaire [SCQ; Rutter, et al. 2003], or best estimate clinical judgment. Table 1 details language scores on the Mullen Scales of Early Development [MSEL; Mullen, 1995] for the HRA+, HRA–, and LRC groups. As mentioned above, we also included fine motor scores from the MSEL, as this section focuses on actions with the hands and arms (e.g., bringing fists to midline or banging blocks in midline) rather than items mainly related to posture and gait, as is true of the gross motor section. Project approval was obtained from the Institutional Review Boards at Boston Children's Hospital and Boston University, and informed consent was obtained from the parents of each infant participant.

Procedures

Audio recordings from the first 30 minutes of the ADOS from the 18-, 24-, and 36-month visits were used as a spontaneous speech samples for each toddler. Sessions were recorded in rooms equipped with two Sony SNC-

Table 1. Descriptive Characteristics by Age and Group

	Low-Risk Controls (LRC)	High Risk + ASD (HRA+)	High Risk – ASD (HRA–)
Number	22	11	22
Sex	5 M, 17 F	7 M, 4 F	13 M, 9 F
36-mo. ADOS Score (mean ± S.D.)	2.8 ± 3.4	8.3 ± 3.5	3.4 ± 2.9
Language T-Scores			
18 months EL	51.0 ± 6.7	44.7 ± 11.7	50.5 ± 10.6
18 months RL	59.6 ± 13.0	41.1 ± 17.3	51.3 ± 13.5
24 months EL	58.1 ± 8.7	49.1 ± 11.0	51.3 ± 7.4
24 months RL	63.1 ± 6.9	47.9 ± 16.6	54.1 ± 5.8
36 months EL	62.7 ± 7.9	53.6 ± 8.6	57.9 ± 6.4
36 months RL	61.2 ± 8.3	48.8 ± 16.6	53.9 ± 7.9

RZ30N cameras and two SHURE SM57 microphones. Audio was extracted from the video files using AoA Audio Extractor [AoAMedia, 2013], downsampled to 16 kHz with Audacity [Audacity, 2013], and visualized and played back with Wavesurfer [Sjölander & Beskow, 2011].

Utterances in the 30-minute samples that contained syllable-initial /b/ and /p/ were extracted, yielding a total sample of 5,389 stops. Both words (i.e., utterances that closely matched the adult form, such as [bʌbo] (“buhbo”) for “bubble”) and word approximations (i.e., utterances that less closely matched the adult form, such as [bʊn] (vowel as in “book”) for “balloon”) were included in the sample. Each utterance was glossed according to the intended word and broadly phonetically transcribed. Gloss was made 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.

Counts and Measurements

Several counts were made of the words in each session for each participant. First, the number of words for each child and the number of /b/ and /p/ tokens in those words from the 30-minute samples were 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] (“bop”). Also included in the count of words and /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.

Time waveforms and wideband spectrograms of the audio files were displayed in Wavesurfer. VOT was then measured for each syllable-initial singleton bilabial stop in identifiable, unobscured, non-whispered words. Using both visual and auditory information, markers were placed by hand at (1) the broadband, aperiodic

burst marking the oral release and (2) the beginning of regular, periodic glottal vibrations marking the voice onset. This is illustrated in Figure 1. The time of each marker was recorded into a spreadsheet and VOT was then calculated as the time interval between the two markers. Occasionally, the VOT of a word with an intended bilabial stop could not be measured or categorized because of a speech error. When this was the case, the type of error was noted and the total number of errors tallied for each child. Error types included stops with no oral closure, stops that were voiced all the way through, stops that were nasalized or fricated (i.e., manner errors), and utterances for which the intended word could not be discerned (e.g., [bæp] (“bap”) or [gæbwʊd] (“gabwood”). For each child who produced at least three non-errored tokens of /b/ or /p/ at each age, the mean and the (within-child) standard deviation of the VOT for each category was calculated. For each child who produced at least three non-errored tokens of /b/ and three non-errored tokens of /p/ at each age, a *t*-test was used to determine whether the two VOT populations were statistically distinct from each other.

Measurement Reliability

An additional judge, blind to group status, independently measured 11 randomly selected audio files (10% of the total). Each file consisted of a full 30-minute recording for a given participant at a particular age. Pearson’s product-moment correlation between VOT measurements for the two judges was $r = .902$, $P < .0005$, with a mean difference of 0.6 ms between judges. These figures are comparable to those of Macken and Barton [1980]: 6 ms difference between judges; Forrest and Rockman [1988]: Pearson’s $r = .95$; Whiteside, Dobbin, and Henry [2003]: Pearson’s $r = 0.978$; and Hitchcock and Koenig [2013]: 2.1 ms difference between judges, Pearson’s $r = 0.996$.

