The emergence of mature gestural patterns in the production of voiceless and voiced word-final stops[a)]

Susan Nittrouer,[b)] Sandy Estee,[c)] Joanna H. Lowenstein, and Jennifer Smith
Utah State University

(Received 25 June 2003; revised 28 September 2004; accepted 12 October 2004)

The organization of gestures was examined in children's and adults' samples of consonant–vowel–stop words differing in stop voicing. Children (5 and 7 years old) and adults produced words from five voiceless/voiced pairs, five times each in isolation and in sentences. Acoustic measurements were made of vocalic duration, and of the first and second formants at syllable center and voicing offset. The predicted acoustic correlates of syllable-final voicing were observed across speakers: vocalic segments were shorter and first formants were higher in words with voiceless, rather than voiced, final stops. In addition, the second formant was found to differ depending on the voicing of the final stop for all speakers. It was concluded that by 5 years of age children produce words ending in stops with the same overall gestural organization as adults. However, some age-related differences were observed for jaw gestures, and variability for all measures was greater for children than for adults. These results suggest that children are still refining their organization of articulatory gestures past the age of 7 years. Finally, context effects (isolation or sentence) showed that the acoustic correlates of syllable-final voicing are attenuated when words are produced in sentences, rather than in isolation. [DOI: 10.1121/1.1828474]

PACS numbers: 43.70.Bk; 43.70.Ep; 43.70.Fq [AL] Pages: 351–364

I. INTRODUCTION

That young children do not produce speech as adults do is not news, but still, many questions remain concerning how children do eventually attain mature patterns of gestural organization. Concerning initial states, one clear trend is that children's earliest utterances are produced with greater constraints of the surrounding phonetic environment imposed on individual places of constriction for both vocalic and consonantal gestures. For example, this enhanced constraint was found in 1-year-olds' speech for segments involving lingual gestures such that close, front vowels and consonants with alveolar constrictions tended to co-occur, while close, back vowels and consonants with velar constrictions tended to co-occur (Davis and MacNeilage, 1990). In fact, this enhanced constraint of the phonetic environment on lingual gestures has been reported for children as old as 7 years (Nittrouer, 1993). Another trend observed in the earliest productions of children is that they avoid multisyllabic productions involving more than one constriction location for consonants, preferring instead to duplicate a single consonantal constriction within these utterances (e.g., Donahue, 1986; Oller, 1980). Finally, MacNeilage and Davis (1991) described children's earliest productions as deriving almost completely from jaw movements, with other articulators being tightly coupled to these movements. So, for example, variation in vowel quality derives from variation in the degree of jaw lowering, rather than from variation in lingual shape. All these examples suggest that one important developmental change that must be accomplished in order for children to acquire mature gestural patterns for speech production is independent control over articulators. Along with this developmental change, appropriate and precise ways of coordinating separate articulatory gestures must be learned to provide the kinds of stable interarticulator relations observed for skilled adult speech production (e.g., Tuller and Kelso, 1984).

Tracking the emergence of mature gestural organization is not easy because of methodological obstacles. Some investigators have described the gestural patterns of children's early speech production using narrow phonetic transcription (e.g., Ferguson and Farwell, 1975; Menn, 1978; Piske, 1997; Waterson, 1971), but these analyses are so intensive that it is difficult to do more than diary studies with them. Recent technological advances have made it possible to use acoustic analysis (e.g., Goodell, 1991; Katz, Kripke, and Tallal, 1991; Nittrouer, 1993; Sussman *et al.*, 1999) and direct kinematic measures (e.g., Green *et al.*, 2000; Smith and Goffman, 1998) with children's speech, but both methods have limitations. With acoustic analysis care must be taken to select stimuli for which there is minimal chance of multiple articulatory patterns producing the same acoustic form (Hoole, Nguyen-Trong, and Hardcastle, 1993). Direct kinematic measures with children are best suited for examination of lip and jaw movements because these articulators are the only ones visible without invasive techniques, which are both less likely to succeed with young children and more likely to spawn concern about risk to individual research participants.

The current study investigated the organization of articulatory gestures for the production of words with final stops that were either voiced or voiceless, and was an extension of Nittrouer (1993), in which acoustic analyses of children's and adults' stop-vowel sequences were reported. That study found that children's jaw movements had attained mature

[a)]Portions of this work presented at the 143rd meeting of the Acoustical Society of America, Pittsburgh, June, 2002.
[b)]email: nittrouer@cpd2.usu.edu
[c)]Affiliation: Boys Town National Research Hospital.

gestural form by 3 years of age, but lingual gestures remained highly constrained by the phonetic environment until at least the age of 7 years. The overarching conclusion of that study was that the path to mature patterns of gestural organization is not uniform. Rather, the rate of learning varies depending on the articulator to be used and the shape of the utterance to be produced.

The question of how children learn to make and coordinate the gestures required to produce words differing in the voicing of final stops is interesting partly because of what is known about children's *perception* of syllable-final stop voicing. The two most widely studied properties that convey information about syllable-final stop voicing are (1) the duration of the vocalic portion of the syllable and (2) the first-formant ($F1$) offset transition. Three studies found that children as old as 6 years of age weighted the $F1$ offset transition more and vocalic duration less than adults in making voicing decisions about syllable-final stops (Greenlee, 1980; Krause, 1982b; Wardrip-Fruin and Peach, 1984), although interpretation differed slightly across studies. In particular, Wardrip-Fruin and Peach suggested that the problem exhibited by children might best be described as difficulty in *integrating* the vocalic duration with the $F1$ cue, rather than as an age-related difference in the *weighting* of these cues. Minor differences in interpretation aside, these developmental perception results are interesting in light of cross-linguistic studies demonstrating that the integration and/or weighting of the cues to syllable-final voicing depends on native language experience. Speakers of languages without syllable-final stops (such as Mandarin and Japanese) or of languages that fail to show a vocalic-length difference depending on the voicing of the final stop (such as Arabic) fail to use vocalic duration as a cue in their phonetic decisions for English words ending with stops (Crowther and Mann, 1992; 1994), unless they receive intensive training (Flege, 1989). Furthermore, several studies have demonstrated that native speakers of languages without final stops, or without a vocalic-duration difference associated with the voicing of final stops, do not differentiate vocalic duration for voiceless and voiced final stops in their own productions of English words as much as native English speakers do (Crowther and Mann, 1994; Flege and Port, 1981). Such results support the hypothesis that native language experience with the various acoustic properties distinguishing voicing in word-final stops is required for speaker/listeners to be able to use and reproduce these properties in their own listening and speaking. Thus, the question arises of how children's productions of words containing syllable-final stops differing in voicing are organized, given that they either do not integrate or do not weight the primary cues responsible for distinguishing this voicing dimension as adults do.

One study investigated vocalic duration in children's speech production and reported that 3- and 6-year-olds demonstrated more of a difference in vocalic duration than adults depending on syllable final voicing (Krause, 1982a). Upon first consideration, this finding seems counterintuitive in light of the cross-linguistic studies: If speakers must learn how to vary vocalic duration according to syllable-final voicing, then why would children show a greater effect? Buder and Stoel-Gammon (2002) answered this question by suggesting that vocalic length may inherently vary with syllable final voicing, and so be present from the onset of speech. According to this position, speakers of languages that lack a vowel-length distinction as a function of syllable final voicing actually must learn not to make the distinction. To support this position, they present data showing that children learning Swedish, which has only an attenuated vocalic-length distinction associated with syllable-final voicing, show the effect in their productions at age 24 months, but not at 30 months. Although this result supports their contention, there is at least one reason for skepticism: Vocalic length of syllables ending with voiced consonants produced by Swedish-learning 24-month-olds was generally greater than 400 ms, which is much longer than typical syllables. Be that as it may, Buder and Stoel-Gammon did not report on other aspects of production, and so it is impossible to obtain a picture of gestural *organization* for these syllables. Neither did they provide estimates of variability in vocalic duration for individual children, and so it is impossible to determine how consistent children were in their productions. Krause did provide estimates of variability, and reported greater variability in vocalic duration for children's than for adults' samples. However, Krause did not examine spectral properties in children's speech samples, and so we cannot evaluate gestural organization for these word tokens. The current study was conducted in order to investigate this larger question of how the several articulatory gestures involved in producing words with final stops differing in voicing are organized.

Most acoustic analyses done on adults' samples of words with voiced and voiceless final stops have focused only on the duration of the preceding vocalic portion, generally defined as the sum of the voiced initial transitions, the steady-state region, and any voiced final transitions. These studies have found that this portion is shorter before voiceless than before voiced final stops in adults' speech (e.g., Chen, 1970; Crowther and Mann, 1992; 1994; Flege and Port, 1981; House and Fairbanks, 1953; Peterson and Lehiste, 1960). Other studies have investigated $F1$ frequency at voicing offset, showing that it is higher before voiceless rather than voiced stops (e.g., Crowther and Mann, 1992; 1994; Summers, 1987). Taken together, the vocalic duration and $F1$ differences indicate that speakers abduct the vocal folds before completing the closing gesture when the final stop is voiceless, but close the vocal tract before voicing ends when the final stop is voiced.

