

The development of acoustic cues to coda contrasts in young children learning American English^{a)}

Jae Yung Song^{b)}

Department of Communication Sciences and Disorders, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53211

Katherine Demuth

ARC Centre of Excellence in Cognition and its Disorders, Centre for Language Sciences and Department of Linguistics, Macquarie University, Sydney, New South Wales 2109, Australia

Stefanie Shattuck-Hufnagel

Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 9 March 2011; revised 21 January 2012; accepted 26 January 2012)

Research on children's speech perception and production suggests that consonant voicing and place contrasts may be acquired early in life, at least in word-onset position. However, little is known about the development of the acoustic correlates of later-acquired, word-final coda contrasts. This is of particular interest in languages like English where many grammatical morphemes are realized as codas. This study therefore examined how various non-spectral acoustic cues vary as a function of stop coda voicing (voiced vs. voiceless) and place (alveolar vs. velar) in the spontaneous speech of 6 American-English-speaking mother-child dyads. The results indicate that children as young as 1;6 exhibited many adult-like acoustic cues to voicing and place contrasts, including longer vowels and more frequent use of voice bar with voiced codas, and a greater number of bursts and longer post-release noise for velar codas. However, 1;6-year-olds overall exhibited longer durations and more frequent occurrence of these cues compared to mothers, with decreasing values by 2;6. Thus, English-speaking 1;6-year-olds already exhibit adult-like use of some of the cues to coda voicing and place, though implementation is not yet fully adult-like. Physiological and contextual correlates of these findings are discussed. © 2012 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.3687467>]

PACS number(s): 43.70.Ep [AL]

Pages: 3036–3050

I. INTRODUCTION

Researchers have long noted that children's early word productions are variable in form, and different from those of adults (e.g., [Smith, 1973](#)). This has given rise to speculation regarding the nature of children's early phonological representations. Although much of what is known about phonological development comes from phonemic transcriptions of child speech, interpretation of transcription data can be challenging, because children sometimes make systematic acoustic contrasts that are not perceived as contrastive by adults. For example, [Macken and Barton \(1980\)](#) showed that some children below the age of 2 went through a period during which they made voicing distinctions for stops in word onsets, but this distinction relied on a non-adult-like voice onset time (VOT) difference that was entirely within the range for adult voiced stop onsets. Thus, children showed a contrast between voiced and voiceless onset stops that was not always perceptible for adults. Children also produce

contextually governed variation at an early age, even when they do not always overtly produce the conditioning context. For example, [Weismer *et al.* \(1981\)](#) found that language-delayed children produced longer vowels before voiced compared to voiceless target codas, even when they did not produce the target consonant closure and release. [Song and Demuth \(2008\)](#) explored these issues further with typically developing 1-2-year-olds. They found that children systematically lengthened the preceding vowel in tokens where they did not produce an audible coda consonant (e.g., *dog* [dɔ̃g] < [dɔ̃ɪ]), providing evidence for a coda consonant representation. The presence of such acoustic "covert contrasts" (e.g., [Scobbie *et al.*, 2000](#)) raises questions about children's emerging phonological representations, and when and how their speech productions begin to assume the same acoustic-phonetic realizations as those produced by adults.

In the present study, we were interested in how children's phonological representations develop over time, and the extent to which detailed acoustic analyses can help to reveal what they know, as well as how adult-like their phonetic implementations are. In particular, we focused on the development of voicing (voiced vs. voiceless) and place of articulation (PoA) (alveolar vs. velar) contrasts in coda stops. It is reported that coda consonants are typically acquired later than onset consonants, and less is known about the acoustics of codas. Coda consonants are of particular interest because

^{a)}Portions of this work were presented at the 157th Meeting of the Acoustical Society of America in Portland, OR (2009), the Child Phonology Conference in Austin, TX (2009), Generative Approaches to Language Acquisition (GALA) in Lisbon, Portugal (2009), and the 161st Meeting of the Acoustical Society of America in Seattle, WA (2011).

^{b)}Author to whom correspondence should be addressed. Electronic mail: songjy@uwm.edu

many inflectional morphemes in English, such as the past tense morphemes *-t/d*, appear in coda position, sometimes creating a complex coda cluster at the end of a word (e.g., *kicked* /kɪkt/). Thus, a study of the acoustics of monomorphemic codas will provide a baseline for future exploration of how and when morphemic coda consonants are acquired.

Studies of adult speech production have suggested that there are many potential cues to stop coda voicing, including the duration of the preceding vowel, presence and duration of a voice bar (i.e., low-frequency periodicity indicating continued vocal fold vibration after oral closure), and the presence and amplitude of aspiration noise produced at the vocal folds after the oral release (Cole *et al.*, 2007; Lisker and Abramson, 1964; Repp, 1979; Wright, 2004). It has also been shown that there are various acoustic correlates to PoA in adult speech, including spectral cues during vowel formant transitions and the spectrum and number of release bursts (Blumstein and Stevens, 1979; Olive *et al.*, 1993). However, only limited information is available about the acoustic correlates of voicing and PoA contrasts for stops in *children's* speech, and what research there is has primarily focused on stops in word-initial, *onset* position. Results of these studies are reviewed below, before turning to the question of children's stops in word-final, *coda* position.

VOT is one of the primary cues to voicing contrasts in onset stops. A number of studies have described children younger than 3–4 years undergoing at least three developmental stages before they produce adult-like VOT (Bond and Wilson, 1980; Kewley-Port and Preston, 1974; Macken and Barton, 1980; Zlatin and Koenigsknecht, 1976): the first stage, where children's undifferentiated VOT values fall within the short lag region that adults use for voiced stops in English; the second stage, where children make VOT distinctions between voiced and voiceless onsets, but both sets of VOT values generally fall within the adult perceptual boundaries for English voiced stop; and the final stage, where children's VOT values become more adult-like, separating into a short lag region for voiced stops and a long lag region for voiceless ones. More recently, Imbrie (2005) showed that children between the ages of 2;6–3;3 had a significantly longer VOT-lag for voiced stops than adults, and that this decreased to more adult-like values over the next 6 months. These findings provide evidence that 2–3-year-olds are still developing appropriate timing and glottal adjustments for onset voicing distinctions.

On the other hand, children appear to produce another cue to the voicing contrast, the vowel duration difference associated with coda voicing, earlier in life. In adult English, the effect of final consonant voicing on the duration of the preceding vowel is strong, with vowels preceding voiced consonants being almost twice as long as vowels preceding voiceless consonants (House, 1961), at least in utterance-final position (Crystal and House, 1988). English-learning children are known to produce the vowel duration cue to coda voicing before the age of three (Buder and Stoel-Gammon, 2002; Krause, 1982). Recent research looking at spontaneous child speech has found this distinction to be in place even before the age of 2 (Ko, 2007).

In a study exploring additional acoustic cues to coda voicing, Shattuck-Hufnagel *et al.* (2011) examined CVC (consonant-vowel-consonant) productions from two children from the Imbrie Corpus (Imbrie, 2005) who were aged 2;5 and 3;2 at the beginning of the study. The target words analyzed contained either velar codas (*bug* vs. *duck*) or bilabial codas (*tub* vs. *cup*), and the data were coded for the presence vs. absence of five acoustic cues that have been associated with the voicing contrast in coda stop consonants in adults. The results showed that both children exhibited systematic cues to coda voicing contrasts. That is, a voice bar during stop closure and an epenthetic vowel after the release appeared more frequently for voiced codas, whereas noise at the end of the vowel and noise after the coda release were produced more frequently for voiceless codas. The fifth cue, glottalization at the end of the vowel, which is more common before voiceless codas in adult speech, did not occur often enough to reveal a difference with respect to the voicing feature of the coda in this study, perhaps because their target words did not include coda /t/, which is the most common context for such coda-related glottalization in adult speech. This fine-grained acoustic analysis based on the feature-cue approach to the signaling of phonological contrasts (Stevens, 2002; Keyser and Stevens, 2006) provided detailed information about the acoustic correlates of the voicing feature in child speech. However, since the study only looked at children, it was not possible to determine precisely how these cue values differed from those of adults.

In contrast to the acoustic studies of voicing discussed above, research on children's development of PoA contrasts have been largely perception- and transcription-based. There is therefore limited understanding of how PoA features are realized acoustically in child speech. For example, Irwin (1947) examined the early emergence of consonants with different PoA using phonological transcriptions; during the first months of life, it was glottals (i.e., /h/) that appeared most frequently, followed by velars. By the age of 2;5, the frequency of glottal consonants gradually dropped, and the frequency of consonants produced at other PoAs, such as alveolars and labials, increased. In one of the few acoustic studies, Imbrie (2005) examined various cues to PoA in children's onset stops, including VOT, burst duration, and number of bursts.

Transcriptional studies have noted that some phonological processes commonly observed in early speech, such as velar fronting (e.g., *go* [g_o] → [d_o]) (Inkelas and Rose, 2007; McAllister, 2009) and consonant harmony (e.g., cat [kæt] → [kæŋ]) (Pater and Werle, 2003), involve changes in consonant PoA. Since children often make such PoA "errors," one might assume that their representation of PoA features is incomplete or non-adult-like. However, an extensive body of psycholinguistic literature on categorical perception shows that infants below the age of 5 months are sensitive to acoustic variations that define various phonological features, such as voicing, manner, and PoA (Eimas, 1974; Eimas and Miller, 1980; Eimas *et al.*, 1971). By the end of the first year of life, infant sensitivity to such acoustic cues has been refined to reflect the phonological structure of the ambient language (Kuhl *et al.*, 1992; Werker and Tees,

1984). Although infants below the age of 2 have been shown to be less successful in demonstrating this sensitivity when the tasks of phonetic discrimination involve word processing or word learning that require attention to meaning (Stager and Werker, 1997), more recent studies demonstrate that infants as young as 19 months show linearly graded sensitivity to the number of feature mismatches when appropriate tasks were used (White and Morgan, 2008). That is, the time that infants spent looking at a picture of the target (e.g., *shoe*) decreased as the degree of mismatch in features increased [e.g., *shoe* (correct pronunciation) > *foo* (change in PoA) > *vo* (change in PoA and voicing) > *goo* (change in PoA, voicing, and manner)].

