1015

AMERICAN
SPEECH-LANGUAGE-
HEARING
ASSOCIATION

Journal of Speech and Hearing Research, Volume 39, 379–389, April 1996

How Children Learn to Organize Their Speech Gestures: Further Evidence From Fricative-Vowel Syllables

Susan Nittrouer
Boys Town National Research Hospital
Omaha, NE

Michael Studdert-Kennedy
Haskins Laboratories
New Haven, CT

Stephen T. Neely
Boys Town National Research Hospital
Omaha, NE

Previous studies with fricative-vowel (FV) syllables have shown that the difference in overall spectrum between fricatives is less in children's speech than in that of adults, but that fricative noises show greater differences in the region of the second formant (F_2) as a function of the upcoming vowel than those of adults at corresponding points in the fricative. These results have been interpreted as evidence that children produce fricatives that are not as spatially differentiated as those of adults and that children initiate vowel gestures earlier during syllable production than adults do (Nittrouer, Studdert-Kennedy, & McGowan, 1989). The goals of the present study were (a) to replicate the previous age-related difference for F_2 with FV syllables; (b) to test the alternative interpretation that age-related differences in fricative F_2 reflect age-related differences in vocal-tract geometry; (c) to determine whether age-related differences in F_2 (and so, by inference, in articulatory organization) might extend beyond the syllable boundaries, perhaps into the schwa of a preceding unstressed syllable; and (d) to determine if gestures other than fricative gestures show less spatial differentiation in children's than in adults' speech. To these ends, F_2 frequencies were measured in schwa-fricative-vowel utterances (consisting of the fricatives /s/ and /ʃ/ and of the vowels /i/ and /ɑ/) from 40 speakers (10 each of the ages of 3, 5, 7 years, and adults) at three locations (for the entire schwa, for 10 ms of fricative noise centered at 30 ms before voicing onset, and for 10 pitch periods from vocalic center). Results of several analyses supported four conclusions: (a) the earlier finding was replicated; (b) age-related differences in vocal-tract geometry could not explain the age-related difference in vowel effects on fricative noise; (c) children master intersyllabic gestural organization prior to intrasyllabic gestural organization; and (d) unlike fricative gestures, children's vowel gestures are more spatially distinct than those of adults.

KEY WORDS: speech development, articulatory organization, vocal-tract geometry

Standard accounts of speech production assume a linear sequence of phonemic segments produced one at a time, following articulatory plans uniquely associated with each successive phoneme or feature (e.g., Abercrombie, 1965; Stevens, 1989; Stevens & Blumstein, 1978). Such accounts, based on long-standing traditions of linguistic description, have had to find ways of dealing with the fact that the acoustic and articulatory correlates of individual segments are influenced by nearby segments. Accordingly, models of "coarticulation" have been developed. Although these models differ in the particulars of how neighboring segments are said to influence each other, all agree that coarticulation is, in a sense, a secondary characteristic of speech production. That is, ideal canonical articulatory patterns are associated with each segment or its component features, but in actual speech these canonical patterns are adjusted to accommodate the patterns associated with neighboring segments or features (e.g., Daniloff & Hammarberg, 1973; Henke, 1967;

0022-4685/96/3902-0379

Hockett, 1955; Perkell, 1986). From this perspective on adult speech production, coarticulation appears as a skill that children learn after they have mastered the articulation of individual segments. That is, children learn the canonical patterns of production associated with each phoneme (or feature), and then learn how to make the appropriate adjustments to accommodate neighboring segments (Katz, Kripke, & Tallal, 1991; Kent, 1983; Sereno, Baum, Marean, & Lieberman, 1987). From this developmental perspective, the acoustic and articulatory records of children's speech should reveal less evidence of coarticulation than adult records.

