

Some observations on the development of anticipatory coarticulation

Bruno H. Repp

Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695

(Received 12 July 1985; accepted for publication 21 January 1986)

The influence of vocalic context on various temporal and spectral properties of preceding acoustic segments was investigated in utterances containing [ə#CV] sequences produced by two girls aged 4;8 and 9;5 years and by their father. The younger (but not the older) child's speech showed a systematic lowering of [s] noise and [t^h] release burst spectra before [u] as compared to [i] and [æ]. The older child's speech, on the other hand, showed an orderly relationship of the second-formant frequency in [ə] to the transconsonantal vowel. Both children tended to produce longer [s] noises and voice onset times as well as higher second-formant peaks at constriction noise offset before [i] than before [u] and [æ]. All effects except the first were shown by the adult who, in addition, produced first-formant frequencies in [ə] that anticipated the transconsonantal vowel. These observations suggest that different forms of anticipatory coarticulation may have different causes and may follow different developmental patterns. A strategy for future research is suggested.

PACS numbers: 43.70.Bk, 43.70.Fq

INTRODUCTION

The development of coarticulation in children's speech production is a topic of great current interest, though data are still scarce. It is commonly assumed that children coarticulate less than adults, especially with regard to anticipatory effects which are said to be "planned," and there is some preliminary evidence from acoustic analyses and from physiological studies to support this notion (see Kent, 1983). A reduction in the extent of coarticulation is taken to reflect an underlying general tendency toward producing speech segment by segment, which decreases with age (Kent, 1983).

In the present pilot study, acoustic measures of several anticipatory coarticulation effects were obtained from two children and their father. Because of this small sample size, the data are intended to stimulate further research rather than to establish firm developmental patterns. Nevertheless, the familial relatedness of the three subjects may have reduced irrelevant individual differences, thus lending the data

somewhat more generality than a sample of three unrelated individuals would have provided.

I. METHODS

A. Subjects

The subjects were two sisters aged 4;8 and 9;5 years and their father (the author). The children are monolingual speakers of American English; the adult is a native speaker of German who speaks English almost exclusively, though not without an accent.

B. Utterances and procedure

Each subject produced six words, *sea, sand, soup, tea, tan, and tooth*, five times in the carrier phrase "I like the" The children repeated each sentence after their father, taking turns at speaking first.¹ The recordings were made in a sound-attenuated booth, with all three talkers facing a single microphone.

C. Acoustic analysis

The children's utterances were low-pass filtered at 9.6 kHz and digitized at a 20-kHz sampling rate with high-frequency pre-emphasis. A 24-coefficient LPC analysis with automatic peak picking and subsequent hand-editing of inconsistencies yielded estimates of formant frequencies. A numerical index of the relative high-frequency content of the spectrum in a given 20-ms analysis frame was provided by the first LPC reflection coefficient, which is the (negative, normalized) average of the cosine-weighted spectrum (see Markel and Gray, 1976). Temporal measures were obtained from oscillographic displays. Means and standard deviations of the various measures were computed across the five tokens of each utterance. The adult's utterances were analyzed similarly, using a 10-kHz sampling rate for digitization and a 14-coefficient LPC model.

II. RESULTS

A. Effects of vocalic context on voiceless interval durations

Table I shows two coarticulatory effects in the temporal domain: [s] noise durations were longest before [i] and shortest before [æ], and [t^h] burst plus aspiration (i.e., acoustic voice onset time or VOT) was longest before [i] also. In separate one-way analyses of variance, the [s] duration differences reached significance for the younger child [$F(2,12) = 4.47, p = 0.0354$] while the VOT differences reached significance for the older child [$F(2,12) = 6.24, p = 0.0139$]. Both effects were highly significant in the adult [$F(2,12) = 59.0, p < 0.0001$, and $F(2,12) = 7.35, p = 0.0083$, respectively]. All three talkers showed similar patterns, however, and the lower reliability of the children's results may be attributed to their greater variability (cf. Smith *et al.*, 1983).²

B. Effects of vocalic context on constriction noise spectra

A lowering of [s] frication and [t^h] release burst spectra due to anticipatory lip rounding for [u] has been observed in adults (Mann and Repp, 1980; Sereno *et al.*, 1985; Soli, 1981; Turnbaugh *et al.*, 1985; Zue, 1976). Visual inspection of average [s] noise offset and [t^h] burst onset spectra (both representing noise immediately preceding the release of the constriction) revealed a clear shift of the ener-

TABLE I. Means and standard deviations (in parentheses) of some voiceless segment durations (ms).