In sum, the previous literature suggests that children below the age of 2 may have sophisticated phonological knowledge about voicing and PoA contrasts in words, even before they can reliably produce them in an adult-like way. One source of information about what a child knows is the pattern of acoustic cues that the child produces; that is, the child may indicate knowledge of a phonemic contrast by producing some of the cues that an adult uses, before developing the ability to produce the full pattern of adult-like cues. However, much is still to be learned about the development of acoustic cues to voicing and PoA contrasts in early child speech, especially with respect to codas.

The primary goal of the present study was, therefore, to provide a more comprehensive understanding of how and when certain acoustic correlates of voicing and PoA contrasts are acquired for stops in word-final coda position. Specifically, we wanted to know (a) which of the possible cues children produce early in acquisition, (b) whether children's production of the cues initially differs from that of adults, and if so, how and when do they become more adult-like, and (c) what the implications of these findings are for the development of phonological representations more generally.

To address these issues, we conducted an acoustic analysis of spontaneous speech productions collected from 6 mother-child dyads speaking American English in a longitudinal study. In particular, we investigated a selected set of *individual non-spectral acoustic cues* to the voicing and PoA features in stop codas, rather than simply noting whether an adult listener heard the target segment or not. This approach grew out of Stevens' (2002) model of speech production and perception based on individual acoustic cues to distinctive features. This model proposes that a given feature contrast may be signaled by a number of different acoustic cues, and that the precise set of cues that a speaker employs may vary depending on the other features in the feature bundle, as well as on the segmental and structural context in which the feature occurs. Performing analyses at the level of acoustic cues therefore provides a much richer and more systematic constellation of observations than performing analyses at the level of the segment (or even the feature) alone. Specifically, it has the potential to capture critical details about phonetic variation patterns in children's productions, how these develop over time, and how these may differ from those of adults.

Study 1 examined cues to the voicing contrast, and Study 2 examined cues to PoA contrasts. The values for

these cues were averaged across participating adults (mothers) and children. For some cues, we examined the average duration or number of cues (e.g., number of release bursts) as a function of voicing and PoA (e.g., the average duration of the vowel before voiced vs. voiceless codas in adult and child speech), and for other cues, we examined the frequency of the presence of a cue as a function of coda voicing and PoA (e.g., the frequency of occurrence of the voice bar during the closure of voiced vs. voiceless codas in adult and child speech).

Although we hypothesized that these cues would vary systematically with voicing and PoA in both mothers' and children's speech, we also expected some differences between the two populations. For example, Imbrie (2005) examined the development of the acoustic patterns of onset stop consonants by taking various durational, amplitude, spectral, formant, and harmonic measurements on 1049 utterances produced by ten 2;6–3;3-year-old children. The acoustic analysis revealed high variability in many of the measures including VOT, F0, intensity, and F2 at vowel center, suggesting that children were overall less consistent than adults in controlling and coordinating their articulatory gestures, vocal fold stiffness, and respiration. In addition, she argued that their smaller articulator size and high subglottal pressure were probably responsible for the high incidence of multiple release bursts for stops. In that study, some aspects of the children's speech became more adult-like over the 6 month observation period, but their gestures were still far from producing adult-like acoustic patterns at the end of the study, when the children were 3;0–3;9. In the present study, we examined the speech of younger 1;6–2;6-year-olds. Since coda consonants are typically acquired later than onsets, we predicted that the codas of these younger children would exhibit even more inconsistent use of some of the cues when compared to the productions of adults. In addition, we expected, based on Imbrie's (2005) observations (e.g., multiple release bursts), that some cues would have exaggerated values, but that the children would move toward more adult-like use of these cues by the end of the study period.

II. STUDY 1: THE DEVELOPMENT OF NON-SPECTRAL ACOUSTIC CUES TO VOICING CONTRASTS IN STOP CODAS

A. Method

1. Subjects and database

The data examined in this study came from the Providence Corpus (Demuth *et al.*, 2006), a collection of spontaneous speech interactions between 6 mother-child pairs from the New England Area.¹ All 6 children (3 boys, 3 girls) were typically developing, monolingual speakers of American English. The parents of 2 children spoke the dialect typical of Southern New England, which is often characterized by the omission of postvocalic /t/. The parental input for the other 4 children more closely resembled Standard American English.

Digital audio/video recordings were collected in the children's homes, approximately 1 h every 2 weeks for 2 years. Recording started between the ages of 0;11–1;4, depending on

when each child started producing words. During the recording, both mother and child wore a wireless Azden WLT/PRO VHF lavalier microphone pinned to their collar as they engaged in everyday activities. The recordings were made using a Panasonic PV-DV601D-K mini digital video recorder. The audio from the video was later extracted and digitized at a sampling rate of 44.1 KHz. Both the mothers' and children's speech were orthographically transcribed using Codes for the Human Analysis of Transcripts (CHAT) conventions (MacWhinney, 2000). The children's utterances were also transcribed by trained coders using International Phonetic Alphabet (IPA) transcription, showing the phonemic representations of words and the position of stressed syllables. Ten percent of the child data from each recording session were retranscribed by a second transcriber. Transcription reliability of overall coded segments ranged from 80% to 97% across files in terms of presence/absence of segments and place/manner of articulation. Voicing is difficult to reliably code in young children's speech, and was therefore not assessed in these reliability measures.

2. Data

We first extracted the highest frequency monosyllabic CVC content words in the Providence Corpus, selecting those ending in alveolar (/t/, /d/) and velar (/k/, /g/) stops. Words starting with glides or liquids (e.g., *red*) were excluded, due to the difficulty of identifying the beginning of vowels in such utterances for duration measures. The final set of target words contained either voiceless codas (*bat*, *cat*, *hot*, *boat*, *feet*, *hat*, *back*, *book*, *duck*,) or voiced codas (*bed*, *food*, *good*, *head*, *side*, *big*, *dog*, *pig*). We examined both child and adult productions of these words when the child was 1;5–1;7, 1;11–2;1, and 2;11–3;1, i.e., during time periods centered on 1;6, 2;0, and 2;6 years of age, respectively. This provided a reasonable number of tokens for each speaker and allowed us to explore developmental effects, both in the child speech and in that of their mothers during the same time periods.

We then extracted all the audio files of the sentences containing these target words. For individual mothers and children at each age, we coded the first 10 acoustically clean tokens of a target word in utterance-final position (e.g., *It's big*) and the first 5 acoustically clean tokens in each of four utterance-medial contexts: before consonant-initial words (e.g., *big spoon*), before glide-initial words (e.g., *big wagon*), before words beginning with a stressed vowel (e.g., *big apple*), and before words beginning with an unstressed vowel (e.g., *big as*). This provided some controlled variability for the medial context, and a maximum of 30 tokens per target word per speaker per age. Unusable tokens included those with poor acoustic quality, often because of overlap with other speaker's vocalizations or background noise. The final set of data included 2928 tokens.

As shown in Table I, the tokens analyzed were relatively evenly distributed across voiceless vs. voiced codas, ages, and positions within the utterance. The children had more utterance-medial tokens at later ages, probably due to the fact that utterance length increases with age. For both mothers

TABLE I. Number of tokens analyzed in Study 1.

	Utterance-final position						Utterance-medial position						Total
	Mothers		Children		Mothers		Children		Mothers		Children		
	1;6	2;0	2;6	1;6	2;0	2;6	1;6	2;0	2;6	1;6	2;0	2;6	
Voiceless	191	120	134	107	62	91	171	188	187	43	40	105	1439
Voiced	141	108	110	60	48	84	215	242	233	51	72	125	1489
Total	332	228	244	167	110	175	386	430	420	94	112	230	2928

and children, more than half of the utterance-medial tokens were followed by consonant-initial words (Mothers: 56%, Children: 61%). About a quarter of the utterance-medial words were followed by words beginning with an unstressed vowel (Mothers: 24%, Children: 18%). Target words followed by glide-initial words (Mothers: 11%, Children: 11%) and by words beginning with a stressed vowel (Mothers: 9%, Children: 10%) were the least common utterance-medial contexts. Importantly, the mothers and children showed similar distributions of utterance-medial words.

Table II further shows a breakdown of tokens contributed by each subject at different points in time. Overall, the numbers were similar among mothers and children respectively, suggesting that the amount of each participant's contribution to analyses was comparable. At the same time, one of the mother-child dyads (Mother 4 and Child 4) had a greater number of tokens compared to other pairs; this is due to the fact that Mother 4 and Child 4 had denser corpora, with weekly recordings during the time of investigation. Child 1 was slower in language development and did not produce many words in the 1;6 and 2;0 samples.

3. Acoustic coding

We then examined the acoustics of each token, coding for individual cues to voicing. To this end, we developed a set of coding conventions using both visual information

TABLE II. Number of tokens contributed by each subject at each age.

	Utterance-final position			Utterance-medial position			Total
	1;6	2;0	2;6	1;6	2;0	2;6	
Mother 1	25	32	56	35	41	64	253
Mother 2	63	36	52	66	39	91	347
Mother 3	67	43	24	83	101	55	373
Mother 4	97	87	53	112	171	96	616
Mother 5	45	10	25	45	27	54	206
Mother 6	35	20	34	45	51	60	245
Total	332	228	244	386	430	420	2040
	1;6	2;0	2;6	1;6	2;0	2;6	Total
Child 1	0	6	65	0	0	33	104
Child 2	46	8	22	19	10	38	143
Child 3	20	30	21	1	33	24	129
Child 4	64	51	48	67	63	71	364
Child 5	12	7	9	2	4	36	70
Child 6	25	8	10	5	2	28	78
Total	167	110	175	94	112	230	888

from the spectrogram and waveform and auditory information. Acoustic coding was carried out by several trained coders using Praat (Boersma and Weenink, 2005) to display and label the speech tokens. To evaluate inter-coder reliability, 10% of the tokens (293) were randomly chosen at age 2;0 and re-labeled by a second coder. The average difference in vowel duration and duration of post-release noise was 8 ms ($SD = 12$) and 16 ms ($SD = 23$), respectively. Agreement for the presence or absence of each of the other 3 cues was good (voice bar: 91%; glottalization at the end of a vowel: 93%; post-release noise: 91%). Pearson r correlations between the measurements of the original and recoded data were over 0.74 for all measures. All correlations were significant ($p < 0.001$), suggesting high inter-coder reliability.