In 1989, Nittrouer, Studdert-Kennedy, and McGowan published the results of a study examining the acoustic structure of children's speech for evidence of coarticulation. Using fricative-vowel (FV) disyllables consisting of the fricatives /s/ and /ʃ/ and of the vowels /i/ and /u/ (i.e., /sisi/, /ʃiʃi/, /susu/, and /ʃuʃu/), second formant (F_2) frequencies were measured in the fricative noise of the second syllable, and at the onset of voicing in the second syllable. We chose both to use disyllables and to make acoustic measures on the early portions of the second syllable precisely because we expected children to display less evidence of coarticulation than adults. We assumed that the second fricative, being subject to both perseveratory and anticipatory effects of the neighboring vowels, would allow us to observe at least some coarticulatory effect in the acoustic structure of children's speech. To our surprise, children's F_2 frequencies actually differed more between the /i/ and the /u/ contexts than those of adults at both measurement points. We interpreted our findings as evidence that the children had initiated their vowel gestures earlier in the syllable and so had "coarticulated" more extensively than the adults.

Such a conclusion is compatible with an alternative view of speech development by which the phonetic segment is the endpoint rather than the starting point of development (Studdert-Kennedy, 1987, 1991; Studdert-Kennedy & Goodell, in press). Standard descriptions of babbling during the premeaningful speech period distinguish reduplicated babble, in which syllable margins and nuclei do not change across a syllable string (e.g., /mɑmɑ/, /dædæ/) from variegated babble, in which margin and/or nucleus vary from one syllable to the next (e.g., /dædɪ/, /dɑgɪ/), and suggest that reduplicated patterns predominate early during this period with variegated patterns predominating later (Oller, 1980, 1986; Smith, Brown-Sweeney, & Stoel-Gammon, 1989; Stark, 1980).[1] Thus, the holistic, undifferentiated syllable appears to be the initial unit of speech production, from which (we hypothesize) segments gradually emerge, first by differentiation of the syllable into its gestural components, then by integration of those gestures into the recurrent articulatory-acoustic patterns that we know as consonants and vowels (Studdert-Kennedy, 1991; cf. Davis & MacNeilage, 1990). The present study is one of a series exploring this alternative view of speech development by acoustic analysis of child speech (Goodell & Studdert-Kennedy, 1993; McGowan & Nittrouer, 1988; Nittrouer, 1993, 1995; Nittrouer et al., 1989).

In these studies we have adopted the concept of "gesture" as formally defined in the articulatory phonology of Browman & Goldstein (e.g., 1986, 1989, 1992): the formation and release of a certain degree of constriction at a particular point in the vocal tract. Within the framework of articulatory phonology, gestures are the fundamental invariant units of production and perception (cf. Studdert-Kennedy & Goodell, in press). The phenomena of "coarticulation" then arise not from the adjustment of invariant units to variable contexts, but as automatic articulatory-acoustic consequences of their "coproduction," that is, of spatiotemporal overlap among gestures (Fowler, 1980; Fowler & Saltzman, 1993). The degree of articulatory-acoustic variation due to coproduction depends both on the extent to which overlapping gestures share articulators (i.e., are mechanically linked) and on the extent of their temporal overlap. (For studies of adult speech within a coproduction framework, see, e.g., Bell-Berti & Harris, 1981; Gracco, 1988; Gracco & Lofqvist, 1994; Harris, Tuller, & Kelso, 1986; Munhall & Jones, 1995; Munhall, Ostry, & Parush, 1985).

For a child learning to talk, then, one task involves learning to "coproduce" gestures in adult fashion, with the precise degree of spatiotemporal overlap that the language requires. If the primary unit of organization does indeed change as speech develops, from syllable to phoneme-sized phonetic segment, we might expect that children's speech that is not yet fully adult-like would display more extensive (or, at least, different) patterns of overlap among intrasyllabic gestures than adults' speech. In fact, many studies using phonetic transcription of children's earliest utterances reveal that a focal difficulty for the child lies in the sequencing and relative timing of the gestures that compose an utterance (e.g., Ferguson & Farwell, 1975; Menn, 1978; Menyuk, Menn, & Silber, 1986; Studdert-Kennedy & Goodell, in press; Waterson, 1971). Our own studies, cited above, are among the first to explore the development of gestural timing by means of acoustic analysis.