Methods other than acoustic analysis have been used to study adults' productions of words differing in the voicing of final stops, and have uncovered other differences in gestural organization for these words. Most kinematic studies have been done using words with final bilabial stops only, likely due to the visibility of these gestures. Using that approach, Gracco (1994) found that both peak jaw lowering and peak jaw velocity occur sooner in the production of /æp/ than of /æb/. Summers (1987) showed similar results for the jaw lowering gesture, and also demonstrated that the jaw achieves a more open position before voiceless, rather than voiced, stops. However, the jaw remained in a relatively stable position for longer periods of time preceding voiced stops. At the same time, Summers found no consistent difference for jaw raising gestures as a function of the voicing of the final stop. Of particular interest to the current acoustic analysis, Summers was able to correlate $F1$ change with jaw position and movement, but strongly so for medial syllabic portions only. For syllable margins, jaw movement and $F1$ frequency did not correlate as well. Specifically, Summers found that $F1$ at voicing offset was higher for voiceless than for voiced final stops, as expected, even though the jaw was in roughly the same position at voicing offset for both voicing conditions in samples from two of his three speakers. Nonetheless, $F1$ will be considered here as the best acoustic estimate of mandibular movement because of the well-documented fact that $F1$ is a function of the cross-sectional area of the front portion of the vocal tract (Schroeder, 1967; Stevens and House, 1955). The jaw and lips are the biggest determinants of this area, which probably explains Summers' results: The lips may have affected $F1$ frequency at voicing offset for two of his speakers, but he did not investigate lip movements. The lips (particularly the lower lip) contribute significantly to adults' closing gestures for bilabial stops (e.g., Green *et al.*, 2000), as studied by Summers, but they would not be expected to contribute much to closing for stops with lingual placements, as primarily studied here.

Raphael (1975) used physiological measures to examine adults' gestural organization for words with syllable-final voiced and voiceless consonants. He purposely avoided words in which the tongue was used for both vowel and final consonant constrictions. Instead, he examined EMG activity for the tongue (genioglossus) during vowel constrictions and for the lips (orbicularis oris and depressor anguli oris) during consonant constrictions in the production of words such as *lap* and *lab*. Results showed that EMG activity for the genioglossus muscle reached its peak at the same time, relative to the onset of voicing, for words with both voiced and voiceless final stops, but that the decay of that EMG activity was more rapid for words with voiceless final stops than for those with voiced final stops. EMG activity for muscles responsible for lip closure reached its peak at the same time relative to the decay of genioglossus EMG activity for words with both kinds of final stops. According to this gestural account, we might predict that there would be no difference in frequency of formants higher than $F1$ at voicing offset for words with voiced and voiceless final stops. However, this conclusion must be tempered mainly because Raphael's study did not involve words requiring tongue gestures for both vowel and final consonant constrictions. The present study examined words with alveolar and velar constrictions for final stops, so that the tongue was required for both vowel and consonant constrictions. The frequency of $F2$ at syllable center and at voicing offset was measured and was used to speculate as to whether there were differences in the coordination of lingual and vocal-fold gestures as a function of voicing or place of the final stop. In particular, we wanted to examine whether children's lingual gestures are similar to those of adults in the production of these words. Both MacNeilage and Davis (1991) and Nittrouer (1993) suggested that young children acquire mature patterns of vocal-tract opening and closing (accomplished primarily with the jaw) earlier than they acquire mature gestural patterns for other articulators. That suggestion was tested in this study by examining jaw gestures (through $F1$ measures) and tongue gestures (through $F2$ measures) in the production of words with voiced and voiceless final stops with several places of vowel and consonant constriction.

Another concern addressed by this study has to do with the organization of articulatory gestures for words when they are embedded in continuous speech. By far, most studies examining either the acoustic or the articulatory attributes of words with voiced and voiceless final stops have used tokens produced in isolation or in short, consistent carrier phrases. For adults, it may be that the articulatory score (and so, the acoustic consequences) differ when speakers produce words in real, meaningful contexts. For children, it may be that greater challenges are encountered in trying to organize gestures over longer, more involved utterances. For these reasons, speech samples were obtained in isolation and in sentences.

In summary, the current study sought to examine the gestural organization of monosyllabic words ending in voiceless or voiced final stops (particularly those with lingual constrictions) spoken by children and adults, both in isolation and in sentences, using acoustic measures. Measures made were the duration of the vocalic portion, and $F1$ and $F2$ frequencies at both voicing offset and syllabic center.

II. METHOD

A. Speakers

Eight speakers (four male and four female) in each of three age groups (adults, 7-year-olds, and 5-year-olds) participated. None of the speakers had ever been treated for a speech, language, or hearing problem. All speakers passed hearing screenings of the pure tones 0.5, 1.0, 2.0, 4.0, and 6.0 kHz presented at 25 dB HL. In addition, children were administered the Goldman–Fristoe Test of Articulation Sounds in Words Subtest (Goldman and Fristoe, 1986). All children, except one, were judged to produce all items in that subtest correctly. The one exception was a 5-year-old girl judged to be substituting /w/ for word-initial /ɹ/. Because none of the test items involved /ɹ/ production and this substitution is common for 5-year-olds, that was not considered to be a reason to exclude her.

B. Equipment and materials

All speech samples were recorded in a soundproof booth, directly onto the computer hard drive, via an AKG C535 EB microphone, a Shure M268 amplifier, and a Creative Labs Soundblaster 16-bit analog-to-digital converter. CSPEECHSP software (Milenkovic, 1997) was used both for recording speech samples and for the acoustic analyses. Speech samples were digitized at a sampling rate of 22.05 kHz using low-pass filtering with a high-frequency cutoff of 11.025 kHz.

Five minimal pairs made up the set of test items: *feet/feed*, *boot/booed*, *pick/pig*, *buck/bug*, and *cop/cob*. Two pairs had alveolar stop closures, two had velar closures, and one

TABLE I. Sample sentences for each word.

(1)	His *feet* were too big for his shoes.
(2)	Sam has to *feed* the dogs everyday after school.
(3)	Her new *boot* slipped off her foot.
(4)	The comedian was *booed* off the stage last night.
(5)	Susie likes to *pick* flowers on Sunday afternoons.
(6)	My grandfather has a *pig* on his farm.
(7)	Tim shot a *buck* on the first day of hunting season.
(8)	A lady *bug* is red and has black spots.
(9)	The man gave the *cop* the wallet he found in the street.
(10)	We had corn on the *cob* for dinner last night.

had bilabial closures. All words had initial consonants that permitted clear identification of voicing onset in the waveform. Vowels varied on the dimensions both of front/back and open/close. A set of pictures (5×5 in.) depicting each word was used to elicit word samples in isolation from children. A set of cards with the words printed on them served to elicit word samples in isolation from adults. Five sets of sentences were constructed for use, following the simple rules that each sentence should be about eight words long, the target word should not occur in phrase-final position, but the position of the word should differ across sets. A sample set of these sentences is presented in Table I. An adult female speaker with a midwestern dialect recorded these sentences onto tape.

C. Procedures

Hearing screenings were administered first. For children, the Goldman–Fristoe Test of Articulation was administered next. The order of collection of words in isolation and in sentences was alternated and randomized for each speaker, with the stipulation that a complete set of the ten words was collected in one condition before moving to the next. Five sets of words in each condition were collected. Words in isolation were elicited from children with picture cards and from adults with printed cards. Words in sentences were obtained by imitation of the sentences heard on tape. Two experimenters participated in the collection of these samples: one worked with the person speaking and one controlled the computer. As a result, samples in each condition were checked before moving to the next condition, and any samples judged to be unacceptable (due to extraneous noise or low amplitude) were immediately recorded again. In all, 100 samples were collected from each speaker: 2 conditions×10 words×5 samples.

D. Acoustic analysis

Each word was separated and saved to its own waveform file. Spectrograms of two productions of each word, in each context, were made so that general information about the acoustic form of each word would be available while making more fine-grained measurements. Figure 1 shows sample spectrograms of *buck* and *bug* spoken in isolation by an adult male. Several temporal and spectral measures were made on each speech token, following the procedures of Nittrouer (1993):

Duration of vocalic portion (*Dv*): time from the onset of voicing for the vocalic portion to either the offset of voicing (for words with voiceless final stops) or the point of vocal-tract closure (for words with voiced final stops). Arrows below the *x* axis in Fig. 1 show the ends of vocalic portions. These measures were obtained from the waveform. Cursors marking both the start and end of the vocalic portion were placed at zero crossings.

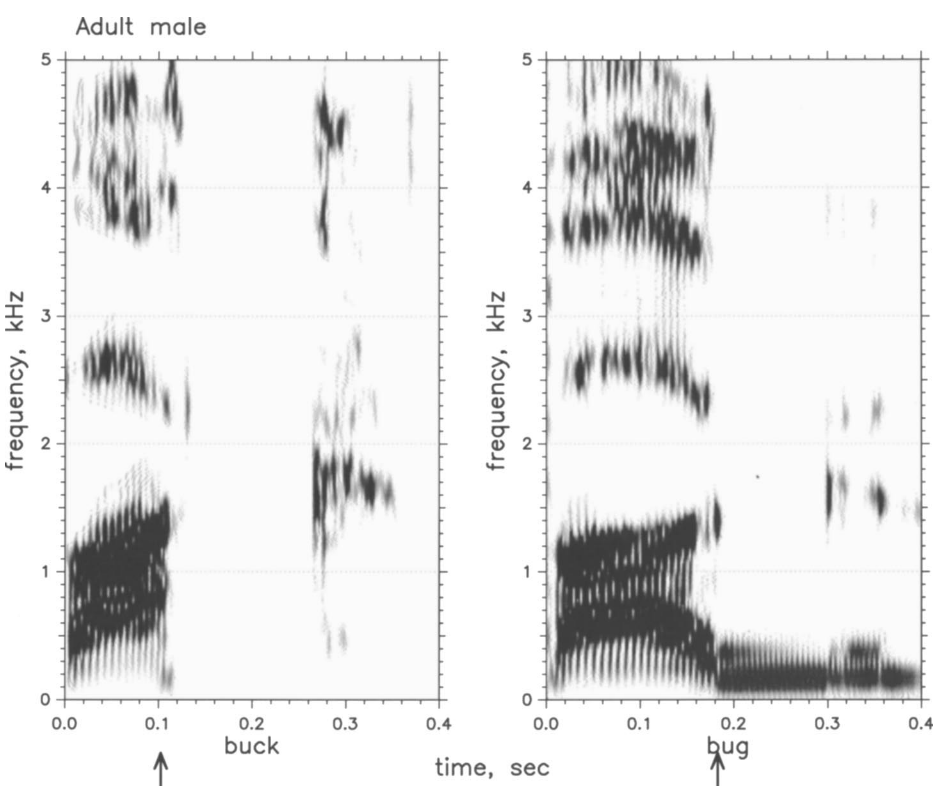

FIG. 1. Sample spectrograms of *buck* and *bug* spoken in isolation by an adult male. The arrows under the *x* axis indicate the points marked as *offset* in the acoustic analyses.