Results

Analysis of Language Scores

The mean score on the MSEL language subtests is 50. All three groups included some participants with language T-scores lower than 40 and some with T-scores greater than 60. Because of the high standard deviations in Table 1, we performed one-way ANOVAs on mean RL and EL score with group as a between-subjects factor to understand whether there were any significant between-group differences. Results revealed significant between-group differences in RL only at 24 and 36 months and in EL only at 36 months. Post-hoc, Bonferroni-corrected analyses revealed that in each case these findings were driven by above-average LRC mean scores, making the LRC group significantly different

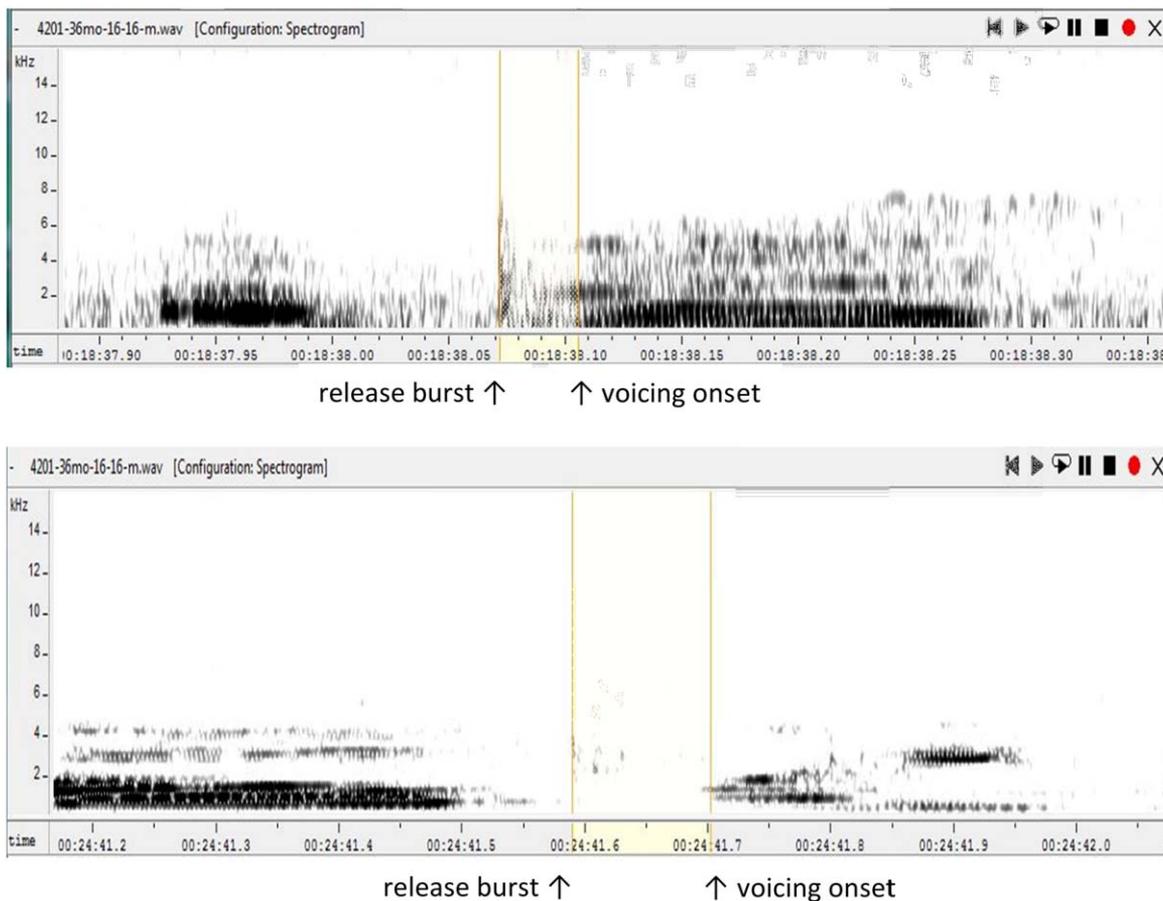


Figure 1. Illustration of VOT measurements. Top panel: Measurement for /b/ in “a big one” is illustrated. Bottom panel: Measurement for /p/ in “open it.” In both cases, the interval of time between the release burst (on the left of each shaded section) and the onset of voicing (on the right of each shaded section) is the voice onset time. (Note: plots are not shown to same scale).

from the HRA+ group. At no ages were the HRA+ and HRA– groups or HRA– and LRC groups significantly different from each other.

Description of Children’s Productions: Counts and Error Rates

Table 2 shows the mean number (and standard deviation) of words, errors, and /b/ and /p/ tokens produced by each group at each age. It also includes the number of children at each age who produced at least three non-errored tokens of /b/ and the number who produced at least three of /p/. A repeated-measures ANOVA on number of words with group as a between-subjects factor showed a main effect of age, $F(2,104) = 9.540, P < .0005$, no main effect of group, and no age \times group interaction. All three groups produced similar numbers of words over time. A 2-way ANOVA with number of errors as the dependent variable and age and group as between-subjects factors revealed no significant main effects of age or group. All three groups produced similar numbers of errors over time. There was also no significant age \times group interaction. Finally, a 3-way ANOVA with number of stop tokens as the

dependent variable, age and group as between-subjects factors, and voicing status (/b/ or /p/) as a within-subject factor showed a main effect of age, $F(2,274) = 11.94, P < .0005$. All groups produced more stops with increasing age. There was also a main effect of voicing, $F(1,274) = 50.7, P < .0005$. All groups produced more /b/ tokens than /p/ tokens. There was no main effect of group and no significant 2- or 3-way interactions.

