

## The acquisition of two phonetic cues to word boundaries\*

MELISSA A. REDFORD

*The University of Oregon, Eugene, OR, USA*

AND

CHRISTINA E. GILDERSLEEVE-NEUMANN

*Portland State University, Portland, OR, USA*

*(Received 18 August 2005. Revised 19 July 2006)*

### ABSTRACT

The study evaluated whether durational and allophonic cues to word boundaries are intrinsic to syllable production, and so acquired with syllable structure, or whether they are suprasyllabic, and so acquired in phrasal contexts. Twenty preschool children (aged 3;6 and 4;6) produced: (1) single words with simple and complex onsets (e.g. *nail* vs. *snail*); and (2) two-word phrases with intervocalic consonant sequences and varying boundary locations (e.g. *this nail* vs. *bitty snail*). Comparisons between child and adult control productions showed that the durational juncture cue was emergent in the four-year-olds' productions of two-word phrases, but absent elsewhere. In contrast, the allophonic cue was evident even in the three-year-olds' productions of single words. Perceptual judgments showed that age- and type-dependent acoustic differences translated into differences in listener behavior. The differential acquisition of the two juncture cues is discussed with reference to the acquisition of articulatory timing control.

### INTRODUCTION

A major research question in the area of child language acquisition is how children come to extract words from running speech when no obvious

---

[\*] We thank Tatiana Furrow as well as Gennifer Fink and Jennifer Peddicord for help with subject recruitment and data collection. We are grateful to Lou Moses for letting us use his lab to run subjects and to the UO Psychology Department for access to the baby and child database for subject recruitment. This work was supported in part by a Summer Research Award from The University of Oregon. Address for correspondence: Melissa A. Redford, Department of Linguistics, 1290 University of Oregon, Eugene, OR 97403. Email: redford@uoregon.edu

boundary markers exist. This so-called segmentation problem is understood as a perceptual problem; the solution is typically thought of in terms of cues afforded by global linguistic phenomena, such as rhythm patterns and phonotactics (Cutler, Mehler, Norris & Segui, 1986; Morgan & Saffran, 1995; Saffran, Aslin & Newport, 1996). Segmentation is not typically perceived as a problem for the development of language production. Whereas children must locate boundaries precisely for comprehension, they are thought to signal them automatically in production as a by-product of learning a language. This view is supported by production studies showing that certain of the recognized cues to word boundaries are acquired globally, even before a child acquires words. For example, the basic rhythm pattern of a language is manifest in babbling as are some basic phonotactic regularities (Boysson-Bardies, Bacri, Sagart & Poizat, 1981; Boysson-Bardies & Vihman, 1991; Davis, MacNeilage, Matyear & Powell, 2000). However, several local or syntagmatic phonetic cues to word segmentation also exist. For example, English listeners use the consonantal duration pattern in an obstruent–sonorant sequence to locate word boundaries in minimal pair sentences such as *help a snail* and *help us nail* (Christie, 1977) and they use stop release duration to locate word boundaries in near minimal pair phrases such as *I stop* and *nice top* (Davidsen-Nielsen, 1974). It may be that these juncture cues cannot be produced without attention to word boundaries in production, which would suggest an interesting problem for speech acquisition. Children would need to acquire two levels of articulatory timing: word-level timing control to realize phonemic contrasts (e.g. voice onset time, as in *pig* versus *big*) and sound sequencing (e.g. *dog* versus *god*); and phrase-level timing control to realize the word boundary patterns. Alternatively, it may be that these patterns are tied to syllable structure, and so only one level of articulatory timing control need be acquired. If this is the case, then children might acquire the specific patterns of timing that distinguish, for example, *a snail* from *us nail*, as soon as they are able to produce both the /s/+sonorant onset cluster in *snail* and the offset–onset sequence in *us nail*. The current study investigated the acquisition of two different juncture cues in order to better understand both the acquisition of articulatory timing control as well as the nature of what is being acquired.

#### *Durational cues to juncture*

Production studies show that consonantal and vocalic duration vary with syllable structure and position. For instance, singleton initial consonants are longer than singleton final consonants in monosyllabic words that occur in phrase-medial position (Boucher, 1988; Cho & Keating, 2001; Keating, Wright & Zhang, 1999; Redford & Diehl, 1999; Turk &

Shattuck-Hufnagel, 2000). Other studies show that internal members of word-onset and -offset consonant clusters are reduced and hence shorter than the external members (Browman & Goldstein, 1988; Haggard, 1973; Klatt, 1976). Perception studies indicate that these patterns can be used by listeners to infer boundaries (Christie, 1977; DeMarco & Harrell, 1995; Quené, 1992; Redford & Randall, 2005; Tuller & Kelso, 1991). For instance, listeners exposed to intervocalic obstruent–sonorant sequences hear an onset cluster to a subsequent vowel when the second consonant in the sequence is especially short relative to the first, and they hear an offset–onset sequence when the second consonant is longer (Christie, 1977; Redford & Randall, 2005).

Children appear to use the durational cue to juncture as effectively as adults. DeMarco & Harrell (1995) showed that adults and eight- and nine-year-old children are able to discriminate minimal word pairs such as *its wings* versus *it swings* with 95% accuracy in a neutral carrier phrase. Although we know of no similar study with younger children, a study conducted by Christophe and colleagues (Christophe, Dupoux, Bertoncini & Mehler, 1994) shows that even very young infants are sensitive to juncture cues. Christophe *et al.* presented three-day-old French infants with bisyllabic stimuli extracted from within words (e.g. *mati* in *mathématicien*) and across word boundaries (e.g. *mati* in *panorama typique*). Using a high-amplitude sucking paradigm, they found that infants were able to discriminate between the two types of stimuli, suggesting that they perceived the phonetic patterns that distinguished the stimuli. Given this result, it seems reasonable to assume that three- and four-year-olds would also have access to phonetic juncture cues, and would have learned to apply these cues to segment speech during language comprehension. The primary question addressed in this study is whether children of this age are able to produce such cues.

### *Juncture in speech acquisition*

Durational cues to juncture have traditionally been explained in terms of the syllable (e.g. Browman & Goldstein, 1988; Campbell & Isard, 1991; Klatt, 1976; Krakow, 1999; Lehiste, 1970). That is, syllables are thought to either provide the temporal frame within which segmental duration is adjusted or the domain within which articulatory timing is specified. This view suggests that speakers inadvertently produce the patterns of segmental duration that cue word boundary location because syllable boundaries align with word boundaries. If this is true, then children might be expected to produce the durational cue to word boundaries as soon as they are able to produce the relevant syllable structures.

The syllable-based explanation for segmental duration patterns is problematic because production experiments that document such patterns typically confound syllable and word boundaries. Such a confound means that the durational patterns cueing boundary perception may be tied to word boundaries rather than to syllable boundaries. Redford (2007) explored this possibility in a study on intervocalic stop-liquid sequences, and showed that the durational patterns marking word boundaries in English do not mark word-internal syllable boundaries. She concluded from this finding that some English boundary patterns may be better explained in terms of listeners' needs than in terms of basic motor speech processes. Kohler (1991) made a similar suggestion for German when discussing the phenomenon of word-initial lengthening. He noted that word-initial consonants might be longer than word-final consonants (in phrase-medial position) because speakers emphasize that portion of the word which contains more information, thereby enabling faster lexical access in the listener.

A listener-oriented explanation for durational cues to juncture suggests that the articulatory timing routines giving rise to such cues may be independent of syllable structure. If this is the case, then children's ability to produce the durational patterns might be acquired separately from their ability to produce different types of syllables. In particular, children may first learn the articulatory routines that govern phonemic patterns and segment sequencing by practicing words in isolation. The child would only begin to acquire the boundary-dependent durational patterns when he or she begins to string words together into multiword utterances. Even then, development of such control may be prolonged because it entails a more complex production routine: one in which the timing parameters are specified separately for within and between word articulation. Timing control over juncture phenomena may also be delayed because young children, who are acquiring the skills for fluent output, may not be sensitive to a listener's need to segment this output into its component parts.