4. Measures

We investigated five non-spectral acoustic cues to coda stop voicing (Fig. 1), all of which have been observed in adult speech (House, 1961; Fant and Lindblom, 1961; Redi and Shattuck-Hufnagel, 2001; Repp, 1979; Wright, 2004). Our coding criteria were defined as follows: (1) Vowel duration: the interval between the onset and offset of clear F2 energy in the spectrogram. (2) Voice bar: continued periodicity after the abrupt drop in amplitude (particularly in F2) that signaled closure for the stop coda. This was often characterized by a low frequency, low amplitude signal with a simpler waveform, reflecting continued vocal fold vibration during the stop closure without radiation of higher-frequency components of the source. (3) Glottalization at the end of a vowel: irregular, creaky-sounding pitch periods at the end of the vowel. This can result from various mechanisms, including strong adduction of the vocal folds. (4) Post-release noise: substantial noise following the coda stop release. The post-release noise sometimes showed the characteristics of frication followed by those of aspiration, but often this distinction was not clear, and sometimes only one of these possibilities was observed. Since our goal was to determine the presence of post-release noise, not its source, we simply determined whether there was substantial post-

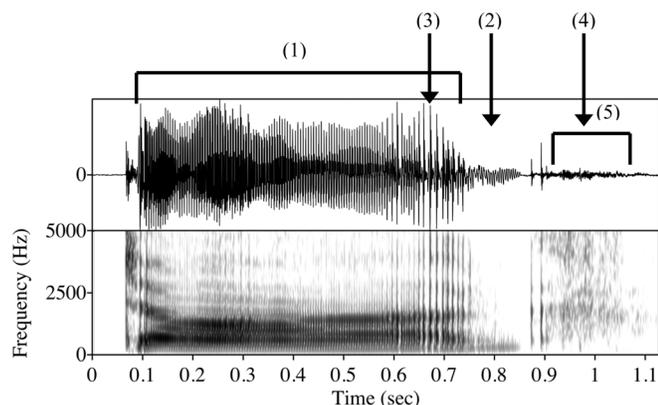


FIG. 1. Representative waveform and spectrogram for the word *dog* produced by a mother, illustrating the five acoustic cues examined: (1) vowel duration, (2) voice bar, (3) glottalization at the end of a vowel, (4) post-release noise, and (5) duration of post-release noise.

release noise or not. (5) Duration of post-release noise: the interval between the onset and offset of noticeable post-release noise. Note that some of these measures involve duration (1, 5), whereas others involve presence or absence of a cue (2, 3, 4).

5. Predictions

Previous studies have shown that preceding vowels are longer and voice bars occur more frequently in the context of tautosyllabic voiced stops than for voiceless stops (e.g., House, 1961; Fant and Lindblom, 1961). Furthermore, in American English vowel glottalization is known to occur more often before voiceless stops (especially /t/) (Redi and Shattuck-Hufnagel, 2001) due to adjustment of the vocal folds to interrupt regular vibration. In contrast to the vocal fold separation that accompanies voiceless stops in onset position, the adjustment for codas is hypothesized to involve strong approximation of the vocal folds, sometimes inducing irregular vibration (glottalization) at the end of the vowel. Past studies have also demonstrated that intraoral air pressure is greater during the production of voiceless stops than voiced stops (Bernthal and Beukelman, 1978; Lisker, 1970), which may contribute to more frequent and longer duration post-release noise.

We therefore predicted that mothers would exhibit (1) longer vowel duration before voiced codas, (2) more frequent occurrence of voice bar during the closure of voiced codas, (3) more frequent glottalization at the end of vowels before voiceless codas, and (4) more frequent and (5) longer post-release noise for voiceless codas. Children were predicted to produce many of the same cues as the mothers, but to exhibit more frequent and longer post-release noise, since they have been reported to generate greater intraoral air pressure than adults during the production of voiceless stops (Bernthal and Beukelman, 1978). We also expected to find some developmental changes, with children approximating more adult-like patterns with age.

B. Results

In order to examine how the five acoustic cues varied with the voicing of the coda, we employed a mixed-effects regression analysis, which incorporates both random and fixed effects. This analysis was particularly appropriate for our spontaneous speech corpus data, because of the flexibility with which it is capable of handling missing values as well as unbalanced numbers of word tokens in individual speakers. Furthermore, mixed-effects models offer the advantage of providing insights into the full structure of the data by examining fixed- and random-effects factors simultaneously (for further information, see Baayen, 2008; Baayen *et al.*, 2008; Quené and van den Bergh, 2004). In the present study, we wanted to capture idiosyncratic differences between speakers by treating them as a random-effect factor. We examined the mothers' and children's data separately, and the mothers and children in each analysis were treated as random effects.

In each of the mixed-effects regression models, the dependent variable was each of the five acoustic cues. Analyses

are carried out using the R statistical computing software (R Development Core Team, 2011). We used mixed-effects logistic regression models for the binary cues (i.e., presence or absence), and generalized linear mixed-effects models for the durational cues. Individual tokens from the speakers were used as individual data points to compute the frequency of occurrence or the average duration of each acoustic cue. This is possible because the mixed-effects regression models take into account the within- and between-speaker variances in the data. For the independent variables, the models included one random-effect factor, speakers (mothers or children), and three fixed-effect factors: voicing (voiced vs. voiceless), position of the target word within the utterance (utterance-medial vs. utterance-final), and the interaction between the two factors (voicing \times position). Earlier reports that individual cues can vary with the position of the target word in its utterance (e.g., Oller, 1973) motivated the consideration of utterance position.

In Secs. I–V below, we discuss detailed results of the models for each of the five cues to voicing. Table III shows the results of mixed-effects regression analyses for mothers and children at each age, where the effect of the random factor (speakers) was controlled. As regards the vowel-final glottalization cue in child speech, only the main effects of voicing and position were examined, because at each age there was one voicing \times position combination with no observation of glottalization, causing the odds ratio for the interaction to be undefined.

1. Preceding vowel duration

Figure 2 shows the average vowel duration (in ms) before voiceless and voiced stop codas in utterance-medial (dotted lines) and utterance-final (solid lines) positions. The thick lines for the children are above thin lines for the mothers in most conditions, showing that the children overall exhibited vowel duration several tens of milliseconds longer than the mothers (utterance-medially: children: 175 ms, mothers: 120 ms; utterance-finally: children: 255 ms, mothers: 220 ms). As shown in Table III, vowel duration, on average, varied as expected with voicing: it was greater before voiced than before voiceless codas for both mothers and children, but this main effect of voicing was caused by a significant interaction between voicing and utterance-position. That is, although vowel duration was reliably longer before voiced codas utterance-finally (solid lines), there was no such difference utterance-medially (dotted lines). This is consistent with previous findings in adult speech corpora showing a reliable effect of voicing on vowel duration only in utterance-final position (Crystal and House, 1988). Finally, the main effect of position was significant in both mothers and even the youngest children, with longer vowel durations utterance-finally compared to utterance-medially.

2. Presence of voice bar

As predicted, for both mothers and children there was a significant main effect of voicing on the presence of voice

TABLE III. Number of tokens from 6 mothers and 6 children, *t*-values, and the significance of the effect of each fixed factor (Note: $^{\circ}$ = marginally insignificant effect ($p = .05$ or $.06$), $* = p < .05$, $** = p < .01$, $*** = p < .001$). For *voicing*, the reference group was voiced codas; for *position*, the reference group was final codas.

	Mothers			Children		
	1;6	2;0	2;6	1;6	2;0	2;6
Number of tokens	718	658	664	261	222	405
<i>(1) Duration of preceding vowels</i>						
Voicing	-3.92***	-7.45***	-6.25***	-7.77***	-5.36***	-9.11***
Position	-8.35***	-15.95***	-12.92***	-5.72***	-8.42***	-11.17***
Voicing \times Position	1.62	6.04***	4.97***	4.57***	3.36***	5.90***
<i>(2) Presence of voice bar</i>						
Voicing	-10.50***	-8.83***	-10.05***	-6.23***	-5.46***	-6.05***
Position	1.01	0.71	-1.45	-1.45	0.80	0.14
Voicing \times Position	0.65	0.53	-0.67	0.36	2.08*	2.03*
<i>(3) Presence of glottalization at end of vowel</i>						
Voicing	4.29***	4.22***	4.20***	4.29***	1.13	1.06
Position	-3.67***	-2.95**	-2.21*	-3.79***	-3.74***	-5.15***
Voicing \times Position	0.43	-0.41	0.20	n/a	n/a	n/a
<i>(4) Presence of post-release noise</i>						
Voicing	-0.06	2.14*	2.56*	0.38	1.70	2.56*
Position	-10.45***	-8.94***	-7.70***	-5.16***	-5.67***	-6.02***
Voicing \times Position	2.32*	2.68**	0.79	1.91 $^{\circ}$	1.34	-0.28
<i>(5) Duration of post-release noise</i>						
Voicing	5.06***	4.01***	4.84***	-0.28	2.06*	3.23**
Position	-10.33***	-10.57***	-9.23***	-6.44***	-5.39***	-5.57***
Voicing \times Position	-3.66***	-2.53*	-3.50***	2.26*	-0.89	-2.21*

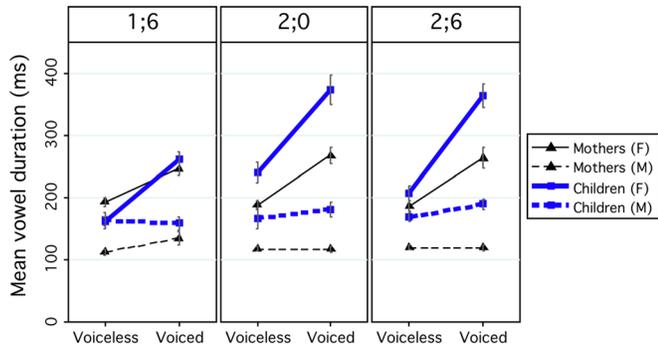


FIG. 2. (Color online) Average vowel duration (ms) before voiceless and voiced stop codas in utterance-medial (M) and utterance-final (F) position at each age. Error bars represent standard error.

bar, with more frequent occurrence of voice bars during the closure of voiced codas (Fig. 3). The probability of a voice bar was not significantly affected by the utterance position for either population. However, there were some significant interactions for the children at 2;0 and 2;6, with a greater difference in the percent voice bar between voiced and voiceless codas utterance-finally compared to utterance-medially. In sum, mothers and children overall showed similar patterns in using the voice bar to cue voicing contrasts, but differed in the effect of voicing \times position interaction.