Voice-Onset Time

Figure 2 shows the mean VOT for /b/ and /p/ as a function of age, for all three groups (for children who produced at least three non-errored tokens of /b/ or /p/). A 3-way ANOVA was performed with mean VOT as the dependent variable, age and group as between-subjects factors, and voicing status as a within-subjects factor. There were no significant main effects of age, group, or voicing; and no significant 2- or 3-way interactions.

Because VOT populations for /b/ and /p/ do not begin to separate until near 3 years of age, and because developmental differences in mean /p/ VOT could be overshadowed by the larger number of /b/ tokens that

Table 2. Counts by Age and Group

	LRC	HRA±	HRA-
18 months mean # words	27.5 ± 30.0	32.5 ± 32.0	38.1 ± 31.8
24 months mean # words	42.1 ± 22.3	53.5 ± 47.5	81.1 ± 66.0
36 months mean # words	86.3 ± 39.5	54.1 ± 22.4	71.8 ± 47.2
18 months mean # errors	9.0 ± 13.3	4.5 ± 3.1	12.0 ± 17.0
24 months mean # errors	9.9 ± 8.4	8.4 ± 12.3	13.4 ± 11.3
36 months mean # errors	6.6 ± 8.6	8.5 ± 10.9	8.0 ± 7.8
18 months mean # of /b/	15.7 ± 18.1	19.1 ± 17.9	25.5 ± 23.8
24 months mean # of /b/	22.7 ± 15.7	30.1 ± 30.0	38.1 ± 35.5
36 months mean # of /b/	39.2 ± 24.8	26.4 ± 12.7	35.2 ± 21.9
18 months mean # of /p/	23.8 ± 4.0	6.7 ± 13.1	5.9 ± 7.2
24 months mean # of /p/	8.6 ± 7.6	9.0 ± 8.0	16.6 ± 25.7
36 months mean # of /p/	22.2 ± 12.6	14.5 ± 7.1	20.2 ± 12.9
18 months # children with ≥ 3 /b/	14	11	15
24 months # children with ≥ 3 /b/	19	11	19
36 months # children with ≥ 3 /b/	19	11	19
18 months # children with ≥ 3 /p/	10	7	13
24 months # children with ≥ 3 /p/	18	9	18
36 months # children with ≥ 3 /p/	19	11	18

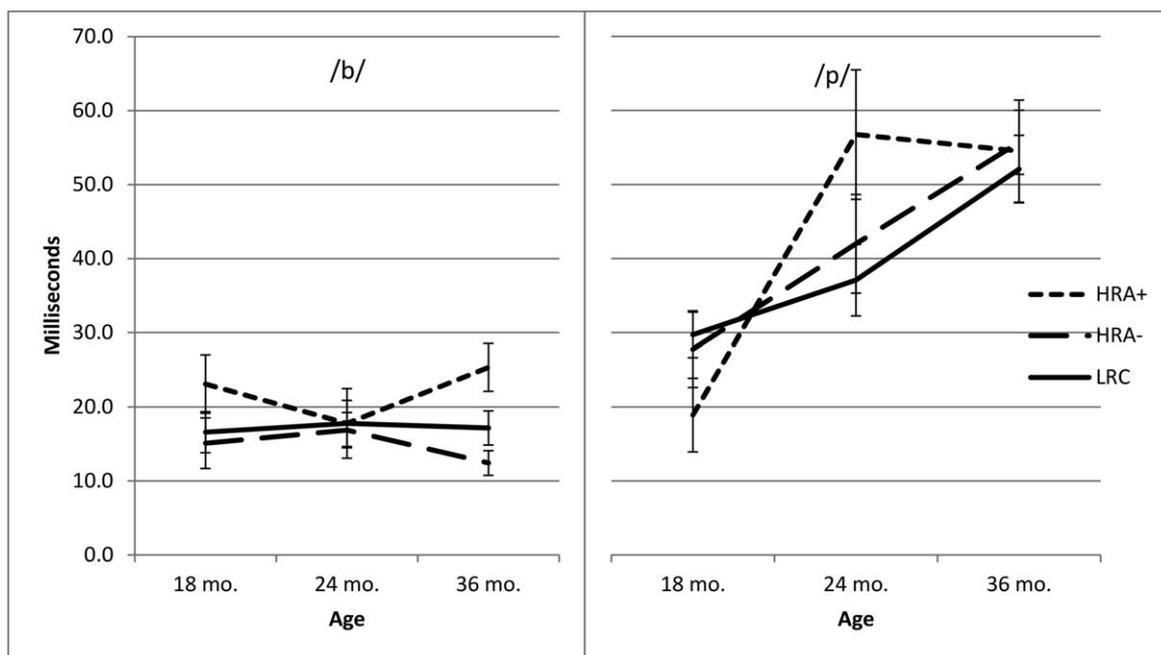


Figure 2. Mean voice-onset time, averaged within groups (HRA+, HRA-, LRC) for /b/ (left panel) and /p/ (right panel) at 18, 24, and 36 months. Bars indicate standard error of the mean.