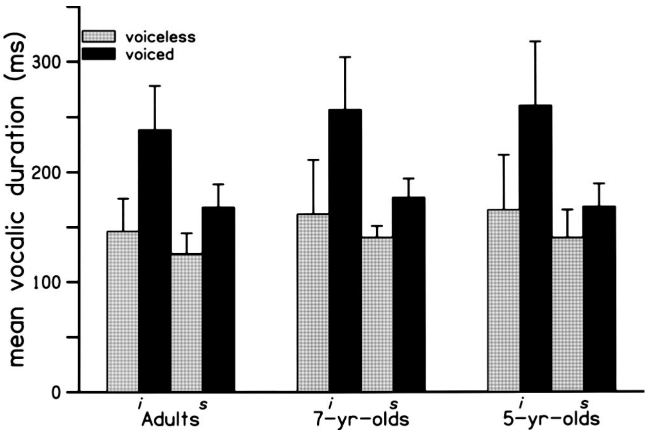

FIG. 2. Mean vocalic duration (Dv) (ms) for each age group, across all tokens of words with voiceless (*feet*, *pick*, *boot*, *buck*, and *cop*) and voiced (*feed*, *pig*, *booed*, *bug*, and *cob*) final stops, spoken in isolation and in sentences. Error bars represent standard deviations (SDs).

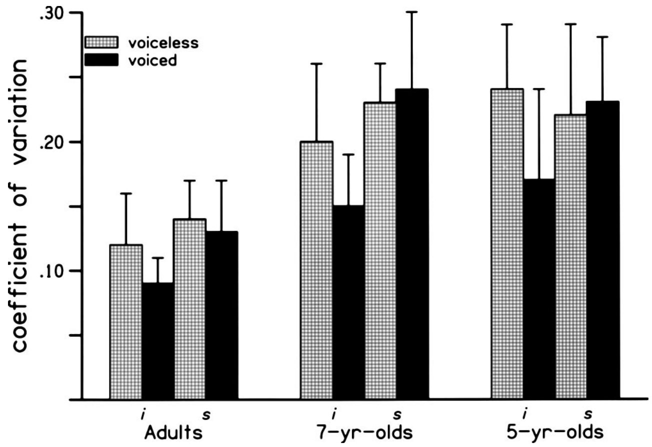

FIG. 3. Mean coefficients of variation (CVs) for Dv for each age group, across tokens of words with voiceless (*feet*, *pick*, *boot*, *buck*, and *cop*) and voiced (*feed*, *pig*, *booed*, *bug*, and *cob*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

F1 at voicing offset (F1off): $F1$ in the last pitch period. The onset and offset of the pitch period were marked at zero crossings. This measure, as well as all formant frequencies, were obtained using LPC analysis with 22 coefficients of the selected sample.

F2 at voicing offset (F2off): $F2$ for the last pitch period.

F1 at temporal center of the vocalic portion (F1center): $F1$ for three pitch periods at the temporal center of the vocalic portion.

F2 at temporal center of the vocalic portion (F2center): $F2$ for three pitch periods at the temporal center of the vocalic portion.

For each measure, a coefficient of variation (CV) was computed across tokens of each word spoken by each speaker by dividing the standard deviation (SD) by the mean. This provided an estimate of variability for individual speakers, and so allowed us to test the hypothesis that children are more variable in their productions than adults are.

Before measurements were made, laboratory staff and the first author discussed procedures for making these measurements, and practiced together to ensure that everyone was using the same procedures. Research assistants, including the second author, made all measurements. If there was a question about the locations of the onsets or offsets of voicing, research assistants consulted with each other or with the first author. The first or second author checked roughly 10 percent of all measurements to ensure that procedures were followed correctly. In all cases, they were.[1]

III. RESULTS

A. Vocalic duration

For the analysis Dv, results were collapsed across words with voiceless final stops (*feet*, *pick*, *cop*, *buck*, and *boot*) and across words with voiced final stops (*feed*, *pig*, *cob*, *bug*, and *booed*). Figure 2 shows mean Dv for words with voiceless and voiced final stops, spoken in isolation and in sentences. Overall, children and adults produced words with similar Dv's. As predicted, it appears that Dv's were shorter for words with voiceless, rather than voiced, final stops. Words with voiced final stops also appear to have had much shorter Dv's when produced in sentences rather than in isolation. Mean Dv's of words with voiceless final stops appear to have been somewhat shorter when produced in sentences instead of in isolation, but the magnitude of this difference is far less than that seen for words with voiced final stops. Across speakers there was a 75-ms difference for words with voiced final stops that were produced in isolation versus in sentences, and a 27-ms difference for words with voiceless final stops produced in the two contexts.

A three-way analysis of variance (ANOVA) with age as the between-subjects factor and context (isolation or sentence) and voicing as the within-subjects factors supported these observations. The main effect of age was not significant, but the main effects of both context and voicing were: for context, $F(1,21)=26.86$, $p<0.001$; for voicing, $F(1,21)=286.32$, $p<0.001$.[2] The only interaction that was significant was context×voicing, $F(1,21)=53.36$, $p<0.001$, supporting the observation that Dv's for words with voiced final stops differed across contexts more than the Dv's of words with voiceless final stops. Overall, it can be concluded that speakers of all ages produced words with similar Dv's, that these Dv's were shorter for words with voiceless rather than voiced final stops, and that when words with voiced final stops were spoken in sentences, Dv was substantially shorter than when words were spoken in isolation.

Figure 3 displays mean CVs for Dv, and shows a complicated pattern of results. In general, variability appears greater for the children's groups than for adults: that is, CVs are larger for children. For words spoken in isolation, it appears that speakers of all ages showed greater variability in Dv for words with voiceless, rather than voiced, final stops. For words spoken in sentences, variability is similar for words with voiceless and voiced final stops, although the relation between the means for the two conditions varies across speaker age. For adults, mean CV for words with voiceless final stops (spoken in sentences) is slightly greater than for words with voiced final stops. For 7- and 5-year-olds, mean CVs for words with voiced final stops (spoken in sentences) are slightly greater than for words with voiceless final stops. Results of a three-way ANOVA provide help interpreting the patterns seen in Fig. 3. The main effect of age was significant, $F(2,21)=22.33$, $p<0.001$, as well as the main effect of voicing, $F(1,21)=11.15$, $p=0.003$. The first of these results is readily apparent from Fig. 3; the second result is harder to discern. Across speakers and contexts, mean CV was 0.19 for words with voiceless final stops and 0.17 for words with voiced final stops. The main effect of context was also significant, $F(1,21)=17.89$, $p<0.001$, reflecting the fact that across speakers and words ending in voiceless and voiced final stops, mean CV was 0.20 for words spoken in sentences and 0.16 for words spoken in isolation. Finally, the two-way interaction of context $\times$voicing was significant, $F(2,21)=8.23$, $p=0.01$. Judging from Fig. 3, it appears that this interaction can be accounted for by the patterns of variability exhibited by 7- and 5-year-olds and described above: When words were spoken in isolation, variability was greater for words with voiceless final stops; when words were spoken in sentences, variability was greater for words with voiced final stops.

A simple effects analysis, holding age constant, was also performed on these CVs to provide further insight into these results. For adults, no significant main effects or interactions were found, but the main effect of voicing was close to significant, $F(1,21)=3.56$, $p=0.073$, as well as the main effect of context, $F(1,21)=3.24$, $p=0.086$. For 7-year-olds, the effect of context was found to be significant, $F(1,21)=15.96$, $p<0.001$. This result indicates that 7-year-olds were more variable for words spoken in sentences than for those spoken in isolation. In addition, the context$\times$voicing interaction was close to significant for 7-year-olds, $F(1,21)=3.03$, $p=0.096$. For 5-year-olds, the main effect of voicing was significant, $F(1,21)=7.23$, $p=0.014$, as well as the context$\times$voicing interaction, $F(1,21)=6.32$, $p=0.02$. This finding supports the impression that variability was slightly greater for words with voiced, rather than voiceless, final stops when words were spoken in sentences, but when words were spoken in isolation, 5-year-olds showed decreased variability for words with voiced, rather than voiceless, final stops. Consequently, several conclusions may be drawn: (1) Children were more variable than adults; (2) All speakers were more variable in their productions of words with voiceless, rather than voiced, final stops; (3) This pattern of greater variability for words with voiceless, rather than voiced, final stops is mostly restricted to words spoken in isolation; and (4) 7-year-olds were more variable for words spoken in sentences than for those spoken in isolation.