The present study continues the series, examining both temporal and spatial properties of child and adult utterances. The first goal of the study was to replicate the finding of Nittrouer et al. (1989), that children's fricative noises in FV syllables display greater differences in the region of F_2 as a function of upcoming vowel than those of adults at corresponding points in the fricative. As noted above, we interpreted this finding as evidence that the children initiated their vowel gestures earlier in the syllable than adults. Hodge (1990) has questioned this interpretation, however, suggesting that the observed differences in F_2 frequencies between children and adults might be due to differences in vocal-tract geometry. A second goal of this study was therefore to test this alternative hypothesis.

Clearly, children's vocal tracts are shorter than those of adults, and so a general raising of formant frequencies would be expected. Children's vocal tracts are not uniformly shorter, however, because the relative lengths of oral and pharyngeal cavities differ across children, women, and men (Fant, 1973; Goldstein, 1980). Consequently, adults and

---

[1] Other data (Davis & MacNeilage, 1995; Mitchell & Kent, 1990) suggest that the proportions of reduplicated and variegated babble remain constant across this time period. However, it must surely be the case that at some point reduplicated forms decline, leaving only utterances in which syllable shape varies across the string.

appropriately be described as a resonance of a specific vocal-tract cavity for these vowels. Therefore, we can examine possible nonlinearities across these vowels due to age-related differences in relative cavity size. Although formant frequencies are, strictly speaking, always resonances of the entire vocal tract, there is one condition under which formant frequencies can be derived by computing resonances of specific vocal-tract cavities. When the vocal-tract configuration approximates a two-tube resonator and the areas of the tubes are greatly different, we can make a valid estimate of formant frequencies from the resonant frequencies of specific cavities (Fant, 1960, 1973; Stevens, 1972, 1989, 1996). Thus, for the vowel /ɑ/, the back cavity is relatively narrow and the front cavity is relatively open, so that F_2 can be computed as a resonance of the front cavity (Fant, 1960). For /i/, the cavities are reversed: the front cavity is relatively narrow, the back cavity relatively open, so that F_2 can be computed as a resonance of the back cavity (Fant, 1973). For /u/, the two cavities are of roughly equal area, with a narrow constriction between them, so that we cannot derive valid formant frequencies by computing resonances of specific cavities. This difficulty for /u/ is exacerbated by liprounding, which lowers the frequency of all vocal-tract resonances (Fant, 1973). Because a goal of the present investigation was to determine the extent to which age-related differences in F_2 frequencies are due to age-related differences in relative cavity lengths, analyses were restricted to the vowels /i/ and /ɑ/.

Procedures

Procedures have been described previously (Nittrouer, 1993). In brief, each subject was recorded individually in a quiet room. Pictures served as prompts for collecting the speech samples.[3] Ten samples of each syllable in the carrier phrase were recorded in blocks of three (i.e., 10 blocks of three syllables each). The three pictures were rearranged between blocks so that they were presented in a different order each time.

The first 5 samples from each subject of each utterance that displayed no extraneous noise were digitized at a 25-kHz sampling rate, with low-pass filtering below 10 kHz. The portion of the utterance between the /s/ noise in "It's" and the closure for the first /b/ in "Bob" was saved to a file. CSpeech software (Milenkovic, 1987) was used for digitizing and for deriving F_2 frequencies. SPECTO (Neely & Peters, 1989) was used to obtain spectrograms, which were used to estimate F_2 frequencies before more exact measurements were made from spectral sections with CSpeech. That is, these preliminary spectrograms gave us a global perspective on each speaker's formant characteristics, thus ensuring that we were not erroneously examining a resonance other than a vocal-tract resonance (such as a subglottal or a nasal resonance) when we derived F_2 frequencies from spectral sections.

Analyses

F_2 frequencies. F_2 frequencies were obtained by LPC analysis for three acoustic sections of the utterance: for the entire schwa; for a 10-ms section centered at 30 ms before voicing onset in the fricative noise; and for 10 pitch periods centered at the temporal midpoint of the vocalic portion (i.e., the steady-state vowel). The section centered at 30 ms before voicing onset was the same as the first location for which F_2 frequencies were measured in Nittrouer et al. (1989), but 20-ms Hamming windows were used in that study. Therefore, the sample points in the 5 ms on each side of the window received decreased weightings.[4] In the present study, the sample points were all weighted the same (i.e., a rectangular window). In effect, then, both studies derived F_2 frequencies for 10-ms sections at this location.