3. Presence of glottalization at the end of the vowel

For mothers, the presence of vowel-final glottalization varied systematically as a function of voicing and position (Fig. 4); vowel-final glottalization was, on average, more frequent before voiceless than before voiced stops, and in utterance-final position compared to medially. The voicing \times position interaction, however, was not significant. Although the effect of voicing on this cue was significant for children at 1;6, surprisingly it was not significant at 2;0 and 2;6. Children produced vowel-final glottalization significantly more often utterance-finally than utterance-medially from 1;6, possibly because of a tendency toward final creak. As noted earlier, an examination of the interaction effect between voicing and position in medial position was not possible due to the sparse data problem in this position. In sum, glottalization varied systematically with position, but less systematically with voicing in early speech.

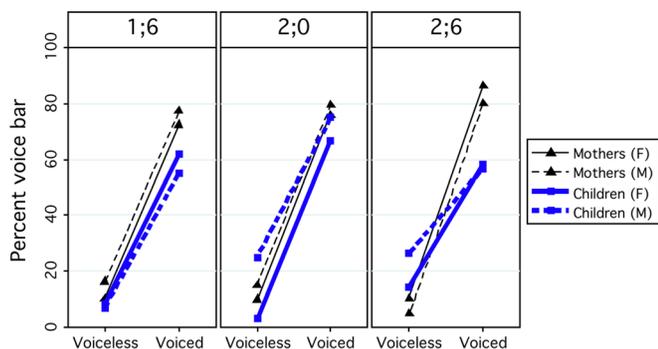


FIG. 3. (Color online) The percent of voice bar before voiceless and voiced stop codas in utterance-medial (M) and utterance-final (F) position.

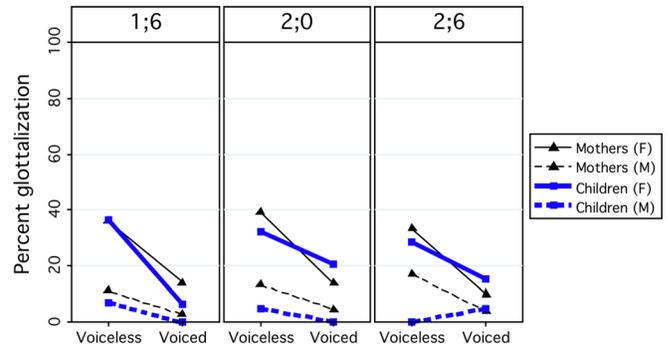


FIG. 4. (Color online) Percent glottalization at the end of the vowel before voiceless and voiced stop codas in utterance-medial (M) and utterance-final (F) position.

4. Presence of post-release noise

For mothers, there was a significant main effect of voicing on post-release noise when the children were 2;0 and 2;6, with more frequent occurrences for voiceless stops (Fig. 5). The lack of voicing effect at 1;6 appears to be due to the high incidence of post-release noise for both voiceless and voiced stops, especially in utterance-final position. This is supported by a significant voicing \times position interaction at 1;6 and 2;0, showing that the difference in percent post-release noise between voiceless and voiced tokens was larger in utterance-medial position than utterance-final position. Finally, the main effect of position was significant at all three ages, with more frequent occurrences of post-release noise utterance-finally compared to medially.

For children, the main effect of voicing was not significant until 2;6 (Fig. 5). The null effect at 1;6 and 2;0 appears to be due to the small difference between voiceless and voiced codas in utterance-final position. However, in medial position, post-release noise was more frequent for voiceless stops than for voiced stops from 1;6, as evidenced by a significant interaction between voicing and position. Children also showed a significant effect of position on this cue at all ages, with more frequent post-release noise in utterance-final position than medially. By 2;6 children decreased their use of post-release noise to near-adult levels in both utterance positions. In sum, mothers and children exhibited more frequent use of post-release noise for voiceless codas, but only in utterance-medial positions when the children were at earlier

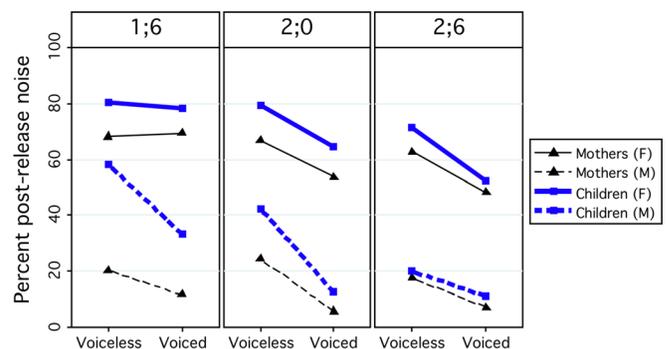


FIG. 5. (Color online) Percent post-release noise for voiceless and voiced stop codas in utterance-medial (M) and utterance-final (F) position.

ages. The lack of an overall voicing effect appears to be due to the frequent occurrence of post-release noise for both voiceless and voiced codas in utterance-final position.

5. Duration of post-release noise

For mothers, all three factors (voicing, position, voicing \times position) significantly affected the duration of post-release noise (Fig. 6). The significant main effect of voicing indicated that post-release noise was, on average, longer for voiceless codas than for voiced codas, but this effect was caused by a significant voicing \times position interaction. That is, the overall effect of voicing was due primarily to differences found in utterance-final position, and less to variation in utterance-medial position. This pattern is consistent with results described above for the duration of the preceding vowel, where voicing-related effects were found mainly in utterance-final position. Lastly, the main effect of position was also significant, with longer post-release noise in utterance-final position than medially.

Children generally showed similar patterns to those of adults, although the main effect of voicing was not yet significant at 1;6; voiced (78 ms) and voiceless (75 ms) codas had equally long durations of post-release noise in utterance-final position (Fig. 6). As was also the case for presence of post-release noise, the duration of post-release noise for voiced codas decreased over time in utterance-final position, revealing a distinction between voiceless and voiced codas from 2;0. In utterance-medial position, voiceless codas had a particularly long post-release noise duration at 1;6 (42 ms), but this decreased by 2;6 (10 ms) to narrow the gap between voiceless and voiced codas to insignificance.

In sum, unlike mothers, who showed longer post-release noise for voiceless than for voiced codas in utterance-final position and less variation utterance-medially, 1;6-year-olds exhibited equally long post-release-noise durations for voiceless and voiced codas in utterance-final position, and longer durations for voiceless than for voiced codas medially. By 2;6, post-release noise duration for voiced codas in utterance-final position and that for voiceless codas in utterance-medial position decreased to resemble adult values. Thus, although children showed a significant effect of voicing \times position interaction at 1;6 and 2;6, the interactions had opposite signs

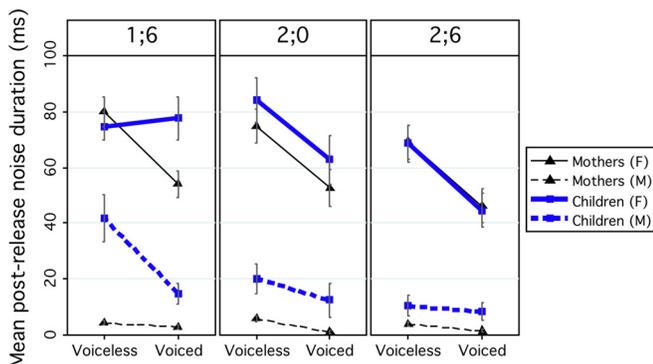


FIG. 6. (Color online) The average duration of post release noise (ms) for voiceless and voiced stop codas in utterance-medial (M) and utterance-final (F) position. Error bars represent standard errors.

at the two ages. This suggests that the post-release noise duration cue to voicing is still developing in early child speech, but that children achieve more adult-like abilities by 2;6.

C. Summary of Study 1

Mothers' speech consistently showed the predicted difference for all 5 cues to the coda voicing contrast, and children showed similar patterns for vowel duration and voice bar cues from 1;6 years. On the other hand, the presence and duration of post-release noise varied less systematically as a function of coda voicing in the children's speech at earlier ages, primarily due to no systematic variation with voicing in utterance-final position. Finally, the effect of voicing on vowel-final glottalization was not reliable in child speech until 2;6.

During our examination of the data in Study 1, we noticed that some of the cues to voicing also seemed to vary systematically with the PoA of the stop coda. Thus, although non-spectral cues are not traditionally thought of as cues to PoA (more often considered to involve formant transitions and release-burst spectra), their pattern of occurrence may contain useful information about PoA as well. Therefore, in Study 2, we examined the effect of PoA on three of the cues in Study 1 (presence of vowel-final glottalization, presence and duration of post-release noise) and two additional coda-release-related cues that have been suggested in the adult literature to vary systematically with PoA: the presence and average number of release bursts.

III. STUDY 2: THE DEVELOPMENT OF NON-SPECTRAL ACOUSTIC CUES TO PoA CONTRAST IN STOP CODAS

A. Method

1. Subjects and database

The subjects and database used in Study 1 were also used in Study 2.