### *The current study*

The current study was designed to investigate when and how children acquire the ability to produce the durational cue to word boundaries in English. We used /s/C sequences to investigate the acquisition of this cue for two reasons. First, phonetic juncture cues may be especially relevant for segmentation of /s/C sequences at word boundaries: English possessive, plural, and third person morphology entails that /s/ occurs very frequently in word-final position and before some other word-initial consonant. Also, /s/ is the only obstruent in English that can combine with both sonorants

and obstruents to form onset clusters, which means that /s/ also occurs very frequently in word-initial position before some other consonant. Second, /s/C sequences allow us to compare the acquisition of the durational juncture cue with the acquisition of a different kind of phonetic juncture cue. When /s/ combines with voiceless stop consonants in English, the stops are realized with significantly less aspiration than when they are singleton onsets (e.g. *spy* [spai] versus *pie* [p<sup>h</sup>ai]). Perception studies indicate that stop aspiration provides a robust cue to boundary location (Davidsen-Nielsen, 1974; Redford & Randall, 2005). Like the durational pattern that cues word boundaries in /s/+sonorant sequences, variation in stop aspiration is usually explained with reference to syllable structure. But unlike the durational pattern, which is a gradient pattern produced by varying closure duration according to the segmental duration and boundary context (e.g. shorter C<sub>1</sub> and longer C<sub>2</sub> for C<sub>1</sub>#C<sub>2</sub>, longer C<sub>1</sub> and shorter C<sub>2</sub> for #C<sub>1</sub>C<sub>2</sub>), stop aspiration variation is categorical. The aspirated and unaspirated allophones are produced with distinct voice onset times when coordinated with vowels (long lag VOT for #CV, short lag VOT for #/s/CV), and the voice onset times for the different allophones do not overlap. Potential differences in the development of control over the durational and allophonic juncture cues could indicate that the underlying articulatory timing routines are also differently specified.

A cross-sectional design was used to investigate the acquisition of the phonetic juncture cues associated with intervocalic /s/C sequences. The experiments focused on the productions of three- and four-year-old children because the production of boundary patterns in utterances with /s/C sequences such as *a snail* versus *us nail* presupposes an ability to produce different initial consonants, word-final /s/ and word-initial /s/C clusters. The ability to produce singleton consonantal onsets emerges early in language acquisition, but the acquisition of final consonants and onset clusters emerges later (Stoel-Gammon & Dunn, 1985: 15–46). In general, final consonants are acquired by most children by age three (Stoel-Gammon & Dunn, 1985: 43) and initial obstruent + approximate clusters are acquired by most children between 2;8 and 3;10 (Grunwell, 1981). However, /s/C clusters are acquired by most children slightly later, between 3;3 and 3;8 (Grunwell, 1981). It is this late acquisition of /s/C onset clusters that led us to investigate the speech of children aged 3;6 and 4;6. Specifically, we expected that most three-year-olds would be able to produce the relevant syllable onsets, but may not necessarily produce the durational and allophonic juncture cues since they would have just acquired mastery over the /s/C clusters. If three-year-olds could not produce the juncture cues, we thought that four-year-olds would be able to since they would have had considerable practice with the clusters by this age. Experiments 1 and 2 compared child and adult productions of word boundary patterns

in single-word utterances and in two-word phrases to determine when preschool children produce the two phonetic cues to juncture in an adult-like fashion. Experiment 3 was conducted to evaluate the perceptual robustness of the age- and boundary-dependent acoustic differences described in Experiment 2.

## EXPERIMENT 1

The first experiment examined whether three- and four-year-olds produce /s/, sonorant and stop consonants differently as a function of onset type as adults do. Child productions of consonants were evaluated as a function of onset type – singleton onset versus /s/+sonorant or /s/+stop onset cluster – and compared with adult productions of the same consonants in the different onset types. The goal was to evaluate whether the durational and allophonic juncture cues, which are attributed to syllable structure in adult speech, are in fact acquired with syllable structure.

### METHOD

#### *Participants*

Ten three-year-olds and ten four-year-olds and their parents participated in the experiment. The three-year-olds ranged in age from 3;4 to 3;7. The four-year-olds ranged from 4;4 to 4;7. The children's parents were contacted by telephone from a call list maintained by the Department of Psychology at the University of Oregon. The telephone contact served not only as a recruitment tool, but also as an initial screening tool. Only children with normal hearing from monolingual, English-speaking households were invited to participate in the experiment. All parents were also interviewed upon arriving for the experiment to determine whether their child had exhibited normal development in language and motor skill acquisition. All the data reported in this study come from children who exhibited normal development as determined by a number of well-known speech and motor milestones (e.g. age of first canonical babble, age of first steps). The parents were also all native English speakers with self-reported normal hearing.

#### *Stimuli*

The stimuli were chosen in order to compare /s/, sonorant and stop aspiration duration in singleton onsets to /s/, sonorant and aspiration duration in /s/+sonorant and /s/+stop onset clusters. Table 1 shows the 15 words used in the experiment to elicit the different consonant types in the different syllable onsets.

## ACQUISITION OF JUNCTURE

TABLE 1. *Single-word stimuli used in Experiment 1*

Singleton onsets			Onset clusters	
/s/	sonorant	stop	/s/+sonorant	/s/+stop
<i>sun</i>	<i>leaf</i>	<i>pig</i>	<i>slide</i>	<i>spider</i>
<i>sock</i>	<i>mouse</i>	<i>puppy</i>	<i>snake</i>	<i>spoon</i>
<i>soap</i>	<i>monkey</i>	<i>kite</i>	<i>snowman</i>	<i>star</i>

Adults read the words off a randomized list that included 105 other word and word-pair stimuli. Some of the additional stimuli were used in Experiment 2, others were included as part of a separate study. The randomized list was broken into four columns, with the target words randomly interspersed across the columns. The adults read the word list one column at a time, completing each column at different points during the experiment. For instance, the first column was often read at the start of the experiment and used to show the child how to speak into the microphone. The fourth column was typically read at the end of the experiment. The second and third columns were read at separate points either during a break in the picture naming task (described below) or in the imitative task (described in Experiment 2) or during a break between the two tasks. This method of recording minimized some of the list effects that are known to occur with this type of elicitation.

Child productions of the 15 words were obtained in a picture naming task; each word was pictured in color on 5 × 7-inch laminated cards. The pictures were obtained from Boardmaker (Mayer-Johnson, Inc.), ensuring that they had been previously tested for ease of recognition. Although the use of picturable words encouraged spontaneous language production, it constrained the set of words from which the stimuli were selected. This constraint resulted in the following asymmetries: the bilabial nasal sonorant was elicited in singleton position, but the alveolar nasal was elicited in /s/+sonorant clusters; the voiceless velar stop was elicited in singleton position, but not as part of a cluster; some of the words were monosyllabic and others were disyllabic. These asymmetries were orthogonal to the comparison between child and adult production, and were expected to be neutral with respect to the comparison of onset singletons versus onset clusters for the following reasons. First, there are no reported differences in the intrinsic durations of singleton bilabial and alveolar nasals (see e.g. Klatt, 1976; Umeda, 1977). Second, the documented differences in voice onset time for singleton voiceless alveolar and velar stop onsets is on the order of 10 milliseconds in English, which is several times smaller than the average 60 millisecond difference between aspirated stops and unaspirated

stops in English (Lisker & Abramson, 1964). Finally, word length is known only to affect the duration of syllable onsets in word-medial position (Klatt, 1976; Oller, 1973). The onsets of interest in this experiment were all word-initial. The word-initial consonants in this study also always occurred as onsets to a stressed syllable regardless of whether the words themselves were mono- or disyllabic.

### *Procedure*

The experiment took place in a child-friendly experimental room, with the experimenter, child and parent all sitting around a child-size table. Parents remained with their children for the duration of the experiment. The child and parent productions were recorded using a portable DAT recorder and a high-quality free-standing microphone oriented towards the child or parent on the table. The responses were transferred to a computer for later acoustic analysis.

A picture naming task was used to elicit the target words. The picture cards were randomly interspersed with 48 other picture cards that were included for a separate study. Children were asked to clearly name the picture presented to them. Spontaneous word productions were the norm. In the few cases where a child did not produce the desired lexical item after several prompts, delayed imitation was used.

### *Measurements*

Consonantal durations were measured in Praat (Boersma & Weenink, 2002) using concurrent displays of the oscillogram and spectrogram. Measurements were taken on all child and adult productions, but the analyses excluded those productions in which a singleton onset was substituted for an /s/C onset. Three children (two three-year-olds and one four-year-old) consistently substituted singleton onsets for onset clusters, and four others (three three-year-olds and one four-year-old) occasionally did. Overall, 12 tokens with /s/+sonorant onsets and 11 tokens with /s/+stop onsets were excluded from the three-year-old analyses, and 5 tokens with /s/+sonorant onsets and 3 tokens with /s/+stop onsets were excluded from the four-year-old analyses.

Measurement criteria for /s/, sonorant consonants and stop release were as follows. The fricative /s/ was defined by the sudden drop/rise in the periodic waveform and by the presence of noisy high-frequency energy. All continuous frication was included in the duration of /s/. This meant that /s/ duration sometimes included evidence of an articulatory transition to the following consonant, for example, a lowering in the average frequency associated with velum or tongue body lowering for a subsequent nasal or



liquid sonorant. Sonorant boundaries were defined on their left edge by the onset of voicing and periodicity. The right edge was defined by a sudden increase in mid-frequency energy and the appearance of F<sub>2</sub>. Only stop burst+aspiration (henceforth aspiration) duration was compared across onset type, since stop aspiration duration is the relevant cue for word boundary identification. Aspiration duration included all voiceless energy from the burst to the onset of the vowel. All measurements were supplemented by auditory judgments.