For each of the three F_2 frequency measures, a mean for each speaker across the 5 tokens of each syllable was obtained and used in subsequent statistical analyses. For each measure, a three-way Analysis of Variance (ANOVA) was conducted, with age as the between-subjects main effect, and with consonant and vowel as the within-subjects main effects. However, these main effects were not of primary interest in this study. Rather, the two-way interactions of Age × Vowel and Age × Fricative were the focus. These two interactions provide information on potential age-related differences in the magnitude of consonant and vowel effects on F_2 frequencies and so provide information on age-related differences in gestural organization.

In addition to these ANOVAs, three planned, orthogonal comparisons were made: Adults versus All Children, 7-year-olds versus 5-year-olds, and 7- and 5-year-olds versus 3-year-olds. There were several reasons for carrying out a priori analyses. First, these comparisons could be made, even if overall F-ratios were not significant. Second, as explained above, the effects of most interest in this study were interactions, rather than main effects. Planned comparisons effectively operate as ANOVAs with just two groups, and so this procedure directly investigates group difference in the magnitude of within-subjects effects. Post hoc *t*-tests, on the other hand, cannot directly examine group differences for within-subjects effects. If one wishes to do so, a new variable that indexes the size of the effect for the

---

[3] Siren and Wilcox (1995) showed differences in anticipatory coarticulation between real words and nonsense items. In this experiment, all stimuli had the status of real words: /si/ and /ʃi/ obviously are real words, and /sɑ/ and /ʃɑ/ had been used as stimuli in a perception task with all speakers from this production study. As part of that task, listeners heard stories about /sɑ/ (a creature from another planet) and /ʃɑ/ (a king from another country), and so these syllables acquired the status of real words.

[4] An alternative to centering the fricative-noise window at 30 ms before voicing onset would have been to center this window at a point in time representing a specific proportion of the fricative, such as 2/3 of the way through. On first consideration, this alternative may seem most appropriate, given that temporal coordination among gestures is generally described in relational terms. For fricative production, however, Soli (1981) demonstrated that adults consistently and clearly initiate movement toward the vowel target at about 30 ms before the onset of voicing. This absolute temporal relation between fricative and vowel gestures may reflect aerodynamic constraints. Stop closures, for example, can be loose or tight, as long as complete closure is maintained, and speed of release can vary somewhat. Fricatives, on the other hand, require a precise degree of constriction to maintain the appropriate degree of turbulent airflow. Variability in the timing and speed of fricative release would cause variability in fricative quality.

TABLE 2. Mean F_2 frequencies for the fricative noise (Hz) and standard deviations, in parentheses, are presented in the first four rows. The FV syllable in which the measurement was made is shown in the first column. The "1989" and "present," in parentheses, refer to F_2 frequencies derived by Nittrouer et al. (1989) and the current study, respectively. Vowel ratios of F_2 are shown for /s/ and /ʃ/ in the last three rows.

FV	3	5	7	Women	Men
/si/ (1989)	2789	2691	2495	2180	1676
(present)	2839	2710	2615	2152	1778
	(207)	(288)	(124)	(171)	(97)
/ʃi/ (1989)	2960	2863	2745	2281	1812
(present)	2979	2914	2845	2279	1886
	(236)	(285)	(165)	(279)	(145)
/sɑ/	2264	2303	2205	1915	1565
	(248)	(203)	(225)	(271)	(116)
/ʃɑ/	2478	2509	2448	2038	1750
	(328)	(250)	(193)	(170)	(166)
si/sɑ	1.27	1.18	1.20	1.14	1.14
	(0.185)	(0.090)	(0.134)	(0.135)	(0.065)
ʃi/ʃɑ	1.22	1.17	1.17	1.12	1.08
	(0.155)	(0.117)	(0.115)	(0.169)	(0.047)
		All children		All adults	
si/sɑ		1.22		1.14	
		(0.144)		(0.100)	
ʃi/ʃɑ		1.19		1.10	
		(0.128)		(0.119)	

The overall ANOVA done on the F_2 frequencies from this study revealed a statistically significant age effect, $F(3,35) = 37.05$, $p < 0.001$, indicating that younger speakers generally had higher F_2 frequencies than older speakers. In addition, the Adults versus All Children planned comparison was the only significant age effect, $F(1,35) = 109.12$, $p < 0.001$. Thus, all children's groups had higher mean F_2 frequencies than the adults, but no developmental shifts were found among the children's groups.