2. Data

The same target word tokens used in Study 1 were also used in Study 2, but this time the words were classified by the PoA of the coda: eleven contained alveolar codas (*bat*, *cat*, *hot*, *boat*, *feet*, *hat*, *bed*, *food*, *good*, *head*, *side*) and six contained velar codas (*back*, *book*, *duck*, *big*, *dog*, *pig*). The distribution and number of tokens analyzed are shown in Table IV. Again, tokens were relatively evenly distributed

TABLE IV. Number of tokens analyzed in Study 2.

	Utterance-final position			Utterance-medial position			Total
	Mothers	Children		Mothers	Children		
	1;6	2;0	2;6	1;6	2;0	2;6	
Alveolar	190	138	156	77	48	96	149
Velar	142	90	88	90	62	79	237
Total	332	228	244	167	110	175	386

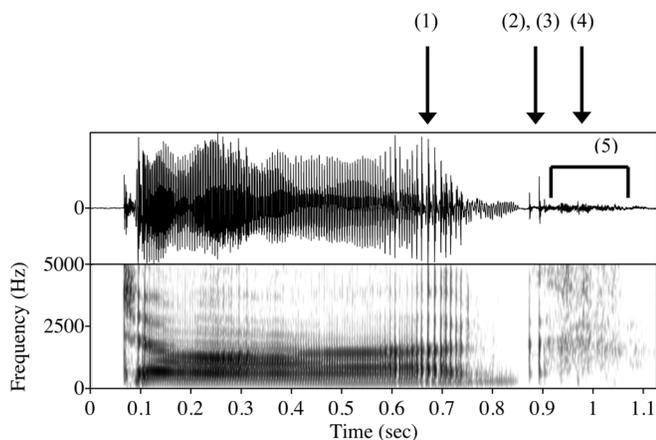


FIG. 7. Representative waveform and spectrogram for the word *dog* produced by a mother, illustrating the five cues examined in Study 2: (1) glottalization at the end of vowel, (2) release bursts, (3) number of release bursts, (4) post-release noise, and (5) duration of post-release noise.

across alveolar vs. velar PoA, ages, and positions within the utterance.

3. Acoustic coding

For three of the cues analyzed in Study 2 (presence of glottalization, presence and duration of post-release noise) the cue labels from Study 1 were used. For the two additional cues (presence and number of release bursts), coding conventions were developed and several trained coders coded the cues using Praat. To evaluate inter-coder reliability, ten percent of the total (293 tokens) was randomly chosen from sample produced at age 2;0 and re-transcribed by a second coder. The average difference in the number of release bursts and the duration of post-release noise was 0.39 ($SD = 0.77$) and 16 ms ($SD = 23$), respectively. The agreement for the presence or absence of each of the remaining 3 cues was good (release bursts: 90%; glottalization at the end of a vowel: 93%; post-release noise: 91%). Pearson r correlations between the measurements of the original and recoded data were over 0.74 for all measures. All correlations were significant ($p < 0.001$), suggesting high inter-coder reliability.

4. Measures

Figure 7 shows an example token produced with the five acoustic cues used in Study 2: (1) Glottalization at the end of a vowel: irregular, creaky-sounding pitch periods at the end of the vowel. (2) Release burst: at least one strong vertical spike in the waveform, signaling the transient at the abrupt release of a stop consonant. (3) The number of release bursts: individual occurrences of such release burst transients (two in Fig. 7, indicated by two sharp spikes in the waveform). This number was zero if there was no acoustic evidence for a coda release transient. (4) Post-release noise: noticeable noise following the stop release. (5) Duration of post-release noise: the interval between the onset and offset of noticeable post-release noise. Note that some of these cues examined frequency of occurrence (1, 2, 4), whereas the others were either numerical counts (3) or durational measures (5).

5. Predictions

As is well documented in the literature (e.g., Pierrehumbert and Frisch, 1994), American English alveolar stops (especially /t/) are often glottalized when syllable-final and unreleased, and often become flaps in certain intervocalic positions. Thus, we expected more frequent glottalization and less frequent oral closure releases for alveolar stops. In addition, multiple bursts are known to be common in the release of onset velar stops, whereas they are less frequent for alveolar stops (Olive *et al.*, 1993). These multiple bursts might occur in the following way. After the tongue dorsum contacts the palate, forming the velar closure, pressure builds up behind that closure, depressing the surface of the tongue in the small back cavity behind that articulatory seal. After the first release of the stop closure, rapid airflow between the tongue surface and the palate can result in lowering of that pressure. These changed aerodynamic forces (combined with tissue elasticity) may allow the surface of the tongue to move upward (Stevens, 2000; Hanson and Stevens, 2006). This upward movement can occur even though the mass of the tongue tissue is moving downward at the release due to muscular activity. We hypothesize that this can sometimes result in a second closure, leading to a second release burst and potentially more such events. Compared to alveolars, velar closure might provide conditions particularly favorable for these effects to operate; first, the tongue dorsum constriction for velars typically has a longer constriction area (front to back in the vocal tract) than alveolars, and second, the constriction area for velars may change less rapidly compared to alveolars, because the dorsum has a large mass that does not move away from the palate as quickly as the tongue tip (Keating *et al.*, 1980).

Based on these previous findings, we predicted that mothers would exhibit (1) more frequent vowel glottalization before alveolar stops, (2) more frequent occurrence of at least one release burst for velar stops, (3) a greater number of release bursts for velar stops, (4) more frequent and (5) longer post-release noise for velar stops. For children, we predicted that, in general, use of these individual cues would vary with PoA in the same way that their mothers' did. Building on findings from Study 1, we also expected a decrease in the production of some of the cues, including frequency and duration of post-release noise, as the children became older. This would be consistent with Imbrie's (2005) finding that the large number of release bursts for both alveolars and velars when children ranged from 2;6 to 3;3 years had decreased to more closely resemble adult values six months later.

B. Results

As in Study 1, we employed mixed-effects regression analyses. In each of these models, the dependent variable was one of the five cues. For the independent variables, the models included one random-effect factor, speakers (mothers or children), and three fixed-effect factors: PoA (alveolar vs. velar), position of the target word within the utterance (utterance-medial vs. utterance-final), and the interaction between the two factors ($PoA \times position$). Table V shows the results of the mixed-effects regression analyses for both the mothers

TABLE V. Number of tokens from 6 mothers and 6 children, t-values, and the significance of the effect of each fixed factor (Note: ° = marginally insignificant effect ($p = .05$ or $.06$), * = $p < .05$, ** = $p < .01$, *** = $p < .001$). For *PoA*, the reference group was alveolar codas; for the *position*, the reference group was final codas.

	Mother			Child		
	1;6	2;0	2;6	1;6	2;0	2;6
Number of tokens	718	658	664	261	222	405
<i>(1) Presence of glottalization at end of vowel</i>						
PoA	-4.34***	-3.39***	-1.88°	-4.42***	1.10	0.12
Position	-4.85***	-3.64***	-2.06*	-4.28***	-3.72***	-3.20**
PoA × Position	-0.49	-0.97	-2.05*	n/a	n/a	-0.15
<i>(2) Presence of release bursts</i>						
PoA	5.24***	4.30***	5.18***	4.03***	1.30	3.59***
Position	-7.18***	-8.72***	-6.25***	-2.24*	-1.19	-4.03***
PoA × Position	-0.91	1.52	-0.47	-0.94	-0.69	-0.57
<i>(3) Number of release bursts</i>						
PoA	7.32***	6.28***	9.84***	4.72***	0.58	6.50***
Position	-6.99***	-6.98***	-5.38***	-0.84	-1.30	-2.94**
PoA × Position	0.04	1.51	-2.36*	-0.84	0.23	-2.14*
<i>(4) Presence of post-release noise</i>						
PoA	5.40***	4.76***	4.80***	3.09**	1.87°	3.31**
Position	-8.04***	-4.82***	-6.34***	-3.36**	-3.04**	-4.80***
PoA × Position	-0.56	2.11*	1.23	-1.97*	-1.79	-1.68
<i>(5) Duration of post-release noise</i>						
PoA	6.59***	8.09***	8.41***	3.75***	3.16**	2.62**
Position	-10.81***	-9.78***	-9.91***	-4.26***	-3.71***	-5.90***
PoA × Position	-4.44***	-5.82***	-6.14***	-1.24	-1.72	-1.50

and children at each age, with t-values and the significance of the effect of each fixed factor (PoA, position, and PoA × position) when the effect of the random factor (speakers) was controlled for. For vowel-final glottalization in child speech at 1;6 and 2;0, only the main effects of voicing and position were examined because no glottalized tokens were observed in one of the voicing × position combinations.

1. Presence of glottalization at the end of the vowel

In the mothers' speech, glottalization at the end of the vowel varied systematically as a function of both PoA and utterance-position (Fig. 8). Glottalization was found more frequently before alveolars than before velars and when words occurred in utterance-final compared to utterance-

medial position (possibly because of a tendency to produce creaky voice quality in utterance-final position). However, the interaction between PoA × position was significant for mothers only when their children were 2;6, when glottalization was greater utterance-medially. For children, the effect of glottalization as a function of PoA was only significant at 1;6; the distribution of glottalization across PoA was not reliably different in child speech by 2;6. In contrast, like adults, children consistently showed more frequent use of glottalization in utterance-final position compared to utterance-medial position. The PoA × position interaction was not significant at 2;6, the only age condition examined.