B. F1off

Two word pairs (i.e., *feet/feed* and *boot/booed*) were not included in the analyses of *F1off* because the vocal tract remains in a fairly closed position throughout the production of these words, and so there is very little change in $F1$ across the word. For *cop* and *cob*, it was difficult to separate $F1$ from $F2$ in many of the children's samples. As a result, *F1off* was analyzed for *pick, pig, buck,* and *bug* only. Figure 4 displays mean *F1off* for words with voiced and voiceless final stops, for male and female speakers separately. As expected, *F1off* was higher for speakers in all three age groups

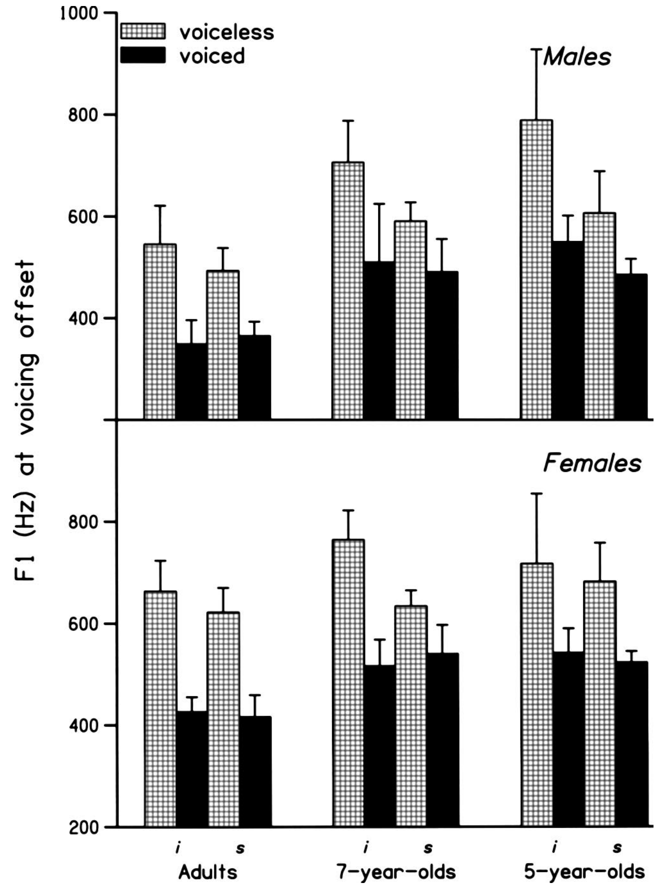

FIG. 4. Mean *F1off* (Hz) for males and females in each age group, for words with voiceless (*pick* and *buck*) and voiced (*pig* and *bug*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

for words ending in voiceless stops than for those ending in voiced stops. In addition, *F1off* specifically for words with voiceless final stops was lower when produced in sentence contexts rather than in isolation. This trend is particularly apparent for children's samples.

A four-way ANOVA, with age and sex as the between-subjects factor and context and voicing as the within-subjects factors, was done on these *F1off* frequencies. Speaker sex was not a variable that was of particular interest in this study, but any time spectral measures are examined it seems prudent to include sex as a factor. All four main effects were significant: age, $F(2,18)=14.92$, $p<0.001$; sex, $F(1,18)=5.21$, $p=0.035$; context, $F(1,18)=32.83$, $p<0.001$; and voicing, $F(1,18)=139.82$, $p<0.001$. The only interaction term that was significant was context$\times$voicing, $F(1,18)=27.26$, $p<0.001$. The main effects of age and sex simply reflect the fact that speakers with longer vocal tracts (i.e., adults and males) have lower $F1$ frequencies, and the main effect of voicing reflects the fact that *F1off* is lower for words with final voiced, rather than for voiceless, stops. The findings of a significant main effect of context and of a significant interaction for context$\times$voicing reflect the same trend: *F1off* is lower for words spoken in sentences than in isolation, and this trend is almost entirely due to *F1off* for words specifically with voiceless final stops being lower when produced in sentences, rather than in isolation. For words with voiced final stops, *F1off* is similar across the two

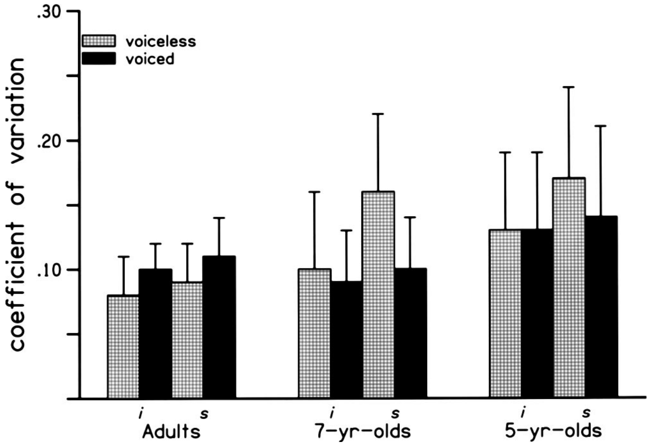

FIG. 5. Mean CVs for *F1off* for each age group, across tokens of words with voiceless (*pick* and *buck*) and voiced (*pig* and *bug*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

contexts. The fact that a significant three-way interaction of context×voicing×age was not found, as might be expected given that this trend seems particularly pronounced for children, is probably due to the fact that all groups showed the effect to some degree. Children simply demonstrated it to a greater extent. To test this suggestion, a simple effects analysis was run on *F1off*, looking at the effects of context and voicing for each age group separately. The context×voicing interaction did not reach statistical significance for adults, but did for both children's groups: 7-year-olds, $F(1,18) = 21.93$, $p < 0.001$; and 5-year-olds, $F(1,18) = 6.37$, $p = 0.021$. Children in both age groups had lower *F1off* frequencies for words with voiceless final stops when those words were spoken in sentences rather than in isolation. In articulatory terms, this result indicates that children's vocal tracts were closed more at voicing offset for these words when they were produced in sentences. This trend must have resulted from one of three possible reasons: (1) children never opened their mouths as much during the production of these words in sentence contexts as in isolation; (2) children began closing their vocal tracts sooner during word production in sentence contexts, compared to isolation; or (3) some combination of these two factors. The analysis of *F1center* will help resolve the issue.

Coefficients of variation were computed for *F1off* of each word used in the analysis above: *pick* and *buck* for the voiceless condition, and *pig* and *bug* for the voiced condition. The amount of variation in *F1off* serves as a metric of how consistently individual speakers coordinated vocal-fold abduction with jaw gestures in the production of these words. Mean CVs are shown in Fig. 5. There is a clear developmental decrease in the amount of variation associated with these measures. This age-related difference appears across voicing and contexts for 5-year-olds versus adults, but for 7-year-olds, CVs for *F1off* do not appear higher than those of adults except for words ending in voiceless stops spoken in sentences. In fact, both 7- and 5-year-olds displayed increased variability for words with voiceless final stops in sentences, compared to the other three conditions.

A three-way ANOVA was performed on CVs for *F1off*, with age as the between-subjects factor and voicing and context as within-subjects factors. Sex was not a factor here because there is no reason to predict that variability would differ for male and female speakers. The main effect of age was significant, $F(2,21) = 5.25$, $p = 0.014$, and the main effect of context approached significance, $F(1,21) = 4.16$, $p = 0.054$. These main effects indicate that children were more variable than adults in how they coordinated vocal-fold abduction with jaw gestures in the production of words with voiceless and voiced final stops, and that overall speakers were more variable when words were produced in sentences. But, this last effect is probably explained in large part by the increased variability found just for children's productions of words with voiceless final stops in sentences. To examine this suggestion, a simple effects analysis was done on these CVs for each age group separately. The term of interest was the context×voicing interaction. This interaction term was not significant for adults, indicating that adults showed similar differences in variability of *F1off* across voicing conditions for words spoken in isolation and in sentences. However, the context×voicing interaction was significant for 7-year-olds, $F(1,18) = 21.93$, $p < 0.001$, as well as for 5-year-olds, $F(1,18) = 6.37$, $p = 0.021$. These results support the suggestion that children were particularly variable in their attainment of *F1off* for words with voiceless final stops spoken in sentences.

C. F1center

This spectral measure effectively examined whether there was an age-related difference in the degree of jaw opening at the middle of the syllable, depending on the voicing of the final stop. Summers' (1987) finding that maximum excursion was greater for words with voiceless, rather than voiced, final stops led to the prediction that, at least for adults, *F1center* would be higher for words with voiceless, rather than voiced, final stops. We wanted to see if the same effect would be found for children. As with the analysis of *F1off*, this examination was performed only on measures from the *pick/pig* and *buck/bug* minimal pairs. Figure 6 shows mean *F1center* for male and female speakers in each group.

A four-way ANOVA was performed on these measures with age and sex as the between-subjects factors and context and voicing as the within-subjects factors. All four main effects were found to be statistically significant: age, $F(2,18) = 8.36$, $p = 0.003$; sex, $F(1,18) = 5.13$, $p = 0.036$; context, $F(1,18) = 11.06$, $p = 0.004$; and voicing, $F(1,18) = 21.12$, $p < 0.001$. In addition, the context×voicing interaction was significant, $F(1,18) = 11.54$, $p = 0.003$, suggesting that the difference in *F1center* across voicing conditions may have been attenuated for words spoken in sentences instead of in isolation. To examine whether children actually showed the difference in *F1center* predicted by Summers (1987), a simple effects analysis was done, holding age constant. All three age groups showed a significant effect of voicing: adults, $F(1,18) = 5.91$, $p = 0.026$; 7-year-olds, $F(1,18) = 5.07$, $p = 0.037$; and 5-year-olds, $F(1,18) = 10.74$, $p = 0.004$. It can thus be concluded that adults and children alike showed the difference in *F1center* predicted by Summers, at least in general. The simple effects analysis was also