It should be noted that the analyses assessed significant differences in absolute durations as a function of speaker (child versus adult) and onset type (singleton versus cluster). This means that it was more important for the measurement criteria to be consistent throughout rather than for the values of individual segments to be in perfect agreement with values obtained using different criteria. To evaluate measurement consistency and accuracy according to the criteria, ten percent of the data was randomly selected and measured by a second rater. The mean differences (and standard deviations) between rater measurements for the three-year-old data were 7.3(±15) milliseconds for the child data and 4.3(±3.9) milliseconds for the adult data. The mean differences for the four-year-old data were 3.4(±2.3) milliseconds for the child data and 4.0(±4.1) milliseconds for the adult data. Reliability was calculated as a correlation between the two raters; an appropriate statistic for determining inter-rater reliability on a continuous variable. Inter-rater correlations were extremely high ( $r=0.98$  and  $r=0.99$  for the three-year-old child and adult data respectively, and  $r=0.99$  and  $r=0.99$  for the four-year-old child and adult data), indicating good measurement consistency and accuracy according to the criteria.

## RESULTS

The data were split to compare children and adult productions within each age group. The purpose of the child-to-adult comparison was to test for adult-like control over consonantal duration as a function of onset type, the manipulated variable. Similarities between child and adult productions would suggest that children have acquired the underlying articulatory timing routines for the different onset structures. Conversely, significant differences would indicate that they had not. The analyses of /s/ duration suggested that four-year-olds had acquired more adult-like timing control than three-year-olds; however, neither three- nor four-year-olds' productions of sonorant consonants were significantly affected by onset type, even though adult productions were. In contrast, children of both age groups showed adult-like mastery over voice onset timing for voiceless stops as a function of onset type. Detailed results are presented below; first

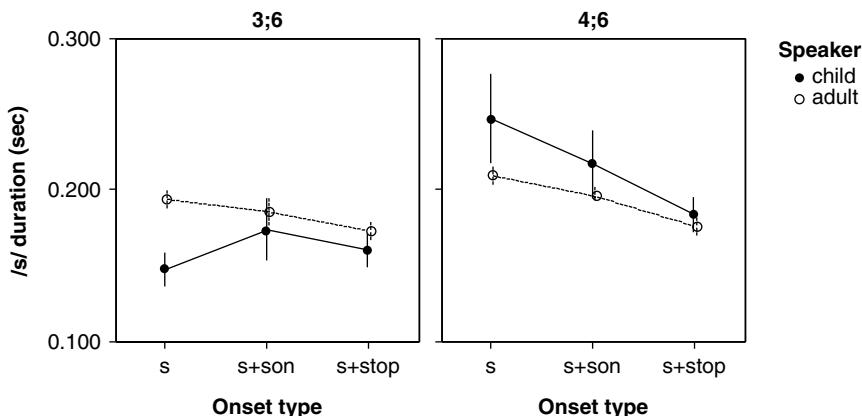


Fig. 1. /s/ duration for child and adult productions of single words with simple /s/ and complex /s/+sonorant and /s/+stop onsets for the different age groups.

for /s/ duration, then for sonorant duration and finally for stop aspiration duration.

*/s/ duration as a function of onset type.* Adult and child productions of singleton /s/ duration were compared with productions of /s/ duration in /s/+sonorant and /s/+stop onset clusters (e.g. *sun* vs. *slide* vs. *spider*). The (2) speaker  $\times$  (3) onset type ANOVA showed a significant effect of speaker in the three-year-old age group ( $F(1, 144) = 5.86, p = 0.017, \eta_p^2 = 0.04$ ), but not in the four-year-old age group ( $F(1, 164) = 2.66, p > 0.1$ ). The effect of onset type was only significant in the four-year-old age group ( $F(1, 164) = 4.37, p = 0.014, \eta_p^2 = 0.05$ ). The interaction between speaker and onset type was not significant for either age group. Figure 1 displays these results.

Figure 1 shows that /s/ duration was longest for singleton onsets and shortest for /s/+stop onsets in the four-year-old group where the effect of onset type was significant. The figure also shows that although the four-year-olds produced the same qualitative pattern as the adults, their /s/ durations were longer and more variable than adult /s/ durations. A post hoc comparison of adult /s/ productions confirms what is evident from the figure; namely, that adults produced the same pattern of long /s/ duration in singleton onsets and short duration in /s/+stop onset regardless of their child's age (i.e. the difference between parents of children aged 3;6 and 4;6 was non-significant).

*Sonorant duration as a function of onset type.* In contrast to the results on /s/ duration, results from the (2) speaker  $\times$  (2) onset type ANOVA revealed significant differences between child and adult productions of sonorants in singleton and /s/+sonorant onsets for both age groups (three-year-olds,  $F(1, 101) = 15.36, p < 0.001, \eta_p^2 = 0.13$ ; four-year-olds,  $F(1, 110) = 16.43,$

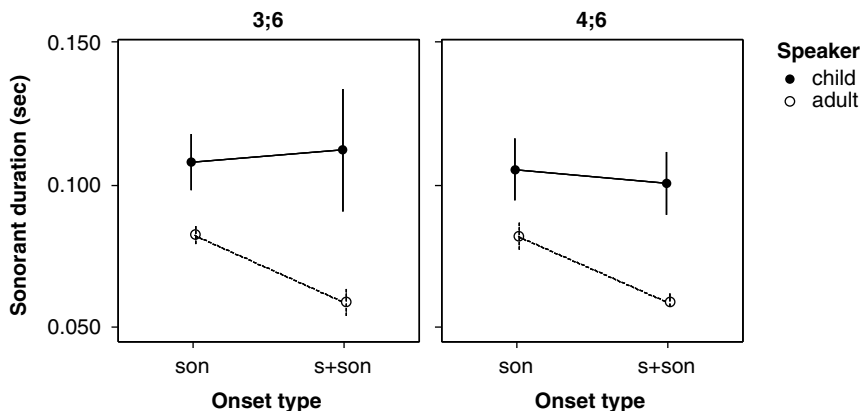


Fig. 2. Sonorant duration for child and adult productions of single words with simple sonorant and complex /s/+sonorant onsets for the different age groups.

$p < 0.001$ ,  $\eta_p^2 = 0.13$ ). There were no other significant effects, even in an analysis that included sonorant type (i.e. liquid versus nasal) as an additional factor.

Figure 2 shows that both three-year-olds and four-year-olds produced sonorants with the same duration in both singleton and /s/+sonorant onsets, in contrast to the adults who produced longer sonorants in singleton position than in clusters. Post hoc analyses on the adult data confirmed that the difference between singleton sonorants and sonorants in clusters was significant for both groups of parents ( $\alpha = 0.0125$ : three-year-old group,  $p = 0.001$ ; four-year-old group,  $p < 0.001$ ).

*Aspiration duration as a function of onset type.* A different pattern of results was obtained for stop aspiration duration, as shown in Figure 3. The (2) speaker  $\times$  (2) onset type ANOVA indicated that the effect of speaker was non-significant in both age groups (three-year-olds,  $F(1, 99) = 1.08$ ,  $p > 0.1$ ; four-year-olds,  $F(1, 111) = 0.19$ ,  $p > 0.1$ ), but both age groups showed a highly significant effect of onset type (three-year-olds,  $F(1, 99) = 50.85$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.34$ ; four-year-olds,  $F(1, 111) = 357.42$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.76$ ): stop aspiration duration was longer in singleton onsets than in /s/+stop clusters. The interaction between speaker and position was not significant in either age group, as is evident from Figure 3.

## DISCUSSION

The results from Experiment 1 indicate that the phonetic correlates of English syllable structure for /s/C sequences, which provide known cues

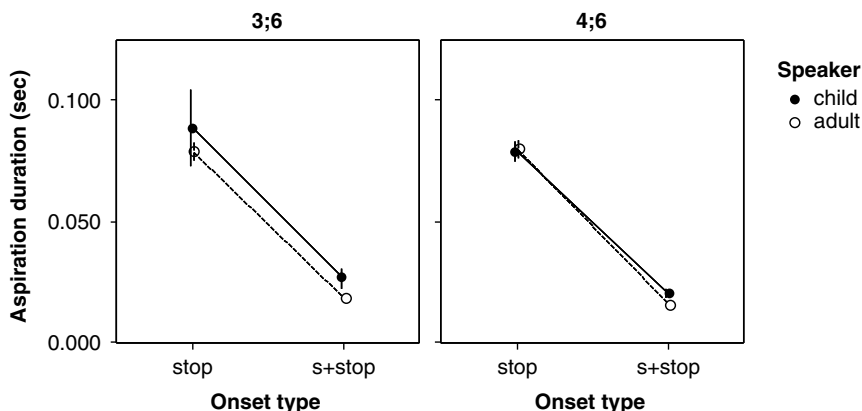


Fig. 3. Stop aspiration duration for child and adult productions of single words with simple stop and complex /s/+stop onsets for the different age groups.

to word boundary location, are acquired differently. The results for /s/ and sonorant duration suggest that the durational patterns are acquired separately from syllable structure, and surprisingly late. In contrast, the results for stop aspiration duration suggest that the different voiceless stop allophones (aspirated versus unaspirated) are acquired with syllable structure. With respect to the question of articulatory timing control sketched in the introduction, the differential acquisition of the durational and allophonic patterns suggest that these may be specified at different levels in the speech plan. Specifically, the early acquisition of allophonic variation with syllable structure may indicate that stop aspiration duration is specified within the word. In contrast, the late acquisition of word-edge durational patterns may indicate that these patterns are specified at the phrase level. If this is correct, then it may be that the durational patterns are acquired in multiword phrases, where juncture is more relevant, before they are evident in the production of single words.