2. Presence of coda release

For mothers, the likelihood of at least one coda release burst varied with both PoA and utterance-position (Fig. 9). As expected, velar stops were released more frequently than alveolar stops, and utterance-final stops were released more frequently than utterance-medial stops. There was no significant interaction between PoA and position. Like adults, children at 1;6 showed significant effects of PoA and position, although the effects were not significant at 2;0. This appears to be due to the interesting patterns found for tokens in utterance-medial position; at 1;6, children released velar codas more frequently than alveolar codas, even though the overall percent release for codas at both PoA was high. At 2;0, however, there was a big drop in percent release for velar stops, resulting in no difference between velar and

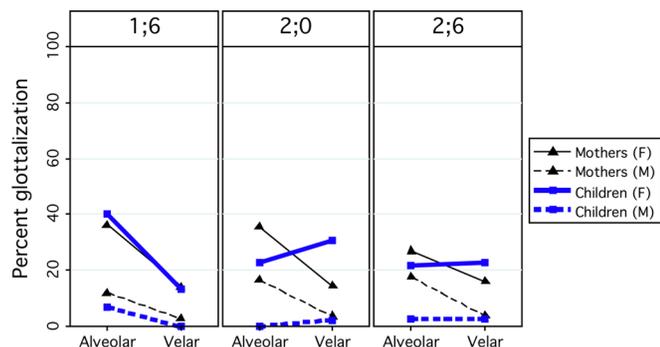


FIG. 8. (Color online) Percent glottalization at the end of vowel before alveolar and velar stop codas in utterance-medial (M) and utterance-final (F) position.

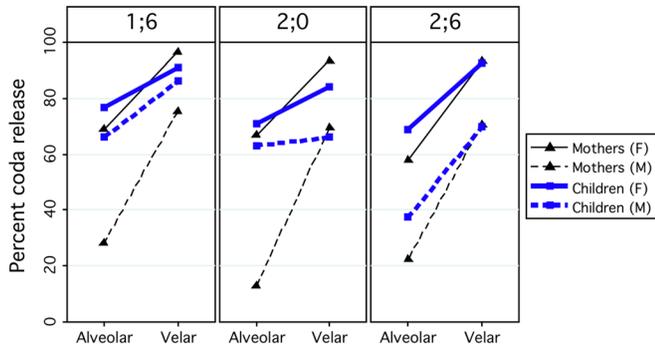


FIG. 9. (Color online) Percent coda release for alveolar and velar stop codas in utterance-medial (M) and utterance-final (F) position.

alveolar codas in medial position. This was followed by a big drop in the percent release for alveolars at 2;6, making the effect of PoA significant again in medial position. Thus, percent release for alveolars and velars in utterance-medial position decreased in a stepwise fashion from 1;6 to 2;6; the value for velars decreased to adult levels first, and only later did the value for alveolars decrease. This pattern seems to be a characteristic of the emergence of adult-like distributions of coda release cues to PoA contrasts in these data.

3. Number of release bursts

In the mothers' speech, the average number of release bursts varied systematically as a function of PoA and position (Fig. 10); the average number of release bursts was greater for velar codas than for alveolars, and in utterance-final compared to utterance-medial position. The interaction between PoA \times position was significant only when their children were 2;6, with a greater difference between alveolars and velars in utterance-final position. Children also had a greater number of release bursts for velars than for alveolars from 1;6, though the difference was not significant at 2;0. Thus again we found a stepwise decrease in the number of release bursts between 1;6 and 2;6, with a larger decrease for velars at 2;0 and then subsequently for alveolars at 2;6. Although velars had a greater number of bursts than alveolars at 1;6, the number of bursts for velars decreased to a level similar to that of alveolars at 2;0, leading to no effect of PoA at this age. The effect of utterance position, which was significant in mothers' speech, was not significant until

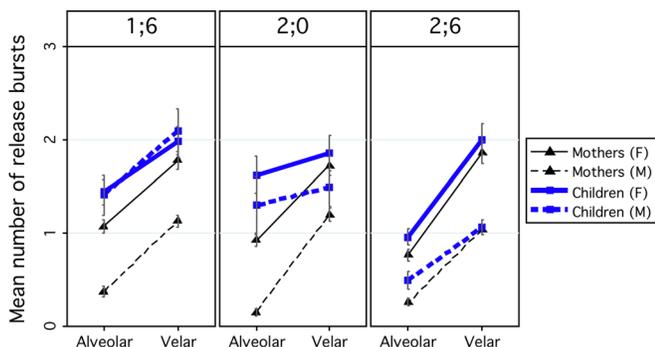


FIG. 10. (Color online) Mean number of release bursts for alveolar and velar stop codas in utterance-medial (M) and utterance-final (F) position. Error bars represent standard errors.

2;6 in the children's speech due to the greater number of release bursts utterance-medially at 1;6 and 2;0. The average number of release bursts in the child speech overall decreased by 2;6, especially in utterance-medial position, more closely approximating adult values for both alveolar and velar codas. Just as for the mothers, the PoA \times position interaction was significant only at 2;6 in child speech, with greater differences utterance-finally.

4. Presence of post-release noise

For both children and mothers, the presence of post-release noise varied as a function of PoA and position (Fig. 11). Both populations showed more frequent production of post-release noise for velar stop codas than for alveolar stop codas, and for tokens in utterance-final compared to utterance-medial position. In general, the PoA \times position interaction was not significant, except at 2;0 for mothers and at 1;6 for children.

Percent post-release noise was higher for children than for mothers at 1;6 and 2;0. As seen above for the presence and average number of release bursts, there was a selective decrease in percent post-release noise as a function of coda PoA. In utterance-medial position, the decrease from 1;6 to 2;0 was bigger for velars than for alveolars, followed by a bigger decrease for alveolars at 2;6. Thus, the originally exaggerated values for percent post-release noise in the child speech decreased to adult values by 2;6, but in a stepwise fashion rather than simultaneously for both PoA. In utterance-final position, children showed greater percent post-release noise than mothers only for alveolars, and this decreased to near-adult levels by 2;6. In sum, although the presence of post-release noise varied systematically as a function of PoA in both children's and mothers' speech, the frequency of post-release noise was overall greater for children, decreasing to more adult-like values by 2;6, with the change in this value for velars preceding the change for alveolars.

5. Duration of post-release noise

For mothers, the duration of post-release noise was, on average, longer for velar stop codas than for alveolar stop codas (Fig. 12). However, as indicated by a significant interaction between PoA and position, the main effect of PoA

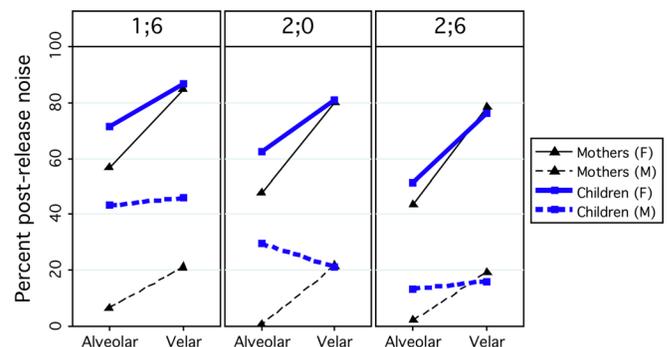


FIG. 11. (Color online) Percent occurrence of post-release noise for alveolar and velar stop codas in utterance-medial (M) and utterance-final (F) position.

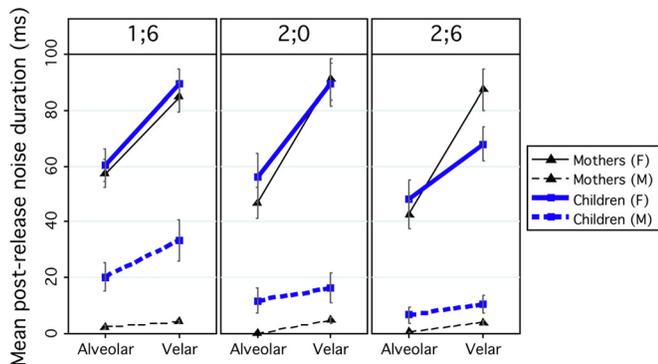


FIG. 12. (Color online) The average duration of post release noise (ms) for alveolar and velar stop codas in utterance-medial (M) and utterance-final (F) position. Error bars represent standard errors.

was due primarily to differences in post-release noise duration between alveolars and velars in utterance-final position. This is in line with the findings for other durational measures that were obtained for the voicing contrast in Study 1. The main effect of position was also significant in mothers' speech, with longer duration of post-release noise for tokens in utterance-final compared to utterance-medial position. The duration of post-release noise varied as a function of both PoA and utterance-position in children's speech as well. However, the interaction between PoA and position was not significant in child speech. In addition, children produced longer post-release noise in utterance-medial position at earlier ages, with the duration decreasing to near-adult levels by 2;6.

C. Summary of Study 2

Mothers' speech consistently showed systematic variation in all five acoustic cues with PoA in the predicted manner. All of these cues also appeared more often or were longer in duration in utterance-final position. Similarly, for the children, most of the acoustic cues showed a generally adult-like distribution with respect to PoA contrasts from 1;6, except for vowel-final glottalization. However, children had more frequent releases, a greater number of release bursts, and more frequent and longer post-release noise at 1;6. These exaggerated values decreased by 2;6, approximating more adult-like values.

IV. DISCUSSION

In the present study we found that children as young as 1;6 produced many of the same types of acoustic cues to stop codas as adults, aligning them appropriately with voicing and place contrasts. However, the range of values for those acoustic cues, and their degree of systematicity, were sometimes strikingly different from that of adults. Drawing on previous findings on the acquisition of onset consonants reviewed in Sec. I, we briefly explore below some of the possible articulatory and other factors that may contribute to these developmental changes in coda position.

First, glottalization at the end of the vowel varied with voicing and PoA less systematically in children's speech.

Previous studies have shown that the size and internal structure of the vocal folds are not adult-like until adolescence (Hirano *et al.*, 1983). Moreover, various pieces of evidence suggest that young children have not yet acquired adult-like control over the larynx. For example, Koenig (2000) found substantial variability in VOT for /h/ production (voiceless portion of an /h/ measured from peak air flow to onset of following voicing) in 5-year-olds as compared to adults. She concluded that this was due to incomplete control of laryngeal factors such as adduction degree and vocal fold tension, and was not merely due to interarticulator timing skill (i.e., how accurately the adduction gesture is timed with respect to other articulatory events). This study suggests that young children have immature control of the larynx, resulting in less consistent performance than that observed in adults. Such immature laryngeal anatomy and control might also explain why the 2-year-olds in our study exhibited inconsistent use of vowel glottalization as a function of voicing and PoA. In addition, these children produced only a small number of tokens with vowel-final glottalization, which made it hard to draw a reliable conclusion about the effect of voicing and PoA.