The main effect of consonant was statistically significant, $F(1,35) = 37.58$, $p < 0.001$, indicating that F_2 frequencies for /ʃ/ were generally higher than those for /s/. Across all speakers and both vowel contexts, mean F_2 for /ʃ/ was 2514 Hz and mean F_2 for /s/ was 2326 Hz. The main effect of vowel context was also statistically significant, $F(1,35) = 125.17$, $p < 0.001$, indicating that F_2 frequencies for the /i/ context were generally higher than those for the /ɑ/ context. Across all speakers and both fricatives, mean F_2 for the /i/ context was 2614 Hz and that for the /ɑ/ context was 2225 Hz.

Of most interest to the present investigation, the Age × Vowel interaction was statistically significant, $F(3,35) = 3.96$, $p = 0.016$. This interaction indicates that the difference in F_2 frequencies between the /i/ and the /ɑ/ contexts is greater for younger than for older speakers, as illustrated in Table 2 by larger vowel ratios for younger than older speakers. Thus, at this location within the syllable (i.e., 30 ms before voicing onset), children's articulatory gestures appear to be further advanced toward their vowel targets than those of adults. The planned comparison of Adults versus All Children was statistically significant for the vowel effect, $F(1,35) = 9.28$, $p = 0.004$, but no other planned comparison was significant. Because the Age × Vowel interaction was significant, post hoc t-tests using a Bonferroni adjustment were done to investigate possible differences in the magnitude of the vowel effect for all possible two-way comparisons of groups. A single metric indexing the magnitude of vowel effect was obtained by taking the mean of the difference of /i/ F_2 minus /ɑ/ F_2 for /s/ and /ʃ/ (i.e., mean of [si-sɑ] + [ʃi-ʃɑ]). No t-test comparing two groups of children was statistically significant. However, each comparison of the adults to a children's group was significant: adults versus 7-year-olds, $t(14) = -2.64$, $p = 0.019$; adults versus 5-year-olds, $t(13) = -2.86$, $p = 0.013$; adults versus 3-year-olds, $t(11) = -3.10$, $p = 0.010$. Using the Bonferroni significance levels, the comparisons of adults versus both 5- and 3-year-olds are significant at the 0.10 level. Thus, it can be concluded that the magnitude of the vowel effect was greater for all children's groups than for adults.

Finally, a t-test was done comparing actual vowel ratios of children and adults. Because there were no significant differences among the children's groups and no Age × Fricative interaction, mean ratios (i/ɑ) across fricatives for all children were compared to mean ratios across fricatives for all adults. This analysis resulted in a significant t, $t(1,37) = -2.12$, $p = 0.04$, indicating that children's vowel ratios were significantly larger than those of adults. Thus, the results of Nittrouer et al. (1989) were replicated.

Vocalic center. Table 3 displays mean F_2 frequencies and vowel ratios for the 10 pitch periods at vocalic center. Significant main effects were those of vowel, $F(1,35) = 1219.61$, $p < 0.001$, and of age, $F(3,35) = 21.10$, $p < 0.001$. Also, the Adults versus All Children comparison for overall age effect was significant, $F(1,35) = 60.88$, $p < 0.001$.

At this point in syllable production, it again appears that children's samples show a greater acoustic difference between vowels than adults' samples do, and the statistical analysis confirms this impression: The Age × Vowel inter-

TABLE 3. Mean F_2 frequencies at vocalic center (Hz) and standard deviations, in parentheses. The FV syllable in which the measurement was made is shown in the first column. Ratios of these frequencies are shown in the last three rows.