## EXPERIMENT 2

This experiment investigated whether preschool children produce the durational and allophonic juncture cues in two word phrases. Experiment 1 indicated that control over aspiration duration emerges with the ability to produce different syllable structures. This result was interpreted to mean that allophonic variation in stop aspiration is controlled at the word level. If this is the case, then there is no reason to suspect that a word boundary would augment or interfere with preschool children's ability to produce this particular juncture cue.

TABLE 2. *The two-word phrases used in Experiment 2*

<i>/s/+sonorant</i>		<i>/s/+stop</i>	
onset cluster	cross boundary	onset cluster	cross boundary
<i>my slide</i>	<i>nice light</i>	<i>my spot</i>	<i>nice pot</i>
<i>icky smelly</i>	<i>ice melting</i>	<i>I stop</i>	<i>nice top</i>
<i>my snot</i>	<i>nice knot</i>	<i>great Scott</i>	<i>nice cot</i>
<i>bitty snail</i>	<i>this nail</i>		

In contrast to the allophonic juncture cue, the durational cue does not appear to be acquired with syllable structure. Experiment 1 indicated that four-year-olds have gained some control over segmental duration in that, like adults, they produced systematic variation in /s/ duration as a function of onset type. However, four-year-olds had not acquired the more robust pattern of variation in sonorant duration. If this result indeed indicates that the durational pattern is not intrinsic to syllable structure, then the pattern must be explained without reference to structure. The alternative explanation presented in the introduction was that the durational patterns are suprasyllabic and listener-oriented in their origins. If this is the case, then it is possible that four-year-olds may produce the patterns in clear speech when juncture is otherwise ambiguous. Specifically, four-year-olds may produce the syllable-related durational patterns to signal boundaries in multiword phrases before they produce them in single-word utterances, where boundary marking is less relevant.

#### METHOD

##### *Participants*

The participants were the same 20 children as in Experiment 1. The adult participants were the 20 parents and the 3 experimenters, who interacted with the children during the imitation task described below.

##### *Stimuli*

The stimuli were the 14 word pairs shown in Table 2. The matched pairs of two-word phrases were controlled for stress on either side of the word, a factor that is known to affect acoustic duration (Klatt, 1976). The stimuli were designed to be somewhat meaningful in the absence of a sentential context. They were also designed using component words expected to be familiar to all three- and four-year-olds. Both design features were to ensure that children would recognize that the stimuli consisted of two separate

words. On the other hand, these design features led to some asymmetry in vowel quality for two of the /s/+sonorant pairs (i.e. *ice melting* vs. *icky smelly* and *this nail* vs. *bitty snail*) as well as to the mistaken matching of the intervocalic onset–offset sequence in *nice cot* (-s#k-) with the post-consonantal onset cluster in *great Scott* (-t#sk-). Although these asymmetries were orthogonal to the principal comparison between child and adult productions, they could conceivably affect segmental duration. Accordingly, the critical comparisons of segment durations by speaker and boundary type for adult and child productions are reported with and without these production data.

### *Procedures*

The target two-word phrases were written down and randomly interspersed with 48 other words (stimuli for a separate study) on a sheet of paper that the experimenter took with her into the experimental room. When the time came for the experimenter to elicit these words, she told the child that they would now be playing a silly word game. The experimenter told the child that she would say some silly words, and the child should repeat these silly words back to her. In a few cases ( $N = 5$ ), the child was uncomfortable with repeating the words back to the experimenter, so the parent would model the words for the child and the child would repeat them back to the parent. All parents also read the phrases in a word list, as described in Experiment 1. The adult productions from the silly word game (experimenter and parents) were compared to parents' read productions to ascertain that all adults produced the words in the same way. However, only the silly word game productions were compared with child productions. These adult productions will be referred to from now on as the modeled productions.

### *Measurements*

The measurement criteria were the same as in Experiment 1. All stimuli were measured, but the analyses only included accurate productions of the /s/C sequences. That is, if a child produced only one consonant in a sequence, then that token was eliminated from the analysis. All other data produced by the child was included in the analysis.

As in Experiment 1, more three-year-olds than four-year-olds produced only one consonant when two were required ( $N = 5$  versus  $N = 2$ ). However, only one three-year-old and one four-year-old systematically produced singletons in place of onset clusters. All other children produced at least some onset clusters correctly. Overall, 23 word pairs were excluded from the three-year-old analyses (15 with onset clusters and 8 with singleton

offsets and onsets) and 8 tokens were excluded from the four-year-olds' analyses (all tokens with /s/C onsets).

Again, 10 percent of the data were randomly selected and measured by a second rater to assess measurement consistency according to the criteria. The mean differences (and standard deviations) between rater measurements for the three-year-old data were  $10.4(\pm 15.3)$  milliseconds for the child data and  $6.9(\pm 6.5)$  milliseconds for the adult data. The mean differences for the four-year-old data were  $6.7(\pm 6.2)$  milliseconds for the child data and  $8.6(\pm 7.8)$  for the adult data. Inter-rater correlations were extremely high, as in Experiment 1 ( $r=0.95$  and  $r=0.99$  for the three-year-old child and adult data respectively, and  $r=0.99$  and  $r=0.98$  for the four-year-old child and adult data respectively).

## RESULTS

As in Experiment 1, the child and adult productions were matched and the data were split by age. In this way, the analyses could preserve information about developmental change while focusing on the child–adult comparison. Again, similarities in child and adult productions would suggest that children had acquired the phonetic boundary patterns, whereas differences would indicate that they had not. The /s/+sonorant and /s/+stop sequences were analyzed separately, in keeping with the different nature of the juncture cues in these two sequences. The analyses showed that four-year-olds produced both the durational and allophonic juncture cues in two-word utterances, albeit somewhat less robustly than adults. In contrast, three-year-olds only produced the allophonic juncture cue appropriately. The results on /s/+sonorant sequences are presented first, followed by the results on /s/+stop sequences.

*Boundary effects on /s/ and sonorant durations.* As a preliminary to the comparison between child and adult productions, the modeled and read productions were compared to evaluate consistency across adult productions. A (2) speaking condition  $\times$  (2) boundary ANOVA showed a significant effect of speaking condition on /s/ duration and on sonorant duration (/s/ duration,  $F(1, 312)=5.34$ ,  $p=0.022$ ,  $\eta_p^2=0.02$ ; sonorant duration,  $F(1, 312)=24.57$ ,  $p<0.001$ ,  $\eta_p^2=0.07$ ): segmental durations were longer overall in modeled speech than in read speech. The interaction between condition and boundary was also significant for sonorant durations ( $F(1, 312)=9.66$ ,  $p=0.002$ ;  $\eta_p^2=0.03$ ), but this was due to a quantitative difference in the pattern of results rather than to a qualitative one. Specifically, sonorants were longer in singleton position in modeled speech than in read speech (mean(*SD*):  $105(\pm 35)$  ms vs.  $85(\pm 30)$  ms), but speakers produced shorter (and roughly equivalent) sonorants in /s/+sonorant clusters under both speaking conditions ( $67(\pm 17)$  ms. vs.  $62(\pm 25)$  ms). The

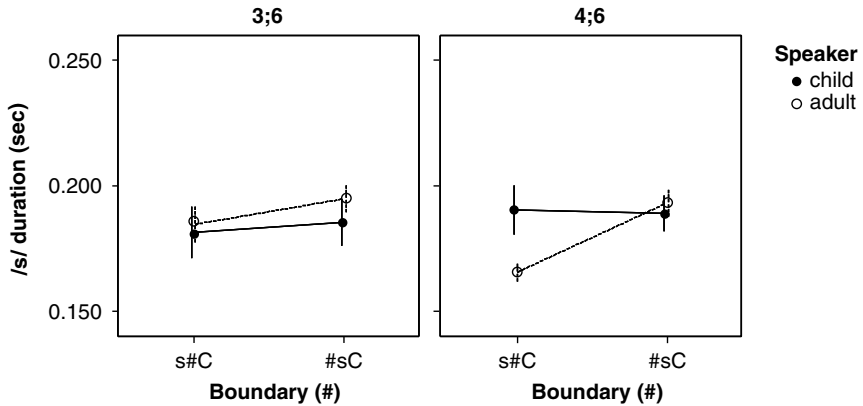


Fig. 4. /s/ duration for child and adult productions of two-word phrases in which /s/ serves either as a singleton offset (s#C) or is part of an onset cluster (#sC). The results for the three- and four-year-old groups are shown in the left- and right-hand panel, respectively.

difference between singleton durations in modeled and read speech is probably attributable to a difference in the tasks: the modeled words were produced by the experimenter or parent for the child during the silly word game, whereas the read words were produced by parents at different intervals during the experiment while the child was kept occupied by placing stickers on a sheet of paper.