	Child A (4;8 yrs)	Child B (9;5 yrs)	Adult
[s(V)] fricative noise			
V = [i]	232 (24)	222 (34)	228 (9)
V = [æ]	184 (25)	189 (21)	173 (9)
V = [u]	207 (27)	202 (17)	197 (9)
[t ^h (V)] burst + aspiration (VOT)			
V = [i]	90 (16)	107 (5)	76 (10)
V = [æ]	75 (12)	89 (10)	64 (15)
V = [u]	84 (21)	84 (16)	50 (7)

TABLE II. Means and standard deviations (in parentheses) of F_2 frequencies at [s] noise offset, at [t^h] burst onset, and in the preceding [ə] (Hz).

	Child A (4;8 yrs)	Child B (9;5 yrs)	Adult
(A)			
at [s(V)] noise offset			
V = [i]	3241 (168)	2385 (92)	1957 (66)
V = [æ]	2899 (186)	2331 (120)	...
V = [u]	2866 (159)	2203 (90)	1547 (51)
at [t ^h (V)] burst onset			
V = [i]	3176 (127)	2492 (144)	2191 (259)
V = [æ]	2998 (63)	2357 (33)	...
V = [u]	3050 (90)	2430 (147)	1757 (116)
(B)			
in [ə] preceding [#sV]			
V = [i]	2846 (123)	2107 (50)	1482 (26)
V = [æ]	2885 (114)	2049 (59)	1421 (15)
V = [u]	2863 (104)	2018 (64)	1490 (75)
in [ə] preceding [#t ^h V]			
V = [i]	2866 (169)	2168 (55)	1467 (18)
V = [æ]	2857 (108)	2154 (24)	1418 (45)
V = [u]	2934 (52)	2077 (47)	1418 (45)

gy maximum towards lower frequencies (5–6 kHz) before [u] as compared to [i] and [æ] (around 8 kHz) in the younger child. Neither the older child nor the adult showed such a shift.

To gain statistical support for these observations, and to examine the time course of the effect in the [s] noise, analyses of variance were conducted on the average first LPC reflection coefficients obtained for three (slightly overlapping) consecutive 60-ms segments of the [s] noises of each talker. For the younger child, there were highly significant effects of vocalic context [$F(2,12) = 14.22, p = 0.0007$] and of time [$F(2,24) = 19.80, p < 0.0001$], as well as a two-way interaction [$F(4,24) = 5.56, p = 0.0026$]. The coarticulatory effect increased with proximity to the vowel but was clearly present throughout the fricative noise. The older child, on the other hand, showed no significant effects, even though spectral variability was lower. The adult talker also showed significant effects of vocalic context [$F(2,12) = 9.89, p = 0.0029$] and of time [$F(2,24) = 5.98, p = 0.0078$], but the pattern was different: The average [s] spectra were lowest before [æ] and highest before [u]; moreover, these differences resided mainly between 1–3 kHz.

The noise spectra were also examined for peaks in the second-formant (F_2) region that anticipate F_2 in the following vowel, a lingual coarticulation effect that is distinct from the global spectral shifts due to anticipatory lip rounding (see Soli, 1981). The F_2 frequency estimates derived from the 20-ms LPC analysis frames closest to [s] noise offset and [t^h] burst onset are reported in Table IIA. There was a significant effect of vocalic context for the younger child [$F(2,24) = 11.28, p = 0.0004$]. In both [s] offset and [t^h] onset spectra, F_2 was highest preceding [i]. The older child, despite more pronounced F_2 peaks and lower variability, showed only a nonsignificant tendency in the same direction [$F(2,24) = 3.32, p = 0.0531$]. The adult's F_2 peaks were significantly higher before [i] than before [u]

[$F(1,16) = 50.36, p < 0.0001$]; before [æ], no reliable $F2$ peaks could be found (cf. Soli, 1981).