children produced, a repeated measures ANOVA on mean /p/ VOT with group as a between-subjects factor was also performed. It showed that, as expected, there was a main effect of age ($F(2,40) = 11.4, P < .0005$), with mean VOT for /p/ significantly greater at 24 months than at 18 months ($P = .012$) and at 36 months than 24 months ($P < .0005$). There was no main effect of group and no age \times group interaction.

Figure 3 shows the within-child standard deviation for /b/ and /p/ by age and group for children who

produced at least three non-errored tokens of /b/ or /p/. A 3-way ANOVA using within-child S.D. of VOT as the dependent variable, age and group as between-subjects factors, and voicing status as a within-subject factor showed a significant main effect of voicing on within-child SD of VOT at 36 months ($F(1,238) = 10.937, P = .001$), but not 18 or 24 months. In addition, there was a significant main effect of age on within-child S.D. of VOT for /p/ ($F(2,238) = 5.780, P = .004$) but not for /b/. The age \times voicing interaction was significant,

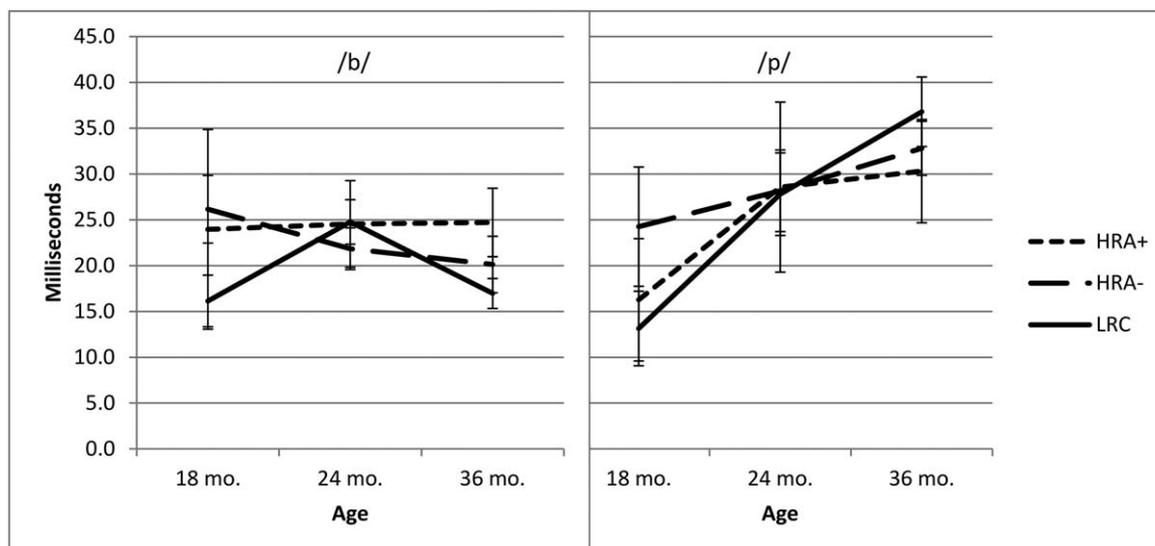


Figure 3. Within-child standard deviation of voice onset time, averaged within groups (HRA+, HRA–, and LRC) for /b/ (left panel) and /p/ (right panel) at 18, 24, and 36 months. Bars indicate standard error of the mean.

$F(4,238) = 4.035, P = .019$. Neither the age \times group nor the group \times voicing interactions were significant, and there was no significant three-way interaction.

Acoustic Distinctiveness Between /b/ and /p/

For toddlers producing at least three tokens of /b/ and at least three tokens of /p/ at each age (see Table 2) a *t*-test was used to determine whether their VOT populations for the two stop types were significantly different ($P < .05$). Thus, the *t*-test indicates whether the /b/ and /p/ populations are statistically distinct for the children in each group. A *P*-value $< .05$ indicates that the two VOT populations are statistically distinct, while a *P*-value $\geq .05$ indicates that the two VOT populations are not distinct—either because the means are close together, because the standard deviations are high, or both. The number of participants per group whose /b/ and /p/ VOT populations were statistically distinct was then entered into a binary logistic regression analysis performed at each age. Because there were (non-significant) between-group differences on number of tokens and number of errors, these variables were also entered into the logistic regression as covariates. The binary logistic regressions were performed twice: first, with risk status (HR vs. LR), EL, RL, and FM as predictors; then, with diagnostic group (HRA+, HRA–, LRC), EL, RL, and FM as predictors. Table 3 shows the number of participants at each age with distinct VOT populations.