FV	3	5	7	Women	Men
/si/	3337	3521	3354	2898	2113
	(228)	(390)	(278)	(115)	(245)
/ʃi/	3293	3518	3230	2856	2088
	(259)	(356)	(162)	(149)	(205)
/sɑ/	1671	1745	1622	1437	1214
	(244)	(148)	(149)	(118)	(76)
/ʃɑ/	1697	1782	1725	1559	1229
	(201)	(148)	(132)	(38)	(82)
si/sɑ	2.04	2.02	2.08	2.03	1.74
	(0.370)	(0.161)	(0.240)	(0.220)	(0.158)
ʃi/ʃɑ	1.95	1.98	1.88	1.83	1.70
	(0.173)	(0.162)	(0.158)	(0.114)	(0.184)
		All children		All adults	
si/sɑ		2.05		1.89	
		(0.266)		(0.236)	
ʃi/ʃɑ		1.94		1.77	
		(0.164)		(0.159)	

39 379–389 April 1996

F_2 Trajectories Through the Utterance

Figure 1 displays F_2 trajectories through the utterances /sV/ (top panel) and /ʃV/ (bottom panel) for each children's group and shows that the trajectories roughly overlap for all children's groups. Figure 2 displays the corresponding F_2 trajectories for men and women. Although the frequency range of F_2 is different for men and women, the trajectories are roughly parallel for the two groups. Finally, Figure 3 shows F_2 trajectories for 3-year-olds and women, with the Y-axis rescaled for the two groups (each scale is logarithmic and spans a frequency ratio of 2.5). The rescaling for this figure was accomplished by aligning F_2 frequencies in the schwa for women and 3-year-olds. The difference between /i/ and /ɑ/ F_2 frequencies in the schwa is roughly equivalent for both women and 3-year-olds. In the fricative, however, 3-year-olds' F_2 frequencies for /i/ and /ɑ/ are more widely separated than those of women. Although the trajectories of only two speaker groups are displayed, the similarity among all children's groups shown in Figure 1 and between women and men shown in Figure 2 indicates that a comparison between any of the children's groups and either men or women would give much the same result. Taken together, these figures (1 to 3) illustrate three important findings. First, for both fricatives the distance between the /i/ and the /ɑ/ trajectories remains relatively stable between the schwa and the fricative for adults, but not for children. Specifically, adults have increased the extent of their vowel production only minimally in the fricative over that which is seen for the schwa: Vowel ratios for adults change between these two locations by roughly 0.05, whereas for children they change by roughly 0.20 (see Tables 1 and 2). Second, most of this age-related difference is due to the fact that children's /ɑ/ trajectories do not rise in the fricative noise, as do adults' /ɑ/ trajectories. Third, trajectories are roughly parallel for adults and children between the fricative noise and vocalic center. That is, there is more similarity for children and for adults in the pattern of F_2 change across the latter two points than across the earlier two points.[8]

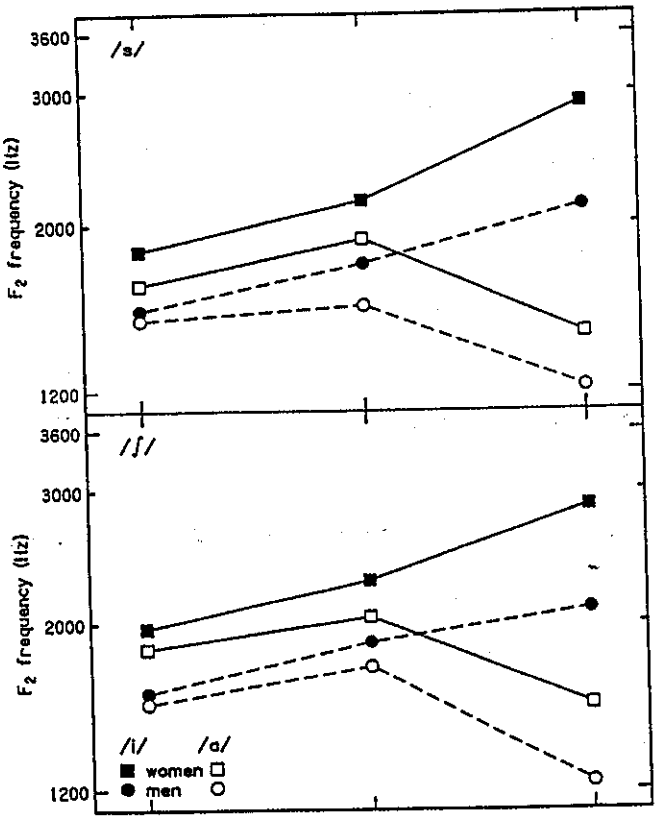

FIGURE 2. F_2 trajectories of adults across the utterance (see caption for Figure 1).