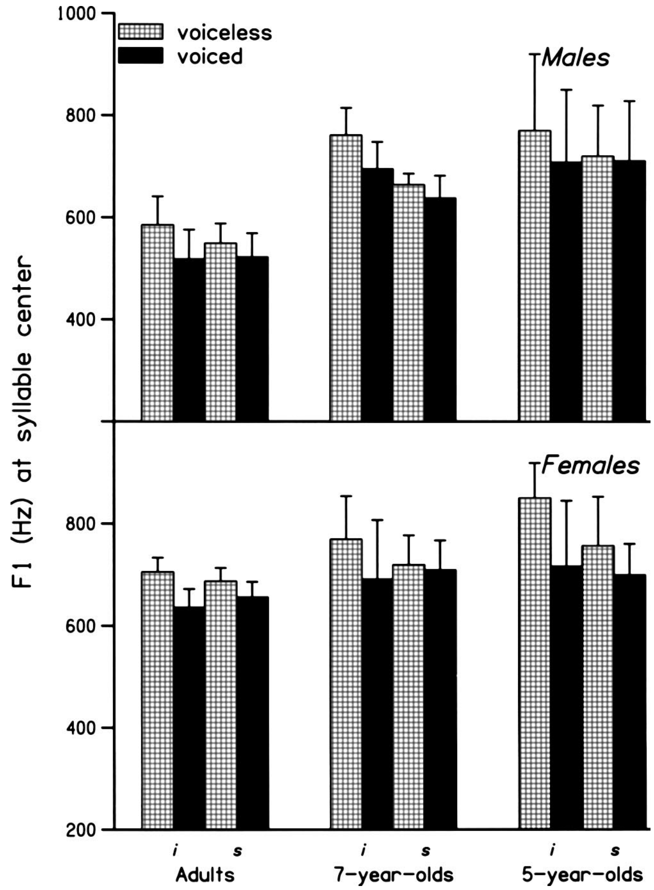

FIG. 6. Mean *F1center* (Hz) for males and females in each age group, for words with voiceless (*pick* and *buck*) and voiced (*pig* and *bug*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

used to help determine whether this voicing-related difference was attenuated more for children than for adults when words were spoken in sentences rather than in isolation. To this end, the context×voicing interaction was examined for each age group separately. For adults, this interaction was not significant, and so we may conclude that adults showed the voicing-related difference in *F1center* to the same extent for words spoken in isolation and in sentences. However, the interaction was significant, or at least close to it, for both children's groups: 7-year-olds, $F(1,18)=3.99$, $p=0.061$; 5-year-olds, $F(1,18)=5.87$, $p=0.026$. Thus, there was an age-related difference in speakers' abilities to preserve the increase in *F1center* for words with voiceless final stops when producing words in sentences: The older the speaker, the better this increase was preserved. In articulatory terms this finding indicates that children do not completely maintain the difference in jaw excursions for words with voiced and voiceless final stops described by Summers when words are produced in sentences. This trend may explain, at least in part, the finding that *F1off* is lower when words with voiceless final stops are produced in sentences rather than in isolation for children's samples, but not for adults' samples. It may be that children never open the vocal tract as much as adults do for words with voiceless final stops produced in sentences. This issue is examined more closely in subsection D below.

As was done with *F1off*, CVs were computed for

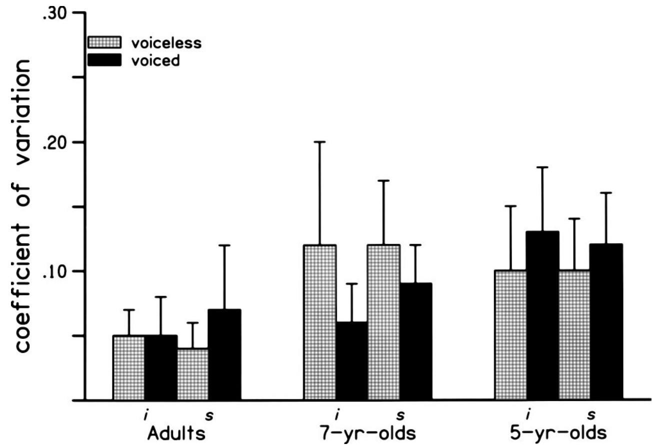

FIG. 7. Mean CVs for *F1center* for each age group, across tokens of words with voiceless (*pick* and *buck*) and voiced (*pig* and *bug*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

F1center. Figure 7 displays mean CVs for *F1center* and reveals several interesting findings when compared to Fig. 5 (which showed CVs for *F1off*). For adults, variability in *F1center* across all conditions is roughly only half as large as what it was for *F1off*. For both children's groups, there is a reduction in variability for *F1center* compared to *F1off*, but this reduction is not as great as for adults. For 7-year-olds, the enhanced variability observed for *F1off* for words with voiceless final stops produced in sentences is seen here for *F1center* as well, but is also observed for these same words produced in isolation. These 7-year-olds clearly demonstrated variability in jaw gestures during the production of words with voiceless stops.

A three-way ANOVA was performed on these CV measures, with age as the between-subjects factor and voicing and context as within-subjects factors. The main effect of age was significant, $F(2,18)=9.05$, $p=0.001$, as well as the age ×voicing interaction, $F(2,21)=6.36$, $p=0.007$. This interaction is likely due to the finding that 7-year-olds showed greater variability when producing words with voiceless, rather than voiced, final stops.

D. Age-related differences in vocal-tract closing gestures for words with voiceless final stops

An analysis was undertaken to examine why *F1off* was lower for children's samples of words with voiceless final stops spoken in sentences, rather than in isolation. At issue was whether this trend could be entirely attributed to the finding that children seemed to be constrained in how much they opened their vocal tracts for words with voiceless final stops when those words were spoken in sentences. To explore this question, the difference between $F1$ frequency at syllable center and at voicing offset was computed on the mean frequencies of words with voiceless final stops (i.e., *pick* and *buck*) for each speaker, for words produced in isolation and in sentences separately. The means for these difference scores are shown in Table II for each age group.

A simple effects analysis was performed on these difference scores to examine the effect of context, holding age constant. The effect of context was not significant for adults,

TABLE II. Mean differences, in Hz, between $F1$ at syllable center and at voicing offset of words with voiceless final stops, spoken in isolation and in sentences. Standard deviations are given in parentheses.

	Isolation	Sentences
Adults	41	60
	(39)	(34)
7-year-olds	29	79
	(96)	(48)
5-year-olds	20	102
	(63)	(40)

was close to significant for 7-year-olds, $F(1,21)=3.48$, $p=0.076$, and was significant for 5-year-olds, $F(1,21)=9.49$, $p=0.006$. Thus, it can be concluded that for children, but not for adults, the change in $F1$ frequency between syllable center and voicing offset was greater for words with voiceless final stops produced in sentences, rather than in isolation. In articulatory terms this means that children began closing the vocal tract before the offset of voicing for words with voiceless final stops produced in sentences. This pattern of articulatory organization can be seen in Fig. 8. Although this figure shows spectrograms of speech samples from just one child, these patterns are typical of what all children did when producing words in sentences. Clearly the $F1$ trajectory for *buck* is different from what is found in Fig. 1, showing spectrograms from an adult's samples. For the most part, this articulatory pattern was not found for children's word samples obtained in isolation, although a few children exhibited the pattern even for words in isolation (note the large SDs in Table II for children's samples in isolation). This articulatory pattern was not observed in adults' samples, regardless of whether samples were obtained in isolation or in sentences.

E. F2off

If the vocal folds are abducted earlier relative to the closing gesture for words with voiceless, rather than voiced, final stops, then $F2$ will differ at voicing offset for these voicing conditions, at least when the tongue forms the consonant closure. Furthermore, $F2$ will either rise or fall going into closure, depending on the locations of vocalic and consonantal constrictions. For these analyses, only words in which $F2$ was presumed to be rising at the end of the syllable (i.e., words with relatively low vocalic $F2$ and relatively high $F2$ near the consonantal constriction) were used. To do otherwise would have severely constrained the likelihood of obtaining a statistically significant difference in *F2off* as a function of final-stop voicing, if one truly exists: In some cases *F2off* would be lower for voiced than voiceless stops; in other cases it would be higher. Accordingly, only the minimal pairs *pick/pig*, *buck/bug*, and *boot/booed* were used in analyses of *F2off*. Figure 9 shows mean *F2off* frequencies, and supports the suggestion that *F2off* is higher for words ending with voiced stops.

A four-way ANOVA was performed on *F2off* with age and sex as the between-subjects factors, and context and voicing as the within-subjects factors. Both between-subjects main effects were found to be statistically significant, as expected: for age, $F(2,18)=29.69$, $p<0.001$; for sex, $F(1,18)=37.01$, $p<0.001$. The main effect of voicing was also significant, $F(1,18)=55.50$, $p<0.001$, indicating that *F2off* was higher preceding voiced rather than voiceless stops. This finding demonstrates that the vocal folds were abducted sooner relative to the lingual closing gesture for words with voiceless, rather than voiced, final stops. The main effect of context was not significant, nor were any of the interactions.

As was done for other measures, CVs were computed for *F2off*, and means are shown in Fig. 10. Both groups of children appear to have been more variable than adults in their attainment of *F2off*. Unlike measures of variability for *Dv* and $F1$ (at both voicing offset and syllable center), 7-year-olds seem to have been as variable as 5-year-olds for *F2off* across voicing conditions and contexts.

A three-way ANOVA with age as the between-subjects

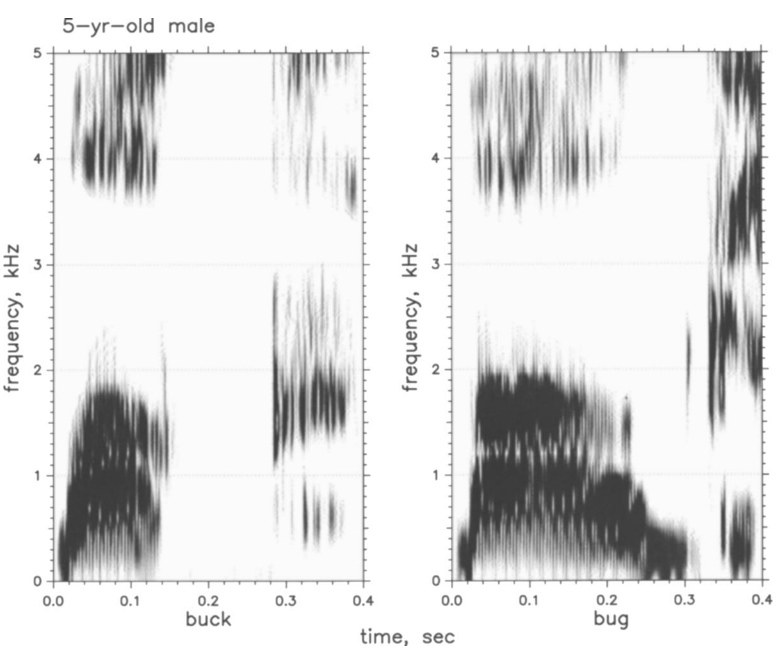

FIG. 8. Sample spectrograms of *buck* and *bug* spoken by a 5-year-old boy, showing that vocal-tract closing can be seen for *buck*.