When the predictor variables were EL, RL, FM, and risk status (HR vs. LR), results showed that no variables significantly predicted acoustic distinctiveness at any age (18, 24, or 36 months).

Table 3. Number of Toddlers with Acoustically Distinct /b/ and /p/ Populations

	LRC	HRA±	HRA–
18 months	1/4 (25%)	0/4 (0%)	1/10 (10%)
24 months	2/14 (14%)	2/8 (25%)	4/13 (31%)
36 months	13/19 (68%)	3/11 (27%)*	13/17 (76%)**

Predictors: Diagnostic status, EL, RL, FM. Covariates: number of tokens, number of errors. (See text for explanation.)

*HRA+ status significantly decreased odds of producing acoustically distinct /b/ and /p/ populations at 36 months ($P = .035$).

**HRA– status significantly increased odds of producing acoustically distinct /b/ and /p/ populations at 36 months ($P = .028$).

When the predictor variables were EL, RL, FM, and diagnostic group (HRA+, HRA–, LRC), only group predicted acoustic distinctiveness, and only at 36 months. Both HRA+ and HRA– status significantly predicted the ability to produce acoustically distinct /b/ and /p/ populations at 36 months. HRA+ status significantly decreased the odds of producing acoustically distinct /b/ and /p/ populations by a factor of 0.2 relative to LRC status, $P = .035$. HRA– status significantly increased the odds of producing acoustically distinct /b/ and /p/ populations by a factor of 1.4 relative to LRC status, $P = .028$.

Discussion

The aims of this study were to determine whether sub-perceptual differences in speech production exist in HRA+ toddlers, and whether any production differences are unique to the high-risk toddlers who develop

ASD or whether they are shared with HRA– toddlers, indicating a separable comorbidity. Language scores on the Mullen Scales 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). However, the HRA+ and HRA– group performed similarly, so group differences on consonant production cannot be ascribed to lower-than-average language ability on the part of the HRA+ or HRA– groups.

All three groups produced similar numbers of words, errors, and /b/ and /p/ tokens. High standard deviations in the figures for words and token counts in Table 2 reflect the range of volubility (“talkativeness”) found in the children in this study, the variation in choices of lexical items that children in each group produced, and the number of tokens per child that were acoustically analyzable. Some children in each group did not produce any words with /p/ or /b/ at some ages. The amount of variability encountered in spontaneous speech samples is always more than that encountered when using elicited samples. However, spontaneous speech samples provide a view of what children habitually produce, which elicited samples do not, and are valuable for that reason. All three groups also followed the typical development trajectory described by Macken and Barton [1980] and Lowenstein and Nittrouer [2008] in producing more /b/ tokens than /p/ tokens at all ages and more tokens of both with increasing age. There were also no significant differences in the rates of speech errors between groups.

Consistent with previous results, mean VOT for /b/ remained unchanged from 18 to 36 months for all groups. Mean VOT for /p/ increased over the course of the study, as expected; groups did not differ on this factor either. The present mean VOT values differ slightly from those of Lowenstein and Nittrouer [2008], but show the same age-related progression. Within-child S.D. of VOT for /b/ was constant with age but increased with age for /p/, as expected. Unlike in Lowenstein and Nittrouer [2008], no overall statistically significant difference was found between the within-child S.D. of VOT for /b/ and that for /p/. However, the values found in this study are consistent with a finding from Bailey and Haggard [1980] and Simon and Fourcin [1978] of an age-related production trend toward lower S.D.s of VOT for voiced stops. Methodological differences between the present study, in which participants were free to move about the room (as required for the ADOS), and previous studies, where participants were in high chairs in sound-proof booths, may account for the greater variability found in this study compared with previous work. In addition, words containing /b/s and /p/s were not selected ahead of time in this study;

analyses were post-hoc on spontaneously produced words. Finally, this analysis included both words produced in isolation and words produced in sentences, which result in different coarticulatory and lexical effects on their production. These factors likely led to greater variability in productions across children than if the words had been selected ahead of time.

Taken together, the word production rates, error rates, /b/-/p/ production rates, VOT values, and within-child SD of VOT values for /b/ and /p/ were similar across groups in this study and show that all groups generally followed the normal course of development. But these measures do not tell us all we need to know about the ability of the children in each group to produce the voicing contrast in stop consonants. The distribution of VOT values for /p/ might be distinct from the distribution of VOT values for /b/ for a particular group of children, but that does not mean that the same is true for each individual within that group. Instead, some children within a group might have distinct VOT populations for /b/ and /p/ (because of widely separated means, narrow standard deviations, or both), while others might have statistically indistinguishable /b/ and /p/ populations (because of narrowly separated means, wide standard deviations, or both). Regardless of the reason for overlapping /b/ and /p/ populations, the lack of a distinction between the two VOT populations suggests that a child may not yet have fully mastered the ability to produce the voiced/voiceless distinction or may be producing the distinction in a different way.