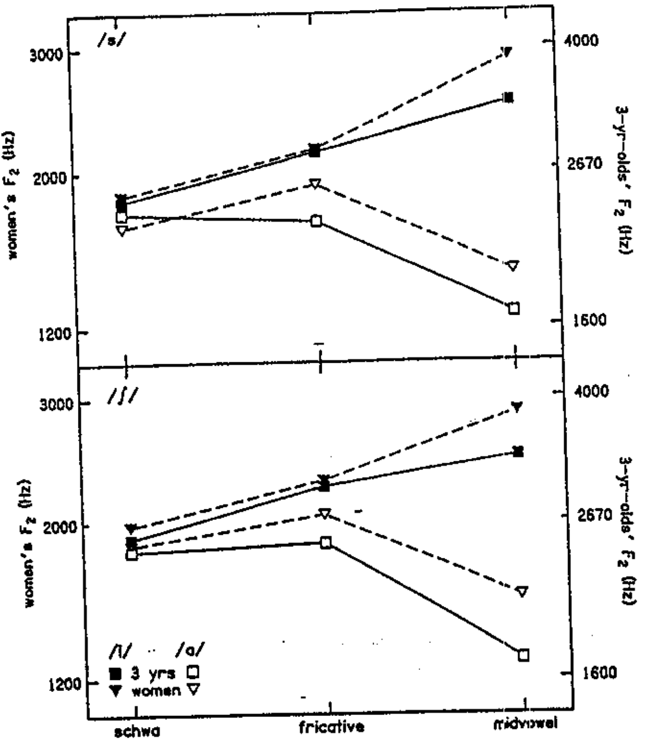

FIGURE 3. F_2 trajectories of 3-year-olds and women across the utterance, with the Y-axis rescaled for each group.

We can quantify this last observation by means of an index, formed from the ratio of F_2 at one measurement point to F_2 at the measurement point before it. A ratio of 1.00 indicates that no change occurred between the two points. Table 5 (top) displays mean ratios of F_2 in the fricative noise to F_2 for the schwa (F_2noise/F_2schwa) for each speaker group. F_2 for /i/ changes roughly the same amount for all speaker groups, but F_2 for /ɑ/ changes more for adults than for children. Table 5 (bottom) displays mean ratios of F_2 at vocalic center to F_2 in the fricative noise (F_2center/F_2noise) for each speaker group. No age-related patterns are seen in these ratios.

---

[8] Trajectories for /i/ between the fricative noise and vocalic center appear to diverge slightly for women and 3-year-olds in Figure 3. However, it will be seen (Table 5) that this is probably a spurious finding; in fact, there is no clear age-related difference in /i/ trajectories between the fricative noise and vocalic center.

the upcoming vowel was /i/. For the /u/ context, on the other hand, children's F_2 frequencies do not indicate appropriate fronting of the tongue for fricative production. Thus, fricative production appears to have been strongly linked to vowel production for these children such that the place of fricative constriction was dependent on the place of constriction for the upcoming vowel. A similar result was reported by Davis and MacNeilage (1990) for younger children producing stop-vowel syllables: They found that place of stop closure was strongly influenced by place of vowel constriction. They emphasized, as we wish to do, that it is the vowel that is influencing the consonant, not the reverse.[9]