The modeled productions were next compared to matching three- and four-year-old productions. A (2) speaker  $\times$  (2) boundary ANOVA revealed no significant effects of speaker or boundary on /s/ durations in the different age groups. However, there was a significant interaction between speaker and boundary in the four-year-old group ( $F(1, 148) = 4.01, p = 0.047, \eta_p^2 = 0.03$ ), which was unchanged when the *melting/smelly* and *nail/snail* productions were excluded from the analysis ( $F(1, 72) = 4.08, p = 0.047, \eta_p^2 = 0.04$ ). Figure 4 shows that the adults produced shorter /s/ in offset position than in onset position when interacting with the four-year-olds, but not when interacting with the three-year-olds. This is because the adults sometimes lengthened word-final /s/ when speaking to three-year-olds, presumably to provide the child with a more salient cue to word boundary location.

A different pattern of results was found in the analyses on sonorant duration, as shown in Figure 5. The (2) speaker  $\times$  (2) boundary ANOVA indicated that sonorant duration varied with speaker and boundary in the three-year-old group (speaker,  $F(1, 143) = 4.54, p = 0.035, \eta_p^2 = 0.03$ ; boundary,  $F(1, 143) = 6.42, p = 0.012, \eta_p^2 = 0.04$ ), but only with boundary in the four-year-old group ( $F(1, 148) = 36.06, p < 0.001, \eta_p^2 = 0.20$ ). The



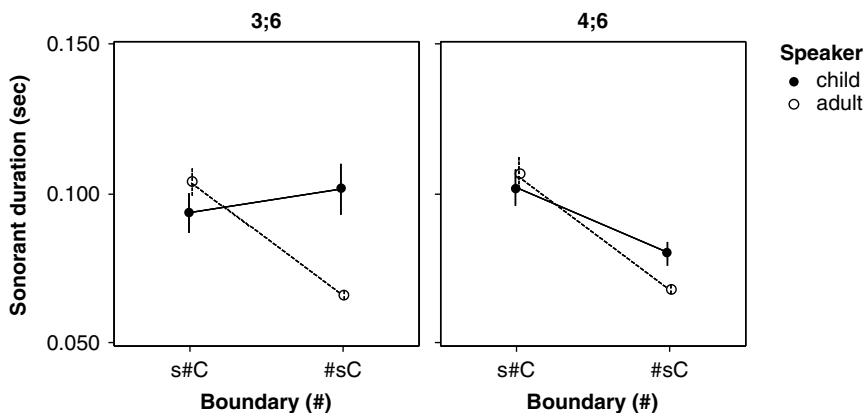


Fig. 5. Sonorant duration for child and adult productions of two-word phrases in which the sonorant serves either as a singleton onset (s#C) or is part of an onset cluster (#sC). The results for the three- and four-year-old groups are shown in the left- and right-hand panel, respectively.

interaction between speaker and boundary was highly significant in the three-year-old age group ( $F(1, 143) = 15.09$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.10$ ), but not significant in the four-year-old age group. These results were not affected by sonorant type (liquid vs. nasal). That is, this factor was not significant nor did it interact with any of the other factors when it was added to the analysis. When the *melting/smelly* and *nail/snail* productions were excluded, the interaction between speaker and boundary remained significant in the three-year-old group ( $F(1, 70) = 13.67$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.16$ ) and approached significance in the four-year-old group ( $F(1, 72) = 3.79$ ,  $p = 0.055$ ,  $\eta_p^2 = 0.05$ ). Also, when these productions were excluded, the simple effect of speaker was no longer significant for the three-year-old age group, though the simple effect of boundary remained in the four-year-old age group ( $F(1, 74) = 20.22$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.22$ ).

The different statistical results for three- and four-year-olds corresponded to a clear qualitative difference between the children's productions. Figure 5 shows that the younger children did not produce a systematic difference in sonorant duration as a function of boundary location, but older children clearly did. The nearly significant interaction between speaker and boundary in one analysis of the four-year-old group was due to a quantitative difference in the pattern – child productions of sonorants in C2 position within a cluster were longer than adult productions of sonorants in the same position.

In sum, the results on /s/ duration show a relatively weak boundary effect in adult speech, so it is perhaps not surprising that preschool children

do not show an effect of boundary either. By contrast, the results on sonorant duration indicate a strong boundary effect in adult speech, but only four-year-olds show a similar effect in their productions. The younger children are still not able to produce the durational pattern that marks word boundary location in /s/+sonorant sequences.

*Boundary effects on /s/ and aspiration duration.* The preliminary (2) speaking condition  $\forall$  (2) boundary ANOVA that compared modeled and read productions of /s/+stop sequences revealed a significant effect of speaking condition on /s/ duration ( $F(1, 231) = 9.70$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.04$ ) and on aspiration duration ( $F(1, 232) = 12.83$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.05$ ). Consonants were longer overall in modeled speech than in read speech. There was also a significant interaction between condition and boundary on aspiration duration ( $F(1, 232) = 13.33$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.05$ ). Like the condition by boundary interaction on sonorant duration, the interaction on aspiration duration was due to a quantitative difference in the pattern – adults produced stops with greater aspiration in singleton onset position when modeling the two-word phrase for a child than when reading the phrase from the word list ( $101(\pm 28)$  ms vs.  $84(\pm 20)$  ms), but they produced stops with equally short aspiration in /s/+stop clusters under both speaking conditions ( $22(\pm 11)$  ms vs.  $20(\pm 15)$  ms). Again, the differences between the speaking conditions are probably attributable to the presence or absence of a child interlocutor.

A (2) speaker  $\forall$  (2) boundary ANOVA comparing child and adult /s/ durations in /s/+stop sequences revealed a similar pattern of results as those obtained for /s/+sonorant sequences (see Figure 4). There was a significant effect of speaker on /s/ durations in the three-year-old group ( $F(1, 104) = 11.44$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.10$ ) and a significant effect of boundary on /s/ durations in the four-year-old group ( $F(1, 113) = 9.65$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.08$ ): the three-year-olds produced /s/ with shorter duration overall than the adults did; and both the adults and the four-year-olds produced shorter /s/ offsets and longer /s/ onsets. The absence of a significant boundary effect on /s/ duration in the three-year-old age group was due to the fact that adults again lengthened word-final /s/ when speaking to the younger children. The interactions between speaker and boundary were not significant for either group. Identical results were obtained when the *cot/Scott* productions were excluded: the effect of speaker was significant for the three-year-old group ( $F(1, 68) = 14.72$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.18$ ); the effect of boundary was significant for the four-year-old group ( $F(1, 74) = 12.81$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.15$ ); and the interaction between speaker and boundary was not significant for either group.

Similarly small group differences were evident in the pattern of results for aspiration duration, but overall the results were as expected (Figure 6). Although the (2) speaker  $\forall$  (2) boundary ANOVA indicated a significant

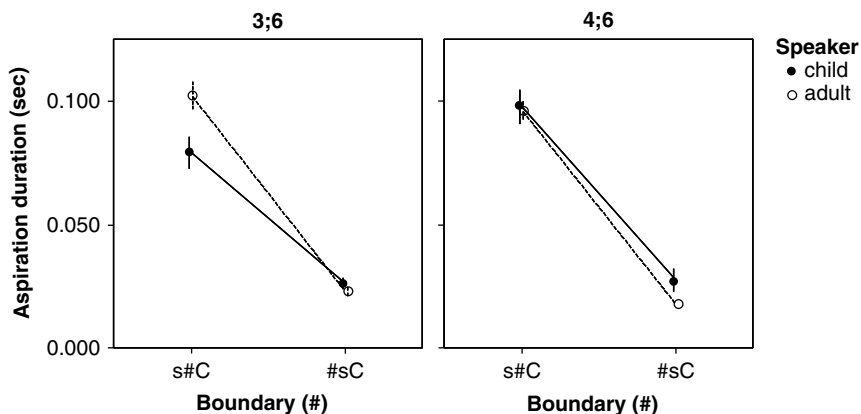


Fig. 6. Stop aspiration duration for child and adult productions of two-word phrases in which the stop either serves as a singleton onset (s#C) or is part of an onset cluster (#sC). The results for the three- and four-year-old groups are shown in the left- and right-hand panel, respectively.

effect of speaker and a significant interaction between speaker and boundary in the three-year-old group (speaker,  $F(1, 104) = 5.89$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.05$ ; speaker  $\times$  boundary,  $F(1, 104) = 7.87$ ,  $p = 0.006$ ,  $\eta_p^2 = 0.07$ ), the effect of boundary was highly significant in both age groups (three-year-old group,  $F(1, 104) = 157.54$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.60$ ; four-year-old group,  $F(1, 113) = 215.01$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.66$ ). The effect of speaker and the interaction between speaker and boundary disappeared from the three-year-old group when the *cot/Scott* productions were excluded, but the effect of boundary remained highly significant in both age groups (three-year-old group,  $F(1, 68) = 103.05$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.60$ ; four-year-old group,  $F(1, 76) = 164.95$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.69$ ).