We found that neither expressive language, receptive language, fine motor ability, nor risk status (HR vs. LR) was significantly related to the likelihood that a child would produce acoustically distinct /b/ and /p/ populations at any age. Only HRA+ or HRA– status were significantly related to the probability that a child produced distinct /b/-/p/ populations at 36 months. In particular, HRA+ status was associated with a significantly lower probability of producing distinct /b/-/p/ populations at this age.

There are several contexts in which the current results can be interpreted. The catalyst/constraint framework of speech development, outlined in Green and Nip [2010], is one. In this framework, speech development is characterized by the facilitating effect of catalysts, defined as factors that spur change toward adult-like speech, and the limiting effect of constraints, defined as factors that temporarily interfere with a child’s ability to produce mature speech. For example, vocal imitation is a catalyzing mechanism that contributes to infants’ ability to produce the sounds of their native languages [Kuhl & Meltzoff, 1996]. Conversely, the inability of infants to control the tongue separately from the jaw is thought to be responsible for the

documented co-occurrence patterns of front vowels with alveolar consonants and back vowels with velar consonants in pre-speech babble [Davis & MacNeilage, 1995]. Because catalysts and constraints interact with each other in a dynamic fashion throughout development, skill plateaus or even regressions may occur. Green, Moore, Higashikawa, and Steeve [2000], for instance, found that 2-year-old children showed less coordination between upper and lower lip movement during speech than either 6-year-olds or 1-year-olds; this regression occurred during a period of rapid vocabulary expansion. In this context, then, the relative and unique difficulty that the HRA+ toddlers showed in producing distinct /b/ and /p/ populations at 36 months could be viewed as a delay or plateau, associated with the presence of a developmental constraint that was not identified in the current study. For example, deficits in imitation are characteristic of children with autism [Edwards, 2014; Rogers, 1999], and this may affect their ability to acquire the voicing contrast. Similarly, Lepistö et al. [2006] identified decreased cortical responses to duration changes in speech sounds in children with ASD and normal language. This difference in speech perception ability might also have affected the ability of the HRA+ participants in the current study to acquire the voicing contrast.

An alternative interpretation comes from the work of Goffman [2010], who found that children with specific language impairment (SLI) show higher variability of oral movement in multi-movement sequences than age peers, though whether the increased variability represents immaturity or an actual disorder in speech motor performance in children with SLI is not known. From this perspective, then, a diagnosis of ASD could be associated with subtle difficulty in producing the finely coordinated movements required for speech, realized in this study as a reduced likelihood of developing the ability to produce acoustically distinct /b/ and /p/ VOT populations at age three.

A final interpretation comes from Grigos, Saxman, and Gordon [2005], who investigated both acoustic and kinematic changes accompanying acquisition of the /b/-/p/ contrast. These researchers found that individual children varied in the degree to which their /p/ VOTs were evenly distributed around the mean. For some children, mean /p/ VOT values were influenced by occasional very large VOT values, but not so for other children. The authors conclude that different children employ different strategies to refine their production of /p/. That interpretation could easily be applied to the present results as well, suggesting that HRA+ toddlers as a group tended to adopt different strategies than the other two groups. The source of the difference is unclear, but may be related to imaging findings from Peeva et al. [2013]. They found that high-functioning (language-normal)

adults with ASD had impaired connections between the supplementary motor area (SMA) and the ventral premotor cortex (vPMC) in the left hemisphere, and suggest that this may be related to reduced or impaired speech output in ASD, even in the absence of language impairment. The current results are consistent with the idea that even when language development is normal in toddlers with ASD, speech production may differ from that of TD toddlers, and these differences may be related to differences in brain structure or organization in ASD.

Limitations and Future Work

There are several limitations to the current study. One is the relatively small sample size and the limited age range of the participants. In addition, the present study's focus on language-normal participants may mean that the current findings do not extend to more severely affected toddlers with ASD.

The findings from the present study should therefore be replicated in a larger population of high-risk infant siblings. One logical extension of the current work would be to examine whether parents or examiners ask for clarification more often from HRA+ toddlers compared with HRA- and LRC toddlers. If so, this would imply that a decreased likelihood of producing distinct /b/ and /p/ populations does in fact compromise intelligibility. Another extension would be to determine whether the relatively subtle between-group differences found here are mirrored by larger between-group differences when more severely affected toddlers with ASD are included. Yet another extension would be to investigate motor differences in bilabial stop production in high-risk toddlers using video facial movement analyses, as has been done for typically developing infants and toddlers [Green et al., 2000; Iuzzini-Seigel, Hogan, Rong, & Green, 2015]; and to investigate a larger range of ages. Finally, imaging studies could be performed to verify whether the acoustic differences found here were indeed associated with differences in brain organization in some children with ASD. Investigations such as these could reveal whether the current findings represent a temporary disruption during a time of increased challenge in other developmental domains, a delay associated with a diagnosis of ASD, or a motor disorder associated with ASD severity. All these investigations would add to our knowledge of how speech production is affected in ASD, regardless of whether a frank speech or language delay is also present, and would inform clinical practice with this population.