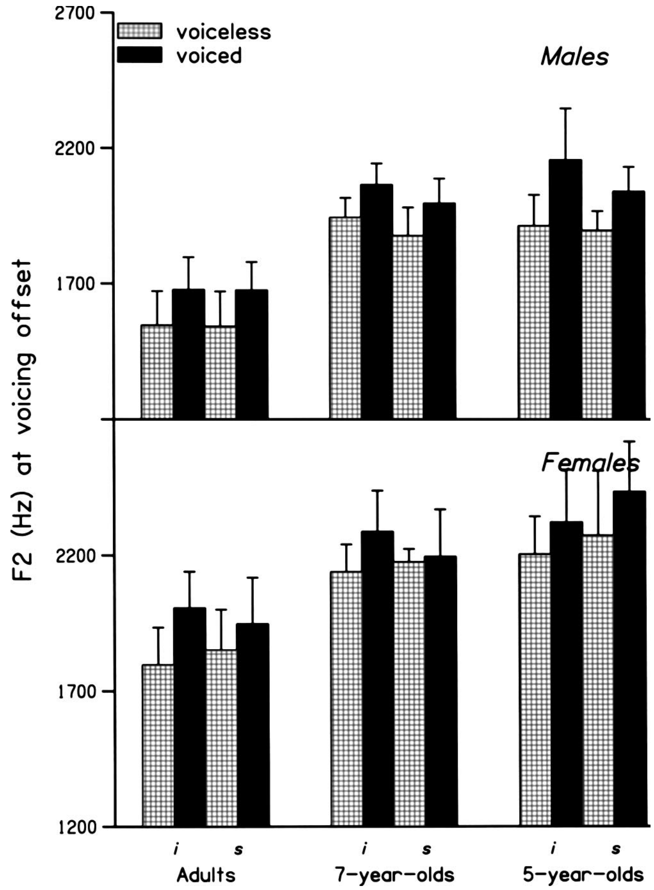

FIG. 9. Mean *F2off* (Hz) for males and females in each age group, for words with voiceless (*pick*, *buck*, and *boot*) and voiced (*pig*, *bug*, and *booed*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

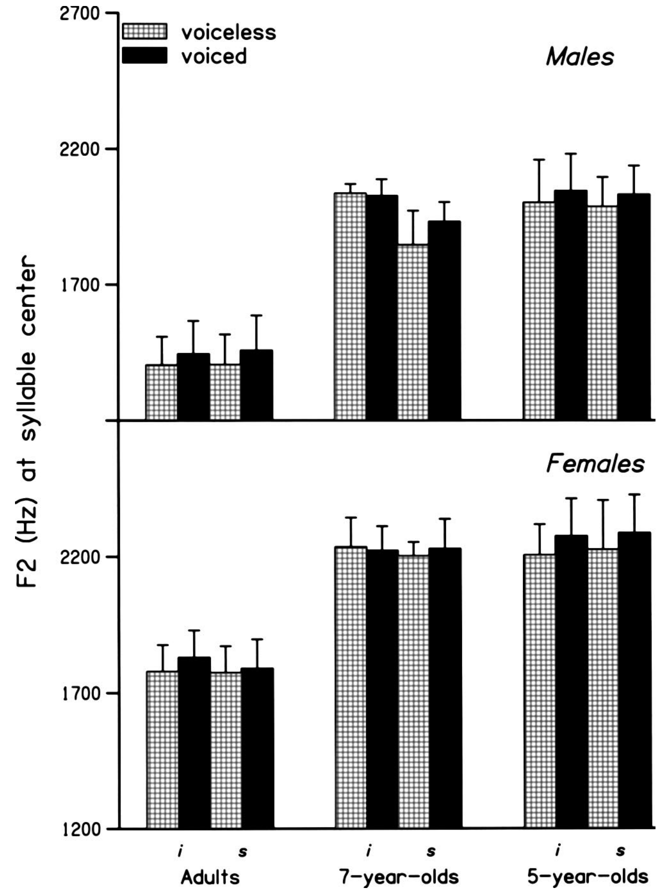

FIG. 11. Mean *F2center* (Hz) for males and females in each age group, for words with voiceless (*pick*, *buck*, and *boot*) and voiced (*pig*, *bug*, and *booed*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

factor and context and voicing as the within-subjects factors was performed. Indeed, the main effect of age was found to be significant, $F(2,21)=11.50$, $p<0.001$, supporting the suggestion that children were more variable than adults. In addition, the main effect of context was significant, $F(1,21)=8.18$, $p=0.009$. This finding reflects the fact that variability for *F2off* was greater for words spoken in sentences rather than in isolation: Across speakers CV was 0.087 for words in sentences versus 0.075 for words in isolation. This result may not reflect a simple increase in variability for words produced in sentences, but instead may indicate that speakers adjusted the precise place of stop closure depending on what gesture was required for the next word in the sentence. The failure to find an age×context interaction indicates that this cross-word coarticulation was no greater in children's than in adults' samples. Similarly, Nittrouer (1993) reported that coarticulation across word boundaries was no greater in children's than in adults' samples.

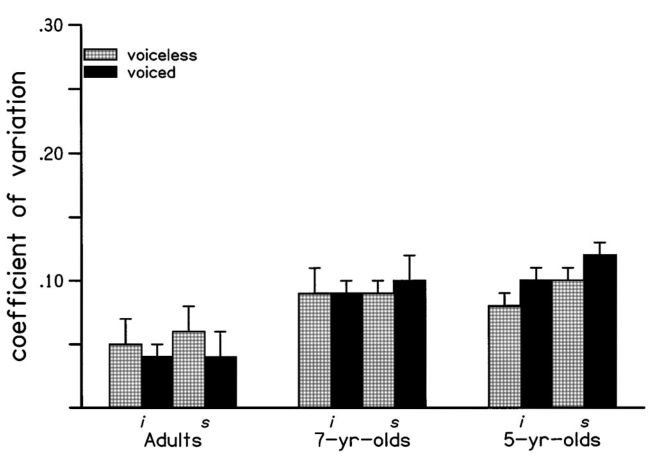

FIG. 10. Mean CVs for *F2off* for each age group, across tokens of words with voiceless (*pick*, *buck*, and *boot*) and voiced (*pig*, *bug*, and *booed*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

F. F2center

Figure 11 shows group means for *F2center*. It appears from this figure that the only factors that affected *F2center* were age and speaker sex. In particular, frequency of *F2center* seems to be largely unaffected by the voicing of the final stop. To evaluate these observations, a four-way ANOVA was performed. The main effects of age and sex were indeed significant: age, $F(2,18)=69.68$, $p<0.001$; and sex, $F(1,18)=37.01$, $p<0.001$. Contrary to impressions from Fig. 11, the effect of voicing was also significant, $F(1,18)=13.43$, $p=0.002$. However, the magnitude of this difference was quite small: across speakers, contexts, and words, *F2center* was just 40 Hz higher in words with voiced final stops than in those with voiceless final stops. This contrasts greatly with the 160-Hz difference observed for *F2off*.

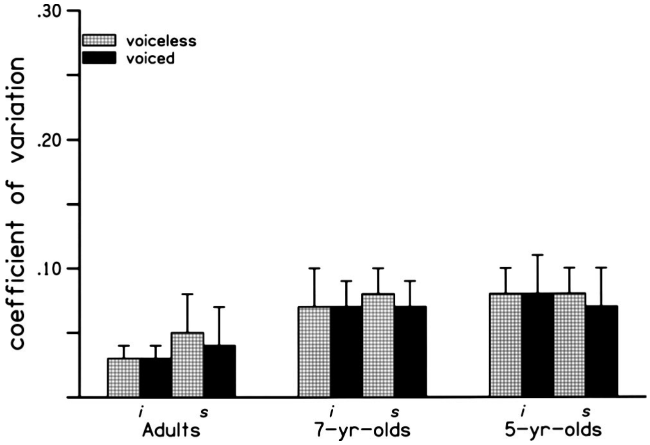

FIG. 12. Mean CVs for *F2center* for each age group, across tokens of words with voiceless (*pick*, *buck*, and *boot*) and voiced (*pig*, *bug*, and *booed*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

As done with all measures, CVs were computed for *F2center*. Figure 12 shows these values. Here, the only effect that appears to be significant is age, and a three-way ANOVA (age×context×voicing) confirmed that impression: Only the effect of age was significant, $F(1,21)=12.41$, $p<0.001$. As with *F2off*, 7- and 5-year-olds appear to have been similarly variable, but more variable than adults.