Figure 6 clearly shows that three- and four-year-olds produced the expected pattern of aspirated singleton stops (s#C) and unaspirated stops in /s/C onsets (#sC). In addition, the figure shows the effect of speaker and the interaction between speaker and boundary that was obtained in an overall analysis of the three-year-old group. The fact that these effects disappear when the *cot/Scott* productions are excluded suggests that differences in the duration of child and adult aspiration durations were relatively unimportant.

## DISCUSSION

The results of Experiment 2 confirm that the allophonic juncture cue is acquired with syllable structure, and suggest that the durational cue is

acquired in a phrasal context. It is also clear from the results of Experiments 1 and 2 that the durational cue to word boundaries is acquired slowly. Although the four-year-olds in Experiment 2 produced singleton sonorant onsets with greater duration than sonorants in /s/+sonorant clusters, the duration differences were less pronounced than the duration differences produced by the adults (see Figure 5). In addition, Experiment 1 indicates that, unlike adults, four-year-olds do not spontaneously produce the duration difference at the edge of isolated words. Further, it is clear from the results of Experiments 1 and 2 that the durational cue is subtler than the allophonic cue to boundary location.

### EXPERIMENT 3

The primary goal of Experiment 3 was to examine whether the production differences observed in Experiment 2 were perceptually robust. The developmental effect was of specific interest: would the difference between the three- and four-year-olds' ability to produce the durational pattern translate into age-dependent differences in listener judgments of boundary location? If so, then the perceptual experiment would support the findings from Experiment 2. If not, then we would have reason to question whether four-year-olds have truly acquired the durational juncture cue.

A secondary goal of Experiment 3 was to determine whether the durational cue and the allophonic cues were the most important phonetic cues to boundary location for /s/C sequences. Would the perceptual results provide an accurate reflection of the magnitude of boundary-dependent differences in durational values for the /s/C sequences of child and adult speech? Or, would boundary-dependent differences in listener judgments be out of proportion to, and so unexplainable from, the measured acoustic differences?

#### METHOD

##### *Participants*

Fourteen undergraduate students from the University of Oregon participated in the experiment for course credit. All students were monolingual, native American-English speakers with normal hearing.

##### *Stimuli*

Only the child and parent (read) two-word phrases were used in Experiment 3. The modeled phrases were not used: (1) to reduce the overall number of stimuli that listeners would need to judge; and (2) because they were generated by fewer speakers overall (3 experimenters and 5 parents as

opposed to the 20 parents who were recorded while reading the stimuli from a word list). A total of 541 stimuli were generated from all the available child and parent productions in the following way.

The utterances were edited to eliminate all lexical cues to boundary location. Specifically, the only portion of the utterance that was preserved was the /s/C sequence with half of the preceding vowel on either side of the sequence. So, for example, the utterance *this nail* was edited to yield [Isne], a VCCV stimulus. The first half of the vowel preceding the consonant sequence (V<sub>1</sub>) and the second half of the vowel following the sequence (V<sub>2</sub>) were deleted to eliminate consonantal transitions that might provide access to the lexical item (e.g. transitions from [ð]) could hint at the original lexical item *this*). Specifically, the midpoints of the vowels on either side of the sequence were identified. Everything to the left of V<sub>1</sub>'s midpoint was deleted, and everything to the right of V<sub>2</sub>'s midpoint was deleted. The overall amplitude of the VCCV stimuli was normalized across all speakers.

By preserving half of V<sub>1</sub> and V<sub>2</sub>, we preserved some information regarding speech rate, but we also preserved information regarding vowel duration and transitions into and out of the sequence. The literature indicates that such information provides poor cues to boundary location (e.g. Boucher, 1988; Turk & Shattuck-Hufnagel, 2000), but in the event that listener judgments on the child speech tokens were more accurate than might be expected from the acoustic measures taken in Experiment 2, we would be able to test the vowels' contribution to the perception of boundary location.

### *Procedure*

Listeners were asked to make boundary decisions on the VCCV stimuli. They were informed that the stimuli were edited versions of two-word utterances, and they were given examples of each of the pairs of two-word utterances that differed in boundary location (e.g. *this nail* versus *bitty snail*). Listeners were then told to decide whether the /s/ of the VCCV stimuli belonged to the first word (e.g. *this*) or to the second word (e.g. *snail*).

A maximum of two listeners at a time were seated in a small experimental room. Each listener sat in front of a desktop computer that controlled the presentation of the stimuli. Stimuli were randomly presented over headphones, and listeners were able to adjust the volume to a comfortable listening level. Listener boundary decisions were recorded as button presses. Listeners were to press the '1' button if they thought that the /s/ belonged to the first word, and they were to press the '2' button if they thought it belonged to the second word.

*Analyses*

The boundary judgments were coded as 0 (s#C) or 1 (#sC). Multiple logistic regression was then used to predict the categorical judgments on /s/+sonorant or /s/+stop sequences according to age group (3;6 or 4;6 or adult), speaker (child vs. adult) and boundary (s#C vs. #sC). The Wald test statistic is reported for the different predictor variables with the related *p* and Exp(B) values. Exp(B) is the odds ratio for different predictor variables. An odds ratio of 1 indicates that different values of the predictor variable did not affect the outcome variable; deviation from 1 indicates the strength and direction of change in the outcome variable given different values of the predictor variable. Planned comparisons were also conducted to evaluate specific differences in listener judgments as a function of boundary location for each of the age groups and sequence types. Additionally, the perceptual judgments were correlated with the durational pattern and with the allophonic pattern to further investigate the relationship between the perception and production results. To parallel Experiment 2, all analyses were conducted only on those judgments pertaining to tokens with both consonants of the /s/C sequence. To further parallel Experiment 2, results are also reported for analyses in which responses on the asymmetric word pairs were excluded (i.e. *melting/smelly*, *nail/snail*, *cot/Scott*).

RESULTS

Overall, the perceptual results paralleled the acoustic results: listeners distinguished between singleton and /s/+sonorant onsets in four-year-olds' speech, but not in three-year-olds' speech; and they were more accurate in segmenting /s/+sonorant sequences produced by adults than those produced by children. Such results suggest that even small differences in consonantal duration have perceptual consequences. The results for /s/+stop sequences were also as expected from the acoustic data: listeners showed more overall accuracy in segmenting /s/+stop sequences than /s/+sonorant sequences, regardless of age. All of the results are presented in more detail below.

*Listener performance on /s/+sonorant sequences.* A multiple logistic regression was used to predict listener boundary judgments on /s/+sonorant sequences according to age group, speaker and boundary. Figure 7 shows the results from this analysis. Listener boundary judgments varied predictably as a function of speaker and boundary (Wald test statistics: speaker = 19.08, *p* < 0.001, Exp(B) = 0.56; boundary = 64.84, *p* < 0.001, Exp(B) = 0.34) and their interaction (Wald test statistic = 13.79, *p* < 0.001, Exp(B) = 1.95). The 3-way interaction with age group was also significant (group × speaker × boundary = 5.46, *p* = 0.019, Exp(B) = 1.81), as shown in Figure 7. These results did not change when responses to the *melting/smelly*

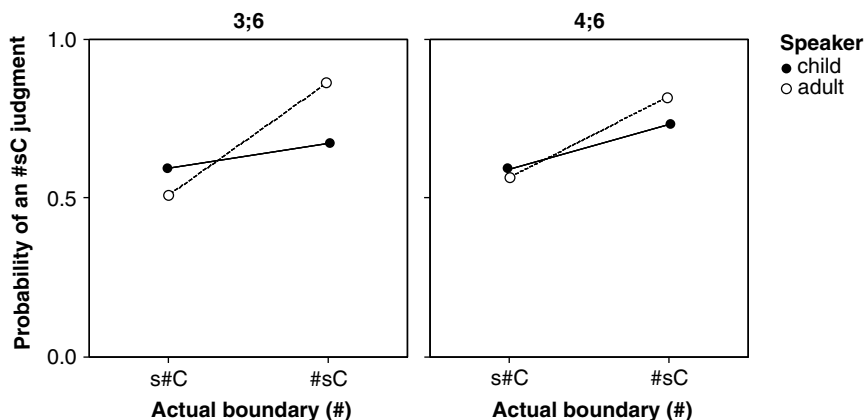


Fig. 7. The probability of an onset cluster judgment as a function of speaker (child or adult), boundary location (#sC or s#C) and age group (three- vs. four-year-olds).

and *nail/snail* word pairs were excluded, though the simple effects of speaker and boundary were somewhat weakened (Wald test statistics: speaker = 6.20,  $p = 0.013$ ,  $\text{Exp}(B) = 0.59$ ; boundary = 39.95,  $p < 0.001$ ,  $\text{Exp}(B) = 0.28$ ; group  $\times$  speaker  $\times$  boundary = 5.13,  $p = 0.023$ ,  $\text{Exp}(B) = 2.42$ ).