Second, the probability and duration of post-release noise did not differ with coda voicing for children at 1;6, but by 2;6 these cues occurred more with voiceless codas, in parallel with the adult results. A close examination of the data revealed that at 1;6, children produced frequent and long post-release noise for both voiceless and voiced codas in *utterance-final* position. However, their use of post-release noise for utterance-final voiced codas decreased by 2;6, differentiating voiced from voiceless codas in this position. Researchers have often observed that young children tend to devoice coda consonants (Smith, 1979; Velten, 1943). Given that post-release noise appears more often with voiceless codas in adult speech, it is probable that the post-release noise occurring with voiced codas in young children's speech was one of the acoustic cues that gave researchers the percept of voiceless codas. On the other hand, in utterance-medial position, even 1;6-year-olds exhibited more frequent and longer post-release noise for voiceless coda stops than for voiced stops; although their use of post-release noise overall decreased by 2;6 in utterance-medial position just as in utterance-final position, they maintained the voiceless-voiced contrast for this cue. The fact that 1;6-year-olds showed systematic differences in post-release noise between voiced vs. voiceless codas in at least one of the utterance positions (i.e., utterance-medially) suggests that they have an adult-like representation of the voicing feature, but that their implementation is not yet adult-like, especially utterance-finally. One possible account of the failure of post-release cues to vary systematically with voicing in utterance-final position in child speech is that the child's control of subglottal pressure and air flow at the ends of utterances is immature, resulting in a long exhale that produces aspiration noise at the vocal folds irrespective of the voicing feature of the coda consonant. Future work that uses spectral analysis to distinguish friction from aspiration sources (e.g., Hanson and Stevens, 2006) in the post-coda-release region may help to test this

hypothesis. Interestingly, although children produced exaggerated cues for both alveolar and velar codas at earlier ages in Study 2, the distribution of these cues differed for alveolar vs. velar codas in both utterance-final and utterance-medial positions. This contrasts with distinctions in voicing in Study 1, where children showed frequent and long post-release noise for both voiceless and voiced codas in utterance-final position.

Third, the younger children in our studies had, overall, more frequent stop releases, a greater mean number of release bursts, and more frequent and longer post-release noise than mothers. These values decreased by 2;6, approximating more adult-like patterns. This phenomenon of decreasing exaggeration of the adult cue pattern was especially prominent in utterance-medial position. There are several possible explanations for this process. One possibility is that it arises from initially immature motor development and control. Previous studies have suggested that speech in this age range is very different from adult speech in terms of respiration, phonation, and articulation. For example, children below age 6 are still learning to coordinate multiple articulators such as the lips and jaw for speech production (Green *et al.*, 2000). Drawing on acoustic data from various aspects of child speech production, Kent (1976) concluded that motor control accuracy improves with age until fully adult-like performance is achieved at about 11 or 12 years.

Moreover, studies on the development of laryngeal and respiratory function during speech have revealed that children generate higher subglottal and intraoral pressure than adults (Stathopoulos and Sapienza, 1993). For example, children aged 4–6 and 10–12 years generate greater intraoral air pressure than adults during the production of bilabial stops (Bernthal and Beukelman, 1978). Netsell *et al.* (1994) also showed that children aged 3;3–4;3 had higher subglottal pressure, higher resistance, and less airflow through the glottis than adults during production of the syllables /pi/ and /pa/. They attributed these patterns to the smaller size of the laryngeal airway and increased expiratory muscle forces in children. Greater subglottal and intraoral pressures may lead to higher airflow during the stop release. Acoustically, this could result in higher-amplitude post-release noise, greater number of bursts, etc. (Imbrie, 2005). Perhaps this could partly explain why some of the acoustic cues were exaggerated in young children's speech in our study, especially those related to the aerodynamics of air pressure and flow. In addition, children's small articulator size and low articulator mass might allow an articulator to vibrate quickly against the palate, even if the pressure drop immediately after constriction release was relatively small. This could also increase the incidence of multiple bursts in children's speech (Imbrie, 2005).

Another possible reason for the decrease in these values over time could be a change in children's speaking rate during the period of examination. Young children are known to have overall greater segment durations and more durational variability than adults (Kent and Forner, 1980; Smith, 1978). Furthermore, children's utterances are usually short (i.e., have fewer words) and they typically speak more slowly than adults (Kowal *et al.*, 1975). Thus, it is

likely that the duration of children's words decreases as they become more fluent, faster speakers. If so, the decrease in the presence and duration of some of the cues, such as the duration of post-release noise, might be the result of an increase in overall speaking rate. We investigated this possibility by examining the change in vowel duration from age 1;6 to 2;6. If vowel duration changes systematically over time, this might be indicative of a speaking rate change during this period. To test this possibility, we measured the duration of lax (*back, gat, bed, big, book, cat, duck, good, hat, head, pig*) and tense (*boat, dog, feet, food, hot, side*) vowels separately before voiced and voiceless codas in utterance-medial and utterance-final positions in individual children's speech. The results, however, showed no significant change in vowel duration at least during the time window examined in the present study. That is, vowel duration remained constant while only the presence and duration of coda-release-related cues decreased over time. This suggests that the decrease in the presence of release bursts and the presence and duration of post-release noise duration in children's speech is not simply due to a change in speaking rate.

Recall, however, that at 1;6, children's use of many of the cues was especially exaggerated in utterance-medial position compared to those of their mothers, with a sharp decrease toward more adult-like values by 2;6. If there is no word following immediately, there might be more time for children to produce an utterance-medial coda, providing the context to produce longer post-release noise, etc. Thus, perhaps the intervals between children's utterance-medial words are longer earlier in development, but decrease as children become more fluent speakers. This is obviously an area for further research.

It is well documented that adult speakers produce various acoustic cues that may assist the listener in the perception of linguistic structure. For example, lengthening of syllables and segments at prosodic boundaries may serve as a powerful cue to the listener for the location of boundaries of constituents such as words, phrases, or sentences (Oller, 1973; Klatt, 1976; Turk and Shattuck-Hufnagel, 2007). Likewise, exaggerated use of some cues at earlier ages might reflect children's efforts to enhance feature contrasts, or on the other hand it might reflect an inability to approximate adult cue values (Demuth *et al.*, 2006; Shattuck-Hufnagel *et al.*, 2011). Investigating the relationship between what children know about the phonological dimensions in the ambient language and what they can produce, given their stage of articulatory development and control, can begin to provide us with deeper insights into the phonological representation of words in young children. Our examination of acoustic cues to voicing and PoA contrasts offers a step in this direction.

V. CONCLUSION

This study provides the first comprehensive, longitudinal examination of non-spectral acoustic cues to coda voicing and place contrasts in children from the age of 1;6, an age where there is little data available on children's

spontaneous speech productions. It found that children showed systematic variation in many non-spectral acoustic cues as a function of voicing and PoA from the beginning of our investigation at 1;6. These early-acquired cues included longer duration of preceding vowels and more frequent voice bars for voiced as compared with voiceless codas, and more frequent release bursts, greater average number of release bursts, and more frequent and longer duration of post-release noise for velars as compared with alveolars. The presence and duration of post-release cues varied less systematically with voicing before 2;6, but this was mainly due to the small variation in utterance-final position. However, there were two aspects of the children's production of these cues that generally differed from adults. First, children's production of vowel glottalization was not reliably adult-like with respect to coda voicing and PoA by 2;6. Second, children had a greater number of bursts, and more frequent and longer post-release noise at 1;6, but both decreased by 2;6, more closely approximating adult values both quantitatively (how much) and distributionally (where).

Previous studies based on phonological transcriptions of child speech have noted that children often make "errors" in the realization of both coda voicing and PoA features. However, the acoustic details underlying these observations have not always been available. Examination of the acoustic cues individually provides a means for resolving some of the ambiguity about the developmental course of acquisition of the voicing and PoA contrasts that arises when listeners are unable to determine the voicing and PoA features of a coda consonant. Our findings provide rare evidence from early speech production that children below age 2 may have sophisticated, adult-like representations of coda voicing and PoA feature contrasts, as well as some knowledge of the distribution of these cues in adult speech. This is important for providing a baseline against which to interpret possible representational and/or implementation problems exhibited by children with language delay. It also offers a step toward developing a more general and complete model of how children acquire mastery of the acoustic cues to the feature contrasts of coda consonants in simple, monomorphemic words. In so doing, it provides a framework for beginning to explore the nature of morphemic coda consonants, and how and when these are acquired.

ACKNOWLEDGMENTS

We thank Stefan Th. Gries, Mark Johnson, and Ivan Yuen for assistance and discussion, Helen Hanson for advice on acoustic measures, and members of the Child Language Lab at Brown University (Melanie Cabral, Karen Evans, Dustin Foley, Heidi Jiang, Elana Kreiger-Benson, Jeremy Kuhn, Melissa Lopez, Matt Masapollo, Deniz Ozzkan, Meghan Patroliia, Miranda Sinnott-Armstrong, Matt Vitorla) for coding assistance. This work was supported by NIH grant R01HD057606 (Demuth and Shattuck-Hufnagel).

¹See the Child Language Data Exchange System (CHILDES; <http://childes.psy.cmu.edu/> (Last viewed 8/1/11)).

- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics* (Cambridge University Press, Cambridge, UK), pp. 241–302.
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). "Mixed-effects modeling with crossed random effects for subjects and items," *J. Mem. Lang.* **59**(4), 390–412.
- Bernthal, J. E., and Beukelman, D. R. (1978). "Intraoral air pressure during the production of /p/ and /b/ by children, youths, and adults," *J. Speech Hear. Res.* **21**, 361–371.
- Blumstein, S. E., and Stevens, K. N. (1979). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," *J. Acoust. Soc. Am.* **66**, 1001–1017.
- Boersma, P., and Weenink, D. (2005). "PRAAT: Doing phonetics by computer," (version 4.4.07) [computer program], <http://www.praat.org/> (Last viewed January 8, 2011).
- Bond, Z. S., and Wilson, H. F. (1980). "Acquisition of the voicing contrast by language-delayed and normal-speaking children," *J. Speech Hear. Res.* **23**, 152–161.
- Buder, E. H., and Stoel-Gammon, C. (2002). "American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing," *J. Acoust. Soc. Am.* **111**, 1854–1864.
- Cole, J., Kim, H., Choi, H., and Hasegawa-Johnson, M. (2007). "Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech," *J. Phonetics* **35**, 180–209.
- Crystal, T. H., and House, A. S. (1988). "Segmental durations in connected-speech signals: Current results," *J. Acoust. Soc. Am.* **83**, 1553–1573.
- Demuth, K., Culbertson, J., and Alter, J. (2006). "Word-minimality, epenthesis and coda licensing in the early acquisition of English," *Lang. Speech* **49**, 137–173.
- Eimas, P. D. (1974). "Auditory and linguistic processing of cues for place of articulation by infants," *Percept. Psychophys.* **16**, 513–521.
- Eimas, P. D., and Miller, J. L. (1980). "Discrimination of information for manner of articulation," *Infant Behav. Dev.* **3**, 367–375.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). "Speech perception in infants," *Science* **171**, 303–306.
- Fant, G., and Lindblom, B. (1961). "Studies of minimal speech and sound units," Quarterly Progress Report No. 2, (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden), pp. 1–11. Retrieved from http://www.speech.kth.se/prod/publications/files/qpsr/1961/1961_2_2_001-011.pdf (Last viewed January 8, 2011).
- Green, J. R., Moore, C. A., Higashikawa, M., and Steeve, R. W. (2000). "The physiologic development of speech motor control: Lip and jaw coordination," *J. Speech Lang. Hear. Res.* **43**, 239–255.
- Hanson, H. M., and Stevens, K. N. (2006). "The nature of aspiration in stop consonants in English," *J. Acoust. Soc. Am.* **119**, 3393. Retrieved from <http://dspace.mit.edu/handle/1721.1/64959> (Last viewed January 8, 2011).
- Hirano, M., Kurita, S., and Nakashima, T. (1983). "Growth, development, and aging of human vocal folds," in *Vocal Fold Physiology*, edited by D. M. Bless, and J. H. Abbs (College Hill Press, San Diego, CA), pp. 22–43.
- House, A. S. (1961). "On vowel duration in English," *J. Acoust. Soc. Am.* **33**, 1174–1182.
- Imbrie, A. K. K. (2005). "Acoustical study of the development of stop consonants in children," Ph.D. thesis, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA.
- Inkelas, S., and Rose, Y. (2007). "Positional neutralization: A case study from child language," *Lang.* **83**, 707–736.
- Irwin, O. C. (1947). "Infant speech: Consonantal sounds according to place of articulation," *J. Speech Disord.* **12**, 397–401.
- Keating, P. A., Westbury, J. R., and Stevens, K. N. (1980). "Mechanisms of stop-consonant release for different places of articulation," *J. Acoust. Soc. Am.* **67**, S93. Retrieved from http://www.linguistics.ucla.edu/people/keating/KeatingWestburyStevens_ASA19_80.pdf (Last viewed January 8, 2011).
- Kent, R. D. (1976). "Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies," *J. Speech Hear. Res.* **19**, 421–447.
- Kent, R. D., and Forner, L. L. (1980). "Speech segment durations in sentence recitations by children and adults," *J. Phonetics* **8**, 157–168.
- Kewley-port, D., and Preston, M. (1974). "Early apical stop production: A voice onset time analysis," *J. Phonetics* **2**, 195–219.
- Keyser, S. J., and Stevens, K. N. (2006). "Enhancement and overlap in the speech chain," *Lang.* **82**, 33–63.

- Klatt, D. J. (1976). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," *J. Acoust. Soc. Am.* **59**, 1208–1221.
- Ko, E.-S. (2007). "Acquisition of vowel duration in children speaking American English," in *Proceedings of Interspeech 2007* (Antwerp, Belgium), pp. 1881–1884.
- Koenig, L. L. (2000). "Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds," *J. Speech Lang. Hear. Res.* **43**, 1211–1228.
- Kowal, S., O'Connell, D., and Sabin, E. (1975). "Development of temporal patterning and vocal hesitations in spontaneous narratives," *J. Psycholinguistic Res.* **4**, 195–207.
- Krause, S. E. (1982). "Developmental use of vowel duration as a cue to postvocalic stop consonant voicing," *J. Speech Hear. Res.* **25**, 388–393.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). "Linguistic experience alters phonetic perception in infants by 6 months of age," *Science* **255**, 606–608.
- Lisker, L. (1970). "Supraglottal air pressure in the production of English stops," *Lang. Speech* **13**, 215–230.
- Lisker, L., and Abramson, A. (1964). "A cross-language study of voicing in initial stops: Acoustical measurements," *Word* **20**, 384–422.
- Macken, M. A., and Barton, D. (1980). "The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants," *J. Child Lang.* **7**, 41–74.
- MacWhinney, B. (2000). *The CHILDES Project* (Erlbaum, Mahwah, NJ), pp. 1–808.
- McAllister, T. K. (2009). *The Articulatory Basis of Positional Asymmetries in Phonological Acquisition*, Ph.D. thesis, (Massachusetts Institute of Technology, Cambridge, MA).
- Netsell, R., Lotz, W. K., Peters, J. E., and Schulte, L. (1994). "Developmental patterns of laryngeal and respiratory function for speech production," *J. Voice* **8**(2), 123–131.
- Olive, J. P., Greenwood, A., and Coleman, J. S. (1993). *Acoustics of American English Speech: A Dynamic Approach* (Springer, New York, NY), pp. 1–396.
- Oller, D. K. (1973). "The effect of position in utterance on speech segment duration in English," *J. Acoust. Soc. Am.* **54**, 1235–1247.
- Pater, J., and Werle, A. (2003). "Direction of assimilation in child consonant harmony," *Can. J. Ling.* **48**(3/4), 385–408.
- Pierrehumbert, J., and Frisch, S. (1994). "Source allophony and speech synthesis," in *Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis*, pp. 1–4.
- Quené, H., and van den Bergh, H. (2004). "On multi-level modeling of data from repeated measures designs: A tutorial," *Speech Comm.* **43**(1/2), 103–121.
- R Development Core Team (2011). *A language and environment for statistical computing*, R foundation for statistical computing, Vienna, Austria. Retrieved from <http://www.R-project.org> (date last viewed 9/27/10).
- Redi, L., and Shattuck-Hufnagel, S. (2001). "Variation in the realization of glottalization in normal speakers," *J. Phonetics* **29**, 407–429.
- Repp, B. H. (1979). "Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants," *Lang. Speech* **22**, 173–189.
- Scobbie, J., Gibbon, F., Hardcastle, W., and Fletcher, P. (2000). "Covert contrast as a stage in the acquisition of phonetics and phonology," in *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, edited by M. Broe, and J. Pierrehumbert (Cambridge University Press, Cambridge, UK), pp. 194–207.
- Shattuck-Hufnagel, S., Demuth, K., Hanson, H., and Stevens, K. N. (2011). "Acoustic cues to stop-coda voicing contrasts in the speech of American English 2–3-year-olds," in *Where Do Features Come From? The Nature and Sources of Phonological Primitives*, Linguistic Series, edited by G. N. Clements, and R. Ridouane (Elsevier, North-Holland), pp. 327–341.
- Smith, N. (1973). *The Acquisition of Phonology: A Case Study* (Cambridge University Press, London, UK), pp. 1–284.
- Smith, B. L. (1978). "Temporal aspects of English speech production: A developmental perspective," *J. Phonetics* **6**, 37–67.
- Smith, B. L. (1979). "A phonetic analysis of consonantal devoicing in children's speech," *J. Child Lang.* **6**, 19–28.
- Song, J. Y., and Demuth, K. (2008). "Compensatory vowel lengthening for omitted coda consonants: A phonetic investigation of children's early representations of prosodic words," *Lang. and Speech* **51**(4), 385–402.
- Stager, C. L., and Werker, J. F. (1997). "Infants listen for more phonetic detail in speech perception than in word-learning tasks," *Nature* **388**, 381–382.
- Stathopoulos, E. T., and Sapienza, C. (1993). "Respiratory and laryngeal measures of children during vocal intensity variation," *J. Acoust. Soc. Am.* **94**(5), 2531–2541.
- Stevens, K. N. (2000). *Acoustic Phonetics* (MIT Press, Cambridge, MA), pp. 1–624.
- Stevens, K. N. (2002). "Toward a model for lexical access based on acoustic landmarks and distinctive features," *J. Acoust. Soc. Am.* **111**, 1872–1891.
- Turk, A., and Shattuck-Hufnagel, S. (2007). "Phrase-final lengthening in American English," *J. Phonetics* **35**(4), 445–472.
- Velten, H. V. (1943). "The growth of phonemic and lexical patterns in infant language," *Lang.* **19**, 281–292.
- Weismer, G., Dinnsen, D.A., and Elbert, M.A. (1981). "A study of the voicing distinction associated with omitted, word-final stops," *J. Speech Hear. Disord.* **46**, 320–327.
- Werker, J. F., and Tees, R. C. (1984). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," *Infant Behav. Dev.* **7**, 49–63.
- White, K. S., and Morgan, J. L. (2008). "Sub-segmental detail in early lexical representations," *J. Mem. Lang.* **59**, 114–132.
- Wright, R. (2004). "A review of perceptual cues and cue robustness," in *Phonetically-Based Phonology*, edited by B. Hayes, R. Kirchner, and D. Steriade (Cambridge University Press, New York, NY), pp. 34–57.
- Zlatin, M. A., and Koenigsnecht, R. A. (1976). "Development of the voicing contrast: A comparison of voice onset time in stop perception and production," *J. Speech Hear Res.* **19**, 93–111.