G. Tongue fronting for alveolar stops

One stated objective of this study was to examine whether there was evidence of a greater synergy between tongue gestures required for vowel and consonant production in children's than in adults' samples. Specifically, it was thought that young children might front the tongue more than adults in anticipation of upcoming alveolar stops. Precisely because this hypothesis was to be tested, a word pair that consisted of the back vowel /u/ and an alveolar stop was used: *boot* and *booed*. In the absence of anticipatory tongue fronting, we would expect *F2center* in *boot* and *booed* to be lower than in *buck* and *bug*. For example, Peterson and Barney (1952) found $F2$ in /u/ to be 300–400 Hz lower than $F2$ in /ʌ/ for men, women, and children. Table III shows mean *F2center* for *boot*, *buck*, *booed*, and *bug* spoken in isolation for all speaker groups in this study. Table IV shows *F2center* for these words spoken in sentences. These tables indicate that *F2center* was actually higher for *boot* and *booed* than for *buck* and *bug*, but not only for children's groups. Adults showed this effect as well. Presumably this was due to anticipatory tongue fronting, which was predicted, but no evidence of an age-related difference in the amount of this tongue fronting was found. This last result differs from predictions.

TABLE III. Mean $F2$ frequency (Hz) at syllable center across words spoken in isolation. Standard deviations are given in parentheses.

	Buck	Boot	Bug	Booed
Males				
Adults	1182	1245	1185	1235
	(68)	(186)	(98)	(223)
7-year-olds	1664	1834	1578	1824
	(164)	(213)	(77)	(94)
5-year-olds	1639	1635	1605	1667
	(180)	(296)	(98)	(303)
Females				
Adults	1491	1653	1532	1623
	(98)	(174)	(97)	(164)
7-year-olds	1794	2127	1823	2032
	(80)	(165)	(188)	(109)
5-year-olds	1833	1863	1819	2073
	(136)	(266)	(145)	(409)

TABLE IV. Mean $F2$ frequency (Hz) at syllable center across words spoken in sentences. Standard deviations are given in parentheses.

	Buck	Boot	Bug	Booed
Males				
Adults	1198	1296	1223	1279
	(70)	(197)	(85)	(291)
7-year-olds	1474	1711	1554	1634
	(82)	(219)	(123)	(104)
5-year-olds	1582	1642	1635	1667
	(87)	(278)	(56)	(287)
Females				
Adults	1426	1635	1478	1573
	(107)	(126)	(114)	(104)
7-year-olds	1740	2152	1829	2017
	(91)	(66)	(89)	(174)
5-year-olds	1849	2065	1889	2043
	(131)	(303)	(119)	(364)

IV. DISCUSSION

The goal of the study reported here was to use acoustic measures to examine the organization of articulatory gestures for words with voiceless and voiced final stops, produced by adults and children in isolation and in sentences. In general, it was found that children distinguished between words with voiceless and voiced final stops in the same ways that adults do. But, children were more variable in their productions, and children's organization of jaw and vocal-fold gestures for words with voiceless final stops produced in sentences differed from their organization of these gestures for words produced in isolation and from the organization of adults' gestures for these words (either in isolation or in sentences).

A. Dv

The durations of the vocalic word portions were similar across tokens produced by speakers of all ages, and these portions were shorter for words with voiceless final stops than for words with voiced final stops. For words with voiced final stops, there was a significant difference in *Dv* depending on whether the word was spoken in isolation or in a sentence. The fact that this particular context effect was attenuated for words with voiceless final stops suggests that there may be some limit to how short the vocalic portion of a word can realistically be: it simply takes time to open the vocal tract as needed for the vowel being produced. It may be that speakers operate near this effective limit when they produce monosyllables with voiceless final stops, regardless of whether they are produced in isolation or in sentences.

At the same time, variability in *Dv* was greater for words with voiceless, rather than voiced, final stops. This observation agrees with the notion that the coordination of gestures involved in producing words with voiceless final stops is likely more difficult than the coordination of gestures involved in producing words with voiced final stops. For words with voiceless final stops, the speaker must abduct the vocal folds at some point in the vocal-tract opening/closing gesture. Accomplishing this goal at precisely the same point across words surely requires great skill. On the other hand, obtaining consistency in *Dv*'s across tokens of words with voiced final stops would seem easier. Voicing offset for words with voiced final stops was defined here as the point at which the vocal tract reaches closure, and so *Dv* effectively measures only the duration of the opening/closing gesture. Nonetheless, 5-year-olds had more difficulty than adults achieving consistency in *Dv*, regardless of whether the goal was simply to produce equally timed opening and closing gestures (as in words with voiced final stops), or to abduct the vocal folds at a precise point in the opening/closing gesture (as in words with voiceless final stops). This enhanced difficulty was observed for 5-year-olds regardless of whether words were spoken in isolation or in sentences. Seven-year-olds, on the other hand, showed more consistency in the timing of these gestures when they were producing words in isolation. When they had the more difficult task of orchestrating the gestural score for the production of an entire sentence, however, they showed increased variability. Apparently children are fine-tuning their speech production skills past the age of 7 years.

It is interesting to compare adults' results for *Dv* across contexts with the cross-linguistic data. In this study adults demonstrated exactly a 100-ms difference in *Dv* for words with voiceless and voiced final stops spoken in isolation (using means across all voiceless and voiced tokens), which is what others have reported for English speakers (e.g., Chen, 1970; Crowther and Mann, 1994). For words spoken in sentences in this study, the voicing effect diminishes to 40 ms, which is similar to differences reported for non-native English speakers producing English words in isolation or in short, consistent carrier phrases (e.g., Crowther and Mann, 1992; 1994; Flege, Munroe, and Skelton, 1992; Flege and Port, 1981).[3] Presumably, in the natural, running speech that non-native speakers customarily hear, the difference in *Dv* for words with voiced and voiceless final stops is closer to the 40 ms measured in this study for words in sentences. Perhaps non-native English speakers imitate exactly what they hear. The question then becomes: Why do native English-speaking adults elongate *Dv* before voiced stops when producing words in isolation? That question is not answerable with these data.

The results reported here do not agree with those of Krause (1982a), who reported that young children showed a *greater* difference than adults in *Dv* for words spoken in isolation depending on the voicing of the final stop. In Krause's study the mean difference between voicing conditions was 97 ms for 6-year-olds and 60 ms for adults. The reason for this discrepancy in results, however, probably has more to do with Krause's findings for adults than with those for children. Adults' mean *Dv* for words with voiceless stops was 209 ms in Krause's study, considerably longer than the 157 ms obtained in the current study (for words spoken in isolation) or the 146 ms reported by Chen (1970). At the same time, Krause reported a mean *Dv* of 269 ms for words with voiced final stops spoken by adults, which is similar to the 251-ms mean found in the current study (for words spoken in isolation) and the 238-ms mean reported by Chen. Therefore, it is not that children's differences in *Dv* for words with voiceless and voiced final stops were unusually large in Krause' study, but rather that adults' differences were somewhat small. In particular, adults in that study produced words with voiceless final stops that had rather long *Dv*s.

B. F1off

The first thing to be said about $F1$ frequency at voicing offset is that its utility in aiding the listener make decisions about the voicing of syllable-final stops would appear to be limited because it does not vary as a function of voicing when the preceding vowel is close. Nonetheless, examination of *F1off* informs us about gestural organization in the production of consonant–vowel–stop sequences when we examine it for words with open vowels.

An age-related difference in gestural organization that emerges from the analysis of *F1off* is that children begin closing the vocal tract before the cessation of voicing for voiceless final stops—especially when trying to organize articulatory gestures over the length of a sentence. Adults, on the other hand, abduct the folds before they begin closing the vocal tract whether they are producing words with voiceless final stops in isolation or in sentences. While children are somewhat restricted in the extent to which they open their vocal tracts (i.e., lower their jaws) when producing words with voiceless final stops in sentences, rather than in isolation, this finding cannot completely account for the age-related difference in patterns of *F1off* across contexts. After reaching maximum jaw opening, adults apparently maintain stable jaw positions until they abduct their vocal folds, as Summers (1987) reported. Children begin to raise their jaws, and this effect is more pronounced for words spoken in sentences rather than in isolation. This finding mandates revision of the conclusion of Nittrouer (1993) that by 3 years of age children have acquired mature patterns of jaw movements. At least for some syllable shapes, it appears that children as old as 7 years have not completely mastered mature jaw patterns. Thus, this finding is one specific example of the suggestion that the emergence of mature gestural patterns is not uniform, that instead children attain mature patterns for some word forms sooner than for others (e.g., Green *et al.*, 2000; Nittrouer, 1993).

Regarding variability, *F1off* was less variable for all speakers than was *Dv*. The reason for this enhanced consistency may have to do with the fact that *Dv* can be influenced by several articulatory parameters, such as how rapidly or slowly the opening gesture is made, how long any steady-state vocalic portion is, and when vocal-fold abduction occurs. However, *F1off* is determined only by the degree of vocal-tract openness at the time the vocal folds are abducted. Apparently speakers tolerate some variability in the several parameters that affect *Dv*, but execute the vocal-fold abduction gesture at a relatively stable point in the opening/closing gesture across tokens of each word. At the same time, the finding of increased variability in *F1off* for children's productions of words with voiceless final stops indicates that their coordination of vocal-fold abduction and jaw gestures for these words was particularly unstable.

C. F1center

The results of this study replicate those of Summers (1987): speakers open the vocal tract more for words with voiceless, rather than voiced, final stops, and so *F1center* is higher for words with voiceless final stops. Although this gestural pattern was attenuated somewhat for children producing words in sentences, it was nonetheless found. Regarding variability, it was found that adults more consistently achieved the same *F1center* than they achieved the same *F1off*. This finding indicates that adults were more consistent in how they organized and produced vocal-tract gestures over the first part of the word than in how they did so over the latter portion of the word. Children did not always show this consistency in gestural organization for early word portions. In particular, 7-year-olds showed greater variability for *F1center* of words with voiceless, rather than voiced, final stops.

It is interesting that the context-related changes observed for the temporal (*Dv*) and spectral (*F1off* and *F1center*) measures are uncorrelated. That is, *Dv* was shorter for words spoken in sentences rather than in isolation, but primarily for words with voiced final stops. On the other hand, *F1off* and *F1center* differed across contexts for children's samples, but only for words with voiceless final stops. Thus, the changes in *Dv* associated with context did not affect spectral measures. In articulatory terms this means that changes in vocalic duration *per se* did not influence the articulatory gestures themselves, or the organization of these gestures.