The 3-way interaction between age group, speaker and boundary suggests that, just as listeners were better able to correctly segment /s/+sonorant sequences produced by adults than those produced by children (i.e. the significant 2-way interaction between speaker and boundary), so too were they more able to correctly segment sequences produced by four-year-olds compared to those produced by three-year-olds. Planned comparisons confirmed this and showed further that whereas boundary predicted listener judgments in the four-year-olds' data, it did not in the three-year-olds' data (Wald test statistics: four-year-olds = 10.67,  $p = 0.001$ ,  $\text{Exp}(B) = 0.67$ ; three-year-olds = 1.08,  $p > 0.1$ ). This result is evident in Figure 8, which compares listener responses on the stimuli derived from child productions.

*Listener performance on /s/+stop sequences.* Once again, a multiple logistic regression was used to predict listener boundary judgments according to age group, speaker and boundary. Figure 9 shows that listener boundary judgments on /s/+stop sequences varied predictably as a function of each of the predictor variables (Wald test statistics: speaker = 43.60,  $p < 0.001$ ,  $\text{Exp}(B) = 0.33$ ; boundary = 308.79,  $p < 0.01$ ,  $\text{Exp}(B) = 0.03$ ) and the 2-way interaction (Wald test statistics: speaker  $\times$  boundary = 50.35,  $p < 0.001$ ,  $\text{Exp}(B) = 5.54$ ), even when responses on the asymmetric *cot/Scott* word pair were excluded (Wald test statistics: speaker = 23.58,  $p < 0.001$ ,

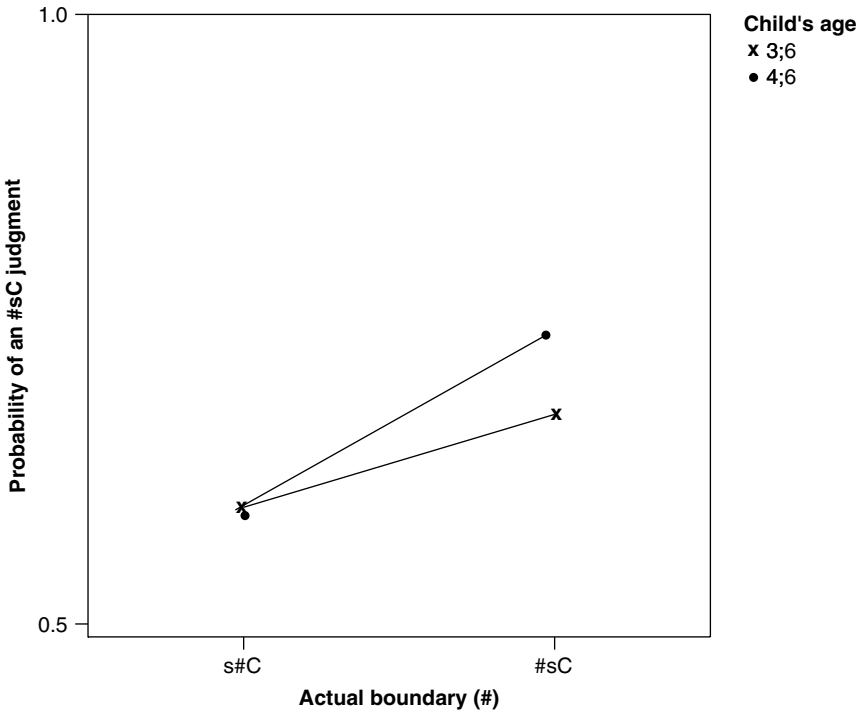


Fig. 8. The probability of an onset cluster judgment as a function of age and boundary location. Children are the only speakers.

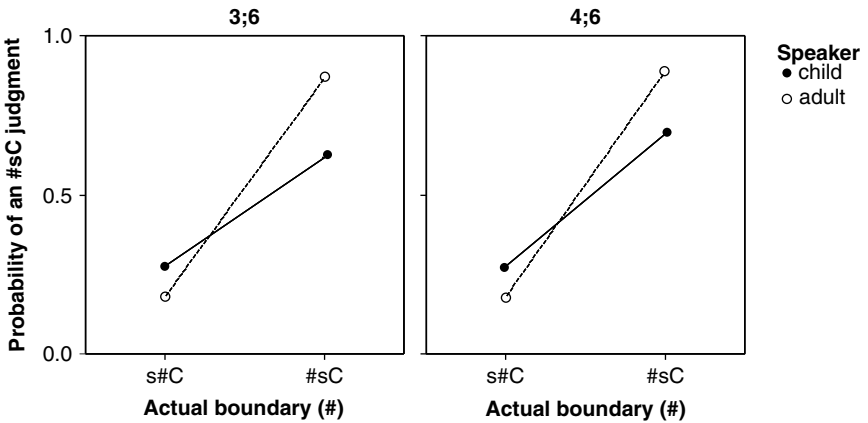


Fig. 9. The probability of an onset cluster judgment shown as a function of speaker (child or adult), boundary location (#sC or s#C) and age group (three- vs. four-year-olds).



Exp(B)=0.33; boundary=199.85,  $p < 0.01$ , Exp(B)=0.03; speaker  $\times$  boundary=30.84,  $p < 0.001$ ). Again, listeners segmented parent productions more accurately than child productions, but listeners were highly accurate even on the child productions. Importantly, age group was not a significant predictor of listener behavior and did not interact with other predictor variables, indicating that listener judgments did not vary with the age of the child speaker or their parents.

In a final set of analyses, we calculated the correlations between sonorant duration and average listener judgments on the stimuli with /s/+sonorant sequences and between aspiration duration and average listener judgments on /s/+stop sequences. The goal was to ascertain whether or not the perceptual results can be reasonably attributed to the segmental duration and allophonic differences associated with singleton versus complex onsets. Although all correlations were significant, the  $r$  values for children were lower than for adults and the  $r$  values for /s/+sonorant sequences were lower than those for /s/+stop sequences. The correlation between the acoustic variable and listener judgments on the child productions of /s/+sonorant sequences was  $-0.21$  ( $p = 0.012$ ) and on the child production of /s/+stop sequences it was  $-0.63$  ( $p < 0.001$ ). For adults, the correlations were  $-0.50$  ( $p < 0.001$ ) and  $-0.85$  ( $p < 0.001$ ) for /s/+sonorant and /s/+stop sequences respectively. In other words, the relative strength of the different correlations patterned with the different effect magnitudes shown in Figures 7 and 9. Such patterning suggests that listener boundary judgments were indeed based on the durational and allophonic juncture cues; these cues were simply more or less present in the different types of stimuli produced by children and adults.

## DISCUSSION

Overall, the results from Experiment 3 support those from Experiment 2. First, listeners more accurately distinguished boundary location in /s/+sonorant sequences of adult speech than in those of child speech, which is consistent with the acoustic results showing that the durational cue to boundary location is stronger in adult speech than in child speech. Second, listener boundary judgments on /s/+sonorant sequences varied systematically with boundary location when these were produced by four-year-olds, but not when they were produced by three-year-olds. Such a result is consistent with the finding from Experiment 2 that the durational juncture cue is emergent in four-year-olds' speech and absent in three-year-olds' speech. Third, listener boundary judgments were most accurate on stimuli with /s/+stop sequences whether these were produced by adults or children. This result is consistent with the acoustic data presented in Experiment 2: the difference between singleton aspirated stops and unaspirated stops in

onset clusters is greater than the difference between singleton sonorants and sonorants in onset clusters. Further, even three-year-olds produce large differences in stop aspiration as a function of onset type, just like adults.