C. Effects of vocalic context on [ə] formant frequencies

Vowel-to-vowel anticipatory coarticulation across an intervening consonant has been observed in adults, especially in [ə] (Alfonso and Baer, 1982; Fowler, 1981). Table IIB shows means and standard deviations of $F2$ frequencies averaged over the whole voiced signal portion corresponding to [ə] in the word *the* as a function of following consonant and vowel. There were no systematic contextual effects for the younger child. The older child, in contrast, showed a systematic decrease of $F2$ as the vowel in the following syllable changed from [i] to [æ] to [u] [$F(2,24) = 7.75, p = 0.0025$] as well as higher $F2$ frequencies preceding [t^h] than [s] [$F(1,24) = 15.85, p = 0.0006$]. Both effects were present throughout the [ə] vowel. The first formant ($F1$) did not show any significant differences for either child. The adult also showed a significant effect of vowel context on $F2$ [$F(2,24) = 5.32, p = 0.0123$]; overall $F2$ frequencies were highest preceding [i] and lowest preceding [æ]. In addition, he showed an effect on $F1$, which was significantly higher (by about 33 Hz) preceding [æ] than preceding [i] and [u] [$F(2,24) = 8.31, p = 0.0018$], thus anticipating the $F1$ differences between these vowels.

III. DISCUSSION

It is not possible to derive any conclusions about general developmental trends from these limited data. Nevertheless, they may serve as a basis for formulating hypotheses about the development of anticipatory coarticulation, to be tested in the future with larger subject groups or in longitudinal studies.

Two coarticulatory effects in the temporal domain were shown by both children and by the adult, though with different degrees of reliability. One of these, the effect of the following vowel on [s] noise duration, may be due to an earlier release of the constriction preceding more open vowels (Schwartz, 1969). DiSimoni (1974) and Weismer and Elbert (1982) have obtained similar differences in preschool children. The other effect apparently shown by all three subjects was that of vowel context on VOT. Related findings in the literature (Fourakis, 1986; Klatt, 1975; Port and Rotunno, 1979; Weismer, 1980) are at least partially consistent with the longer VOT's preceding [i] observed here. These effects may have kinematic or aerodynamic causes that make them difficult to avoid at any age.

A third effect that was probably present in all three talkers, though it was not quite significant in the older child, concerns differences in the location of $F2$ peaks at the release of a fricative constriction or of a stop occlusion. These differences probably reflect differences in tongue body position in anticipation of the upcoming vowel (Soli, 1981), although anticipatory lip rounding may also play a role. Similar effects were found in a 3;6-year-old child by Sereno *et al.*

(1985), and in several 3- and 5-year-old children by Turnbaugh *et al.* (1985). This may be another obligatory effect; without any anticipation, the vowel might sound abnormally diphthongized.

By contrast, certain other coarticulatory effects may be optional and subject to developmental trends. Changes in $F2$ of [ə] in anticipation of the later-occurring vowel clearly were shown only by the older child and the adult. This effect probably reflects differences in tongue body position (Alfonso and Baer, 1982); note that it was not prevented by an intervening alveolar consonant which also involves the tongue (see Recasens, 1984). This relatively long-range anticipatory lingual coarticulation across an "obstacle" may be a skill that is acquired relatively late as a child gets acquainted with the fine details of spoken language. The same might be said about the vocalic context effect on $F1$ frequency in [ə], which was shown by the adult alone and may reflect anticipatory adjustments in jaw elevation. Note that, to the extent that these articulatory postures are not maintained during the intervening consonant constriction, they must indeed be considered "planned."

The most unusual finding concerns the overall weighting of constriction noise spectra. A lowered [s] noise or [t^h] release burst spectrum before rounded vowels such as [u] most likely reflects an effect of anticipatory lip rounding, although changes in tongue body position could also play a role (Carney and Moll, 1971). Such an effect was observed very clearly in the younger child but not in the older child, and it was reversed in the adult. While the reversal may be atypical (it could reflect back cavity resonances brought into play by leaky [s] constrictions characteristic of this adult speaker), it is interesting to note that Nittrouer (1985), in a recent thorough developmental study, has observed that fricative-vowel coarticulation (in terms of global spectral shifts in the noise) does decline with age. The present data are consistent with such a trend, even though its reasons are far from clear at present.

IV. CONCLUSIONS

The various patterns of results observed in this pilot study suggest that phenomena commonly lumped together under the heading of "coarticulation" may have diverse origins and hence different roles in speech development. Some forms of coarticulation are an indication of advanced speech production skills whereas others may be a sign of articulatory immaturity, and yet others are neither because they simply cannot be avoided. Therefore, it is probably not wise to draw conclusions about a general process called coarticulation from the study of a single effect. Indeed, such a general process may not exist. It is suggested that future research adopt the multi-pronged approach illustrated by this pilot study to examine the interrelationships among diverse coarticulatory phenomena, their individual causes, and their patterns of development.