Acknowledgments

We thank Charles Nelson, co-PI of the Infant Sibling Project, and its staff members, past and present, for their hard work in collecting these data. We are also

deeply grateful for the effort of the dedicated families who have committed years of their lives to the Infant Sibling Project, making this work possible.

Grant Information: NIH R21DC08637, NIDCD 1R01DC010290-01, Simons Foundation 137186, Autism Speaks Pilot Grants Program 1323.

References

- American Psychiatric Association. (2013). American psychiatric association task force on DSM-5: Diagnostic and statistical manual of mental disorders: DSM-5. Washington, DC: APA.
- AoAMedia. (2013). AoA Audio Extractor: Free audio-extracting software. Retrieved March 27, 2016 from <http://www.aoa-media.com/audioextractor.htm>.
- Audacity. (2013). Free, open source, cross-platform software for recording and editing sounds. Retrieved March 27, 2016 from <http://audacity.sourceforge.net>.
- Bailey, P., & Haggard, M. (1980). Perception-production relations in the voicing contrast for initial stops in 3-year-olds. *Phonetica*, 37, 377–396.
- Bedford, R., Pickles, A., & Lord, C. (2015). Early gross motor skills predict the subsequent development of language in children with autism spectrum disorder. *Autism Research*, 9, 993–1001.
- Boucher, J. (1976). Articulation in early childhood autism. *Journal of Autism and Childhood Schizophrenia*, 6, 297–302.
- Bartolucci, G., Pierce, S., Streiner, D., & Eppel, P. (1976). Phonological investigation of verbal autistic and mentally retarded subjects. *Journal of Autism and Childhood Schizophrenia*, 6, 303–315.
- Bartolucci, G., & Pierce, S. (1977). A preliminary comparison of phonological development in autistic, normal, and mentally retarded subjects. *British Journal of Disorders of Communication*, 12, 137–147.
- Bartak, L., Rutter, M., & Cox, A. (1975). A comparative study of infantile autism and specific development receptive language disorder. I. The children. *British Journal of Psychiatry*, 126, 127–145.
- Cleland, J., Gibbon, F., Peppé, S., O'hare, A., & Rutherford, M. (2010). Phonetic and phonological errors in children with high functioning autism and Asperger syndrome. *International Journal of Speech-Language Pathology*, 12, 69–76.
- Davis, B., & MacNeilage, P. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research*, 38, 1199–1211.
- Diehl, J., & Paul, R. (2012). Acoustic differences in the imitation of prosodic patterns in children with high-functioning autism spectrum disorders. *Research in Autism Spectrum Disorders*, 6, 123–134.
- Diehl, J., & Paul, R. (2013). Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children with autism spectrum disorder. *Applied Psycholinguistics*, 34, 135–161.
- Diehl, J., Watson, D., Bennetto, L., McDonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high functioning autism. *Applied Psycholinguistics*, 30, 385–404.
- Edwards, L. (2014). A meta-analysis of imitation abilities in individuals with autism spectrum disorders. *Autism Research*, 7, 363–380.
- Forrest, K., & Rockman, B. (1988). Acoustic and perceptual analysis of word-initial stop consonants in phonologically disordered children. *Journal of Speech and Hearing Research*, 31, 449–459.
- Gamliel, I., Yirmiya, N., Jaffe, D., Manor, O., & Sigman, M. (2009). Developmental trajectories in siblings of children with autism: Cognition and language from 4 months to 7 years. *Journal of Autism and Developmental Disorders*, 39, 1131–1144.
- Goffman, L. (2010). Dynamic interaction of motor and language factors in normal and disordered development. In B. Maasen & P. van Lieshout (Eds.), *Speech Motor Control: New Developments in Basic and Applied Research*. Oxford: Oxford University Press.
- Green, J., & Nip, I. (2010). Some organization principles in early speech development. In B. Maasen & P. van Lieshout (Eds.), *Speech Motor Control: New Developments in Basic and Applied Research*. Oxford: Oxford University Press.
- Green, J., Moore, C., Higashikawa, M., & Steeve, R. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research*, 43, 239–256.
- Grigos, M., Saxman, J., & Gordon, A. (2005). Speech motor development during acquisition of the voicing contrast. *Journal of Speech, Language, and Hearing Research*, 48, 739–752.
- Hitchcock, E., & Koenig, L. (2013). The effects of data reduction in determining the schedule of voicing acquisition in young children. *Journal of Speech, Language, and Hearing Research*, 56, 441–457.
- Iuzzini-Seigel, J., Hogan, T.P., Rong, P., & Green, J.R. (2015). Longitudinal development of speech motor control: Motor and linguistic factors. *Journal of Motor Learning and Development*, 3, 53–68.
- Iverson, J., & Fagan, M. (2004). Infant vocal-motor coordination: Precursor to the gesture-speech system?. *Child Development*, 75, 1053–1066.
- Iverson, J., & Wozniak, R. (2007). Variation in vocal-motor development in infant siblings of children with autism. *Journal of Autism and Developmental Disorders*, 37, 158–170.
- Kent, R. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology*, 5, 7–23.
- Kjelgaard, M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes*, 16, 287–308.
- Koenig, L. (2000). Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research*, 43, 1211–1228.
- Kuhl, P., & Meltzoff, A. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, 100, 2425–2438.
- Landa, R., Gross, A., Stuart, E., & Bauman, M. (2012). Latent class analysis of early developmental trajectory in baby siblings of children with autism. *Journal of Child Psychology and Psychiatry*, 53, 986–996.