From our evolving perspective on speech development, the results of Nittrouer et al. (1989) may be re-interpreted as showing that the intrasyllabic gestures in FV syllables were "coproduced" to a greater extent in children's than in adults' utterances. That is, children's fricative and vowel gestures showed both greater mechanical coupling and greater temporal overlap. As with the present study, that investigation used fricatives and vowels that involved primarily lingual gestures. In 1993, Nittrouer found that children were able to differentiate intrasyllabic gestures as adults do when the consonant and vowel made use of articulators that had greater anatomical independence from each other, such as when the jaw is the primary articulator for the consonant gesture and the tongue body is for the vowel gesture. The present study provided evidence, as Nittrouer et al. had, that children lack independent control (in comparison to that of adults) over intrasyllabic gestures when the articulators required to produce the consonant and vowel are anatomically proximal. Thus, one aspect of speech development may very well involve gaining greater independence among gestures that use articulators of increasingly closer anatomical proximity. Similar developmental trends have been reported for other skilled actions. For example, Thelen (e.g., 1985) has reported that when infants move their legs, flexion (or extension) of the hips, knees, and ankles occurs at the same time. That is, the movements of any one of these "articulators" is closely tied to the movements of the others. In fact, Thelen concluded that: "The global movements of the newborn period are refined into more precise coordinative structures: but at the same time, these increasingly differentiated movements are assembled into larger functional units" (p. 8). With an upward shift in age range, this statement could describe the development of movement control for speech production. For children's earliest utterances, global movements of the vocal tract result in syllable production: that is, a constriction at some point in the vocal tract is released to a more open vocal-tract configuration. As the child matures and gains experience producing vocal-tract movements they are refined into more precise coordinative structures (i.e., groups of articulators that function together to achieve a goal) that can then be produced relatively independently and recombined in new and different ways.

In summary, results of this analysis are consistent with the hypothesis that learning to coordinate articulatory gestures as required to impose adult-like phonemic structure on utterances is a long process, extending well into childhood. Unlike earlier theories of developmental speech production that suggested the task facing the child is learning how to "coarticulate" consonant and vowel gestures, we suggest that the child's task is learning how to produce consonant and vowel gestures with the spatial distinctiveness and the relational timing that characterizes adult speech.

Acknowledgments

This work was supported by NIH Grant DC-00633 to the first author, and was conducted, in part, at Haskins Laboratories and at the University of Nebraska at Omaha. Winifred McGowan, Carol Manning, and Gina Meyer helped with acoustic and statistical analyses. Richard S. McGowan provided comments on an earlier draft.

References

Abercrombie, D. (1965). *Studies in phonetics and linguistics.* London: Oxford University.

American National Standards Institute (1969). *Specifications for audiometers.* (ANSI S3.6-1969). New York: ANSI.

Bell-Berti, F., & Harris, K. S. (1981). A temporal model of speech production. *Phonetica, 38.* 9–20.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonological Yearbook, 3.* 219–252.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology. 6.* 201–251.

Browman, C. P., & Goldstein, L. (1992). Gestural structures and phonological patterns. In I. G. Mattingly & M. Studdert-Kennedy (Eds.). *Modularity and the motor theory of speech perception.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Daniloff, R. G., & Hammarberg, R. E. (1973). On defining coarticulation. *Journal of Phonetics. 1.* 239–248.

Davis, B. L., & MacNeilage, P. F. (1990). Acquisition of correct vowel production: A quantitative case study. *Journal of Speech and Hearing Research, 33.* 16–27.

Davis, B. L., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research, 38,* 1199–1211.

Fant, G. (1960). *Acoustic theory of speech production.* Mouton: The Hague.

Fant, G. (1973). *Speech sounds and features.* Cambridge, MA: MIT Press.

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

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8,* 113–133.

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech. 36,* 171–195.

Goldman, R., & Fristoe, M. (1969). *Goldman-Fristoe Test of Articulation.* Circle Pines, MN: American Guidance Service.

Goldstein, U. G. (1980). *An articulatory model for the vocal tracts of growing children.* Unpublished doctoral dissertation. Massachusetts Institute of Technology. Cambridge.

Goodell, E. W., & Studdert-Kennedy, M. (1993). Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: A longitudinal study. *Journal of Speech and Hearing Research. 36.* 707–727.

Gracco, V. L. (1988). Timing factors in the coordination of speech movements. *Journal of Neuroscience. 8.* 4628–4639.

Gracco, V. L., & Lofqvist, A. (1994). Speech motor coordination

---

[9] Note, however, that Nittrouer (1993) found that a /t/ closure had an influence on F_2 frequency much longer into the following vowel in the CVs of 3- to 7-year-olds than in those of adults.