ACKNOWLEDGMENTS

This research was supported by NICHD Grant HD-01994 to Haskins Laboratories. I am grateful to Catherine

Best, Sarah Hawkins, Joanne Miller, Susan Nittrouer, Daniel Recasens, Sigfrid Soli, Michael Studdert-Kennedy, Douglas Whalen, and Grace Yeni-Komshian for helpful comments on earlier drafts.

¹Apart from overall timing and intonation, it seems unlikely that the children directly mimicked any phonetic features of the adult's productions. Rather, it is assumed that the children generated their utterances from lexical representations of the (known) target sentences.

²The effects of vowel context on [t^h] closure duration and on the total [t^h] voiceless interval seemed less systematic. In a combined analysis of [s] and total [t^h] durations, however, none of the talkers showed a significant consonant × vowel interaction, so that the effect of vowel context on the two voiceless interval durations may have been similar (cf. Weismer, 1981). It might also be noted that the average durations of the [s] and [t^h] voiceless intervals were virtually identical in all three talkers (cf. Weismer, 1980).

Alfonso, P. J., and Baer, T. (1982). "Dynamics of vowel articulation," *Lang. Speech* 25, 151-173.

Carney, P. J., and Moll, K. L. (1971). "A cinefluorographic investigation of fricative consonant-vowel coarticulation," *Phonetica* 23, 193-202.

DiSimoni, F. G. (1974). "Influence of vowel environment on the duration of consonants in the speech of three-, six-, and nine-year-old children," *J. Acoust. Soc. Am.* 55, 360-361.

Fourakis, M. (1986). "A timing model for word-initial CV-syllables in Modern Greek," *J. Acoust. Soc. Am.* (in press).

Fowler, C. A. (1981). "Production and perception of coarticulation among stressed and unstressed vowels," *J. Speech Hear. Res.* 40, 127-139.

Kent, R. D. (1983). "The segmental organization of speech," in *The Production of Speech*, edited by P. F. MacNeilage (Springer, New York), pp. 57-90.

Klatt, D. H. (1975). "Voice-onset-time, frication, and aspiration in word-initial consonant clusters," *J. Speech Hear. Res.* 18, 686-706.

Mann, V. A., and Repp, B. H. (1980). "Influence of vocalic context on perception of the [ʃ]-[s] distinction," *Percept. Psychophys.* 28, 213-228.

Markel, J. D., and Gray, A. H., Jr. (1976). *Linear Prediction of Speech* (Springer, Berlin).

Nittrouer, S. (1985). "The role of coarticulation in the perception and production of speech by young children (3 to 7 years)," unpublished Ph.D. dissertation (City University of New York).

Port, R. F., and Rotunno, R. (1979). "Relation between voice-onset time and vowel duration," *J. Acoust. Soc. Am.* 66, 654-662.

Recasens, D. (1984). "Vowel-to-vowel coarticulation in Catalan VCV sequences," *J. Acoust. Soc. Am.* 76, 1624-1635.

Schwartz, M. F. (1969). "Influence of vowel environment upon the duration of /s/ and /ʃ/," *J. Acoust. Soc. Am.* 46, 480-481.

Sereno, J. A., Baum, S. R., Marean, G. C., and Lieberman, P. (1985). "Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children," *J. Acoust. Soc. Am. Suppl.* 1 77, S7-S8.

Smith, B. L., Sugarman, M. D., and Long, S. H. (1983). "Experimental manipulation of speaking rate for studying temporal variability in children's speech," *J. Acoust. Soc. Am.* 74, 744-749.

Soli, S. D. (1981). "Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation," *J. Acoust. Soc. Am.* 70, 976-984.

Turnbaugh, K. R., Hoffman, P. R., Daniloff, R. G., and Absher, R. (1985). "Stop-vowel coarticulation in 3-year-old, 5-year-old, and adult speakers," *J. Acoust. Soc. Am.* 77, 1256-1257.

Weismer, G. (1980). "Control of the voicing distinction for intervocalic stops and fricatives: Some data and theoretical considerations," *J. Phonet.* 8, 427-438.

Weismer, G. (1981). "Temporal characteristics of the laryngeal devoicing gesture for voiceless consonants and fricative-stop clusters: Influences of vowel environment and speaker age," *J. Acoust. Soc. Am. Suppl.* 1 69, S68.

Weismer, G., and Elbert, M. (1982). "Temporal characteristics of 'functionally' misarticulated /s/ in 4- to 6-year-old children," *J. Speech Hear. Res.* 25, 275-287.

Zue, V. W. (1976). "Acoustic characteristics of stop consonants: A controlled study," Lincoln Lab. Tech. Rep. No. 523 (Lexington, MA).