- Lepisto, T., Silokallio, S., Nieminen-von Wendt, T., Alku, P., & Naatanen, R., et al (2006). Auditory perception and attention as reflected by the brain event-related potentials in children with Asperger syndrome. *Clinical Neuropsychology*, 117, 2161–2171.
- Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384–422.
- Lowenstein, J., & Nittrouer, S. (2008). Patterns of acquisition of native voice onset time in English-learning children. *Journal of the Acoustical Society of America*, 124, 1180–1191.
- Lord, C., Risi, S., Lambrecht, L., Cook, E., Jr, Leventhal, B., DiLavore, P., et al. (2000). The Autism Diagnostic Observation Schedule - Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30, 205–223.
- McCann, J., Peppé, S., Gibbon, F., O'hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *Journal of Language & Communication Disorders*, 42, 682–702.
- McCleery, J., Tully, L., Slevc, L., & Schreiber, L. (2006). Consonant production patterns of young severely language-delayed children with autism. *Journal of Communication Disorders*, 39, 217–231.
- Macken, M., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, 7, 41–74.
- Mullen, E. (1995) Mullen scales of early learning. Circle Pines, MN: American Guidance Service, Inc.
- Peeva, M., Tourville, J., Agam, Y., Holland, B., Manoch, D., & Guenther, F. (2013). White matter impairment in the speech network of individuals with autism spectrum disorder. *NeuroImage: Clinical*, 3, 234–241.
- Pennington, B., & Bishop, D.V.M. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology*, 60, 238–306.
- Rapin, I., Dunn, M., Allen, D., Stevens, M., & Fein, D. (2009). Subtypes of language disorders in school-age children with autism. *Developmental Neuropsychology*, 34, 66–84.
- Rogers, S. (1999) An examination of the imitation deficit in infancy. In Nadel, J. & Butterworth, G. (Eds.), *Imitation in infancy* (pp. 254–283). New York, NY: Cambridge University Press.
- Rutter, M., Bailey, A., & Lord, C. (2003). *Social communication questionnaire*. Los Angeles: Western Psychological Services.
- Sauer LeBarton, E., & Iverson, J. (2013). Fine motor skill predicts expressive language in infant siblings of children with autism. *Developmental Science*, 16, 815–827.
- Schoen, E., Paul, R., & Chawarska, K. (2011). Phonology and vocal behavior in toddlers with autism spectrum disorders. *Autism Research*, 4, 12–11.
- Seigel, B. (2004). *Pervasive developmental disorders screening test—2nd edition (PDDST-II)*. San Antonio, TX: Psychological Corporation.
- Shriberg, L. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research*, 36, 105–140.
- Shriberg, L., Paul, R., McSweeney, J., Klin, A., Cohen, D., & Volkmar, F. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger Syndrome. *Journal of Speech, Language, and Hearing Research*, 44, 1097–1115.
- Simon, C., & Fourcin, A. (1978). Cross-language study of speech pattern learning. *Journal of the Acoustical Society of America*, 63, 925–935.
- Sjölander, K., & Beskow, J. (2011) Wavesurfer: An open-source speech tool. *Proceedings of the Sixth International Conference on Spoken Language Processing* 4, 464–467.
- Stevens, K. (1998) *Acoustic phonetics*. Cambridge, MA: MIT Press.
- Tager-Flusberg, H. (1981). Sentence comprehension in autistic children. *Applied Psycholinguistics*, 2, 5–24.
- Velleman, S., Andrianopoulos, M., Boucher, M., Perkins, J., Averback, K., Currier, A., et al. (2009). Motor speech disorders in children with autism. In R. Paul & P. Flipsen (Eds.), *Speech sound disorders in children: In honor of Lawrence D. Shriberg* (pp. 141–180). San Diego: Plural.
- Wang, M., Lekhal, R., Aaro, L., Holte, A., & Schjølberg, S. (2014). The developmental relationship between language and motor performance from 3 to 5 years of age: A prospective longitudinal study. *BMC Psychology*, 2, 34.
- Whiteside, S., Dobbin, R., & Henry, L. (2003). Patterns of variability in voice onset time: A developmental study of motor speech skills in humans. *Neuroscience Letters*, 347, 29–32.
- Whiteside, S., Henry, L., & Dobbin, R. (2004). Sex differences in voice onset time: A developmental study of phonetic context effects in British English. *Journal of the Acoustical Society of America*, 116, 1179–1183.