#### GENERAL DISCUSSION

This study indicates that children acquire the durational and allophonic cues to word boundaries in /s/C sequences at different times. Preschool children acquire allophonic variation in stop aspiration duration with the ability to produce simple and complex syllable onsets. Even the youngest preschool children produced singleton stop onsets with much greater aspiration than cluster-internal stops in single words as well as in two-word phrases. In contrast to the allophonic cue, preschool children acquire the durational juncture cueing pattern well after they have acquired the relevant syllable structures. Although most of the three-year-olds could produce /s/C onset clusters and all could produce /s/C offset-onset sequences, three-year-olds did not distinguish boundary location in /s/+sonorant sequences. Unlike adults, the three-year-olds produced singleton sonorant onsets and sonorants in onset clusters with equal durations. In contrast to the younger children, the four-year-olds distinguished boundary location by producing longer singleton sonorants and shorter cluster-internal sonorants like adults. But, unlike adults, these older children produced the durational cue only in two-word phrases. Even then, the duration differences produced by four-year-olds were not as large as in adult speech and so the pattern was less effective at cueing word boundary perception than the adult pattern. In the rest of this section, we discuss these overall results with respect to the general aims of the study.

The aims of the current study were to understand the acquisition of phonetic juncture cues and the nature of what is being acquired. Two specific hypotheses were proposed: (1) the cues are intrinsic to syllable structure and so are acquired with the ability to produce singleton and complex onsets; (2) the cues are suprasyllabic, existing to mark word boundaries in English, and so are acquired in multiword utterances after the child has had extensive practice with the relevant syllable structures. The results support both hypotheses. In particular, the allophonic cue appears to be tied to syllable structure, in that it is acquired with the ability to produce singleton stop onsets and /s/+stop onset clusters. On the other hand, the durational cue appears to be suprasyllabic, in that it is acquired late and first in two-word phrases with ambiguous boundaries.

The hypothesis that the durational cue to word boundaries is suprasyllabic contrasts with the traditional, syllable-based explanation for this cue. More importantly, the different explanation implies a different model of articulatory timing control. The syllable-based explanation implies a

single level of control where segmental duration patterns are attributed either to syllable-size temporal frames (Campbell & Isard, 1991; Klatt, 1976; Lehiste, 1970) or to intergestural articulatory timing routines within the syllable (Browman & Goldstein, 1988; Krakow, 1999). By contrast, a suprasyllabic explanation suggests layered control over articulatory timing. Segmental duration patterns are explained to result from focal changes in articulatory timing around a boundary between two words.

Although a model with layered timing control is more complex than one with just a single level of control, layered control can also be modeled quite simply. For instance, Byrd & Saltzman (2003) argue that articulatory timing above the level of the word can be achieved by controlling one additional parameter, which they label  $\pi$ . The  $\pi$  parameter is described as a clock-like mechanism that phases intergestural timing. The clock slows at the boundary between two units (e.g. words) and picks up speed after the boundary. Further, Byrd and colleagues have recently suggested that this simple manipulation in clock rate is sufficient to account for the durational patterns that distinguish singleton onsets from complex onsets (Byrd, Lee, Riggs & Adams, 2005).

If layered timing control is as simple as controlling one additional parameter, then we might wonder why this control is acquired so slowly. Even the youngest preschool children in this study had been speaking in multiword phrases for over a year, so why were they unable to slow intergestural timing across a word boundary in /s/+sonorant sequences? And, why were the four-year-olds not able to do so as successfully as the adults? The problem is not one of intergestural timing per se. After all, the allophonic cue requires fine control over the timing of the laryngeal gesture (voicing) relative to the timing of the stop release – and three-year-olds are clearly capable of this. Instead, the answer may reside in the different natures of the allophonic and durational cues. The allophonic cue is signaled by stop aspiration duration, which varies systematically with syllable position, but not with other variables such as speech style or position within the utterance. Thus, when a child hears the voiceless stop in *spider*, he always hears an unaspirated voiceless stop and learns to pronounce it accordingly (i.e. the word is [spɪɪdɔ̃] and never [sp<sup>h</sup>ɪɪdɔ̃]). By contrast, the durational cue is a relational cue: sonorant duration only signals one type of onset structure or another when it is long or short compared to adjacent segmental durations. Further, the extent to which sonorant duration is likely to be produced as long or short relative to other segments will depend on speech style and on the ambiguity of word boundary location. In addition, the results from Experiment 3 suggest that the durational cue is relatively subtle when compared to a categorical cue such as the presence or absence of stop aspiration. All this suggests that the acquisition of the durational cue to boundary location is slowed because

the child is not learning the pattern of a particular word so much as a pattern that exists only under certain conditions. To properly instantiate the cue, it is likely that the child must be sensitive to the cue and its effect in order to create a focal change in articulatory timing around a boundary. A prediction that follows from this line of thought is that when older children (or adults) speak casually or quickly, they do not monitor word boundaries and so do not highlight them. Instead, coarticulatory pressures take over and consonants are 'resyllabified' across boundaries. Such a prediction could be tested in future work by investigating the effects of speech style and rate on segmental duration patterns in child and adult speech.

In summary, the results show that the allophonic cue to word boundaries is acquired with syllable structure, but the relational cue is acquired slowly and in multiword phrases. The slow acquisition of the durational cue, in particular, suggests that phrase-medial word boundaries pose a problem for the acquisition of language production just as they do for the acquisition of language comprehension.

## REFERENCES

- Boersma, P. & Weenink, D. (2002). Praat: doing phonetics by computer (version 4.0.34) [Computer program]. Retrieved November 2002 from <http://www.praat.org/>
- Boucher, V. (1988). A parameter of syllabification for VstopV and relative timing invariance. *Journal of Phonetics* **16**, 299–326.
- Boysson-Bardies, B. de, Bacri, N., Sagart, L. & Poizat, M. (1981). Timing in late babbling. *Journal of Child Language* **8**, 525–39.
- Boysson-Bardies, B. de & Vihman, M. M. (1991). Adaptation to language: evidence from babbling and first words in four languages. *Language* **67**, 297–319.
- Browman, C. & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica* **45**, 140–55.
- Byrd, D., Lee, S., Riggs, D. & Adams, J. (2005). Interacting effects of syllable and phrase position on consonant articulation. *Journal of the Acoustical Society of America* **118**, 3860–73.
- Byrd, D. & Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* **31**, 149–80.
- Campbell, W. N. & Isard, S. D. (1991). Segment durations in a syllable frame. *Journal of Phonetics* **19**, 37–47.
- Cho, T. & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* **29**, 155–90.
- Christie, W. M. (1977). Some multiple cues for juncture in English. *General Linguistics* **17**, 212–22.
- Christophe, A., Dupoux, E., Bertoncini, J. & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America* **95**, 1570–80.
- Cutler, A., Mehler, J., Norris, D. & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* **35**, 385–400.
- Davidson-Nielsen, N. (1974). Syllabification in English words with medial *sp*, *st*, *sk*. *Journal of Phonetics* **2**, 15–45.
- Davis, B. L., MacNeilage, P. F., Matyear, C. L. & Powell, J. K. (2000). Prosodic correlates of stress in babbling: an acoustical study. *Child Development* **71**, 1258–70.

- DeMarco, S. & Harrell, R. M. (1995). Perception of word junctures by children. *Perceptual and Motor Skills* **80**, 1075–82.
- Grunwell, P. (1981). The development of phonology: a descriptive profile. *First Language* **3**, 161–91.
- Haggard, M. (1973). Correlations between successive segment durations: values in clusters. *Journal of Phonetics* **1**, 111–16.
- Keating, P., Wright, R. & Zhang, J. (1999). Word-level asymmetries in consonant articulation. *UCLA Working Papers in Phonetics* **97**, 157–73.
- Klatt, D. (1976). Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *Journal of the Acoustical Society of America* **59**, 1208–21.
- Kohler, K. J. (1991). The phonetics/phonology issue in the study of articulatory reduction. *Phonetica* **48**, 180–92.
- Krakow, R. (1999). Physiological organization of syllables: a review. *Journal of Phonetics* **27**, 23–54.
- Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.
- Lisker, L. & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word* **20**, 384–422.
- Morgan, J. L. & Saffran, J. R. (1995). Emerging integration of sequential and supra-segmental information in preverbal speech segmentation. *Child Development* **66**, 911–36.
- Oller, D. K. (1973). Speech segment duration in English. *Journal of the Acoustical Society of America* **54**, 1234–47.
- Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics* **20**, 331–50.
- Redford, M. A. (2007). Word-internal versus word-peripheral consonantal duration patterns in three languages. *Journal of the Acoustical Society of America* **121**, 1665–78.
- Redford, M. A. & Diehl, R. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *Journal of the Acoustical Society of America* **106**, 1555–65.
- Redford, M. A. & Randall, P. (2005). The role of juncture cues and phonological knowledge in English syllabification judgments. *Journal of Phonetics* **33**, 27–46.
- Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* **274**, 1926–8.
- Stoel-Gammon, C. & Dunn, J. (1985). *Normal and disordered phonology in children*. Baltimore, MD: University Park Press.
- Tuller, B. & Kelso, J. A. S. (1991). The production and perception of syllable structure. *Journal of Speech and Hearing Research* **34**, 501–8.
- Turk, A. & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics* **28**, 397–435.
- Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America* **61**, 846–58.