D. F2off

Results for *F2off* reveal that $F2$ frequency at voicing offset can provide information regarding the voicing of syllable-final consonants. For the words analyzed here, with rising $F2$ at syllable offset, *F2off* was generally higher for voiced than for voiceless final stops. Although not reported, a similar trend would be expected and was observed in casual inspection of words with falling $F2$ at syllable offset. For example, across all speakers *F2off* was 2687 Hz for *feet* spoken in isolation and 2461 Hz for *feed* spoken in isolation. Variability in children's samples was similar for *F2off* and *F1off* (except that 7-year-olds showed variability comparable to that of 5-year-olds in all conditions for *F2off*), but adults demonstrated decreased variability for *F2off* compared to *F1off*. Particularly for adults' speech, *F2off* appears to provide very reliable information about the voicing of the final stop.

E. F2center

One predicted finding that was not observed had to do with tongue gestures. It had been predicted that children might demonstrate greater synergy between tongue gestures required for vowel and consonant production than adults. This could have shown up as greater tongue fronting in anticipation of the alveolar stops in *boot* and *booed* in children's than in adults' samples. If present, the acoustic consequence of this gestural pattern would have been a greater difference between *F2center* for *buck/bug* and *F2center* for *boot/booed* in children's than in adults' samples. However, the degree of tongue fronting was similar for adults and children. At the same time, children were more variable in their attainment of *F2center* than adults were.

As was found with $F1$ measures, there was no effect of shortened vocalic segments for words with voiced final stops spoken in sentences, rather than in isolation, on F2 measures. Neither *F2off* nor *F2center* showed a significant context effect.

In summary, several conclusions can be drawn from these analyses. First, children as old as 7 years of age still organize their gestures for the productions of words differently from adults for some syllable shapes. Second, children are generally more variable in their execution of linguistic gestures than are adults. Overall, learning to coordinate the various gestures involved in producing speech with appropriately timed events is a difficult task that extends well into childhood. Third, the acoustic correlates of syllable-final voicing are attenuated somewhat when words are produced in sentences, rather than in isolation. In general, this finding serves as a reminder that we must be careful about generalizing results obtained for speech samples produced in isolation to our understanding of speech produced in natural contexts. A final conclusion to emerge from these data is that the acoustic correlates of speech production are spread throughout the word. This finding highlights a fact long understood by speech scientists, and yet frequently overlooked in applications to technology and clinical work: there are no discrete acoustic segments that correspond to linguistic units in the speech stream.

ACKNOWLEDGMENTS

This work was supported by research Grant No. R01 DC00633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health, to Susan Nittrouer. The contributions of Lesley Larive and Tammy Mertes to the collection and analysis of these stimuli are gratefully acknowledged.

[1]Because the methods of acoustic analysis were so regimented there could be no computation of inter-rater reliability. As long as each experimenter adhered strictly to the outlined procedures the measurements obtained across experimenters were identical. All indications were that experimenters adhered to those procedures. Thus, inter-rater reliability was effectively 1.0.

[2]Although the alpha level of 0.05 is typically set, many investigators recognize the potential interest of "marginally" significant statistical tests (i.e., those with p values slightly above 0.05). For that reason, all tests with resulting p values of less than 0.10 will be reported throughout this paper. If an exact F- or t ratio is not given, it can be assumed that the value had an associated p of greater than 0.10.

[3]One study (Flege and Port, 1981) did report voicing differences for *Dv* in samples from English-speaking adults of just 40 ms. Interestingly, the carrier phrase in that study ("I say______again to Bob") was slightly longer than others, which are commonly just three words (e.g., "Say_______again.")

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.

Chen, M. (**1970**). ‘‘Vowel length variation as a function of the voicing of the consonant environment,’’ Phonetica **22**, 129–159.

Crowther, C. S., and Mann, V. A. (**1992**). ‘‘Native language factors affecting use of vocalic cues to final consonant voicing in English,’’ J. Acoust. Soc. Am. **92**, 711–722.

Crowther, C. S., and Mann, V. A. (**1994**). ‘‘Use of vocalic cues to consonant voicing and native language background: The influence of experimental design,’’ Percept. Psychophys. **55**, 513–525.

Davis, B. L., and MacNeilage, P. F. (**1990**). ‘‘Acquisition of correct vowel production: A quantitative case study,’’ J. Speech Hear. Res. **33**, 16–27.

Donahue, M. (**1986**). ‘‘Phonological constraints on the emergence of two-word utterances,’’ J. Child Lang **13**, 209–218.

Ferguson, C. A., and Farwell, C. B. (**1975**). ‘‘Words and sounds in early language acquisition,’’ Language **51**, 419–439.

Flege, J. E. (**1989**). ‘‘Chinese subjects’ perception of the word-final English /t/–/d/ contrast: Performance before and after training,’’ J. Acoust. Soc. Am. **86**, 1684–1697.

Flege, J., and Port, R. (**1981**). ‘‘Cross-language phonetic interference: Arabic to English,’’ Lang Speech **24**, 125–146.

Flege, J. E., Munro, M. J., and Skelton, L. (**1992**). ‘‘Production of the word-final English /t/–/d/ contrast by native speakers of English, Mandarin, and Spanish,’’ J. Acoust. Soc. Am. **92**, 128–143.

Goldman, R., and Fristoe, M. (**1986**). *Goldman–Fristoe Test of Articulation* (American Guidance Service, Circle Pines, MN).

Goodell, E. W. (**1991**). ‘‘Gestural organization in the speech of 22- to 32-month-old children,’’ Ph.D. dissertation, University of Connecticut.

Gracco, V. L. (**1994**). ‘‘Some organizational characteristics of speech movement control,’’ J. Speech Hear. Res. **37**, 4–27.

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.

Greenlee, M. (**1980**). ‘‘Learning the phonetic cues to the voiced–voiceless distinction: A comparison of child and adult speech perception,’’ J. Child Lang **7**, 459–468.

Hoole, P., Nguyen-Trong, N., and Hardcastle, W. (**1993**). ‘‘A comparative investigation of coarticulation in fricatives: Electropalatographic, electromagnetic, and acoustic data,’’ Lang Speech **36**, 235–260.

House, A. S., and Fairbanks, G. (**1953**). ‘‘The influence of consonant environment upon the secondary acoustical characteristics of vowels,’’ J. Acoust. Soc. Am. **25**, 105–113.

Katz, W. F., Kripke, C., and Tallal, P. (**1991**). ‘‘Anticipatory coarticulation in the speech of adults and young children: Acoustic, perceptual, and video data,’’ J. Speech Hear. Res. **34**, 1222–1232.

Krause, S. E. (**1982a**). ‘‘Developmental use of vowel duration as a cue to postvocalic stop consonant voicing,’’ J. Speech Hear. Res. **25**, 388–393.

Krause, S. E. (**1982b**). ‘‘Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults,’’ J. Acoust. Soc. Am. **71**, 990–995.

MacNeilage, P. F., and Davis, B. L. (**1991**). ‘‘Acquisition of speech production: Frames, then content,’’ in *Attention and Performance XIII. Motor Representation and Control*, edited by M. Jeannerod (Erlbaum, New York).

Menn, L. (**1978**). ‘‘Phonological units in beginning speech,’’ in *Syllables and Segments*, edited by A. Bell and J. B. Hooper (North-Holland, New York).

Milenkovic, P. (**1997**). CSPEECHSP (University of Wisconsin, Madison).

Nittrouer, S. (**1993**). ‘‘The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study,’’ J. Speech Hear. Res. **36**, 959–972.

Oller, D. K. (**1980**). ‘‘The emergence of the sounds of speech in infancy,’’ in *Child Phonology: Production*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Academic, New York).

Peterson, G. E., and Barney, H. L. (**1952**). ‘‘Control methods used in a study of the vowels,’’ J. Acoust. Soc. Am. **24**, 175–184.

Peterson, G. E., and Lehiste, I. (**1960**). ‘‘Duration of syllable nuclei in English,’’ J. Acoust. Soc. Am. **32**, 693–703.

Piske, T. (**1997**). ‘‘Phonological organization in early speech production: Evidence for the importance of articulatory patterns,’’ Speech Commun. **22**, 279–295.

Raphael, L. J. (**1975**). ‘‘The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English,’’ J. Phonetics **3**, 25–33.

Schroeder, M. R. (**1967**). ‘‘Determination of the geometry of the human vocal tract by acoustic measurements,’’ J. Acoust. Soc. Am. **41**, 1002–1010.

Smith, A., and Goffman, L. (**1998**). ‘‘Stability and patterning of speech movement sequences in children and adults,’’ J. Speech Lang. Hear. Res. **41**, 18–30.

Stevens, K. N., and House, A. S. (**1955**). ‘‘Development of a quantitative description of vowel articulation,’’ J. Acoust. Soc. Am. **27**, 484–493.

Summers, W. V. (**1987**). ‘‘Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses,’’ J. Acoust. Soc. Am. **82**, 847–863.

Sussman, H. M., Duder, C., Dalston, E., and Cacciatore, A. (**1999**). ‘‘An acoustic analysis of the development of CV coarticulation: A case study,’’ J. Speech Lang. Hear. Res. **42**, 1080–1096.

Tuller, B., and Kelso, J. A. S. (**1984**). ‘‘The timing of articulatory gestures: Evidence for relational invariants,’’ J. Acoust. Soc. Am. **76**, 1030–1036.

Wardrip-Fruin, C., and Peach, S. (**1984**). ‘‘Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants,’’ Lang Speech **27**, 367–379.

Waterson, N. (**1971**). ‘‘Child phonology: A prosodic view,’’ J. Ling. **7**, 179–211.