# THE CONTROL OF TIMING IN CHILDREN'S SPEECH

Sarah Hawkins, Dental Research Center, University of North Carolina, Chapel Hill, NC 27514, U.S.A.

It is generally acknowledged that temporal and prosodic variables significantly affect speech intelligibility. For example, adult listeners derive considerable information about the syntax and stress pattern of sentences, when segmental cues are either distorted by spectral rotation or absent because the sentence is hummed. The role of prosody in defining syntactic boundaries has been demonstrated with stylised synthetic intonation contours and with prosody pitted against syntax in cross-spliced sentences. The duration alone of both phonemes and larger units can be crucial to speech intelligibility. Additionally, adults appear to be particularly sensitive to the rhythmic onset of stressed syllables, both when listening to speech and when tapping to the rhythm of their own speech. The listener appears to anticipate when stresses will occur and focuses attention at these times. (Documentation of the above points may be found in papers for the other Symposia of this Congress and in Cohen and Nooteboom (1975).) This integrative and predictive role of prosodic cues figures prominently in recent models of speech perception. For example, Pisoni and Sawusch (1975) suggest prosodic cues may form an interface between low-level segmental information and higher levels of syntax and meaning. Martin (1972) has elaborated the notion of the predictive role of rhythm in speech perception, pointing out that efficient perceptual strategies such as attention-cycling between input and output can be facilitated when the signal need not be monitored continuously.

What relevance have these observations about adult speech to children's perception and production of speech? Although the adult listener may be assumed to attend only minimally to much of the acoustic signal, this cannot be assumed for the child. The young child lacks the linguistic and nonlinguistic experience that would allow him/her to "fill in" a large proportion of the message on the basis of knowledge shared with the speaker. Our sparse knowledge of children's perceptual abilities supports this distinction between adults' and children's perception of speech. For example, although infants of less than 16 weeks can discriminate between stimuli differing in some durational aspects, such as VOT (e.g. Eimas et al., 1971) and syllable duration (Spring and Dale, 1977), children as old as 4-6 years do not necessarily use these durational cues in the same way as adults (Zlatin and Koenigsknecht, 1975; Simon and Fourcin, 1978; Higgs and Hodson, 1978). We know more about children's speech production than their perception; the last 5 or 6 years have provided data on children's timing in phrases, words, syllables (Hawkins and Allen, 1978), segments, and subsegmentals such as VOT (for additional references see below). Many of these studies have demonstrated that while some aspects of children's speech timing resemble those of adult speech from quite an early age, (about 2-4 years), other aspects do not mature until much later (up to about 9-11 years).

The question becomes when and why there are differences between adults and children. Do timing rules appear in children's speech simply as a consequence of increasing neuromuscular coordination, and only gradually come to serve a perceptual function? Or does the child learn the perceptual function of such cues and attempt to produce them in his/her own speech? In the latter case the age when adult timing relationships appear would depend partly on neuromuscular abilities and partly upon the age when their perceptual function is recognised.

This paper discusses ways in which we might distinguish between the above possibilities, using data from children's speech production. The aim is to provide a conceptual framework that will be useful in thinking about children's timing control, as a first step towards formulating a theory of the development of speech timing.

I begin from the position that the child's perception of speech neither is essentially mature before s/he begins to speak (cf. Smith, 1973), nor develops concurrently with production (cf. Waterson, 1970, 1971a,b). Rather, I shall assume that while the young speaker perceives some parts of the speech signal quite maturely, s/he perceives other parts only as unanalysed 'noise'. This view is similar to that of Ingram (1974), except that I assume that the position of the 'noise' may not be fixed in the signal in a segment-by-segment manner. The approach is polysystemic: the child's systems of perception and production both may be described in terms of quasi-independent subsystems, any or all of which may be in a state of considerable flux at a given time.

I assume also that processes manifested in the child's speech will appear in adult speech and most (but perhaps not all) processes of adult speech will appear in child speech. What distinguishes the two is the <u>domain of influence</u> of each process. Thus in adult speech we may find evidence for both hierarchically integrated "comb" models and sequentially ordered "chain" models of timing (Bernstein, 1967; Kozhevnikov and Chistovich, 1965; Ohala, 1975). Phonemes may be integrated into syllables, for example, consistent with the "comb" model, while the timing of successive higher units may be relatively independent of each other, consistent with a "chain" model. In the young child's speech, such integration implying a "comb" model may not be evident at the syllabic level, but similar integration may occur with less complex units. This reasoning suggests that the child's task in learning to speak fluently is not so much that of learning new routines as of applying similar routines to increasingly more complex domains, thereby integrating the elements of these domains into functional units (c.f. Turvey's (1977) action plans). During this learning, the role played by different processes may change. Auditory feedback in children's speech, for example, appears to have a qualitatively different effect than in adults' speech (e.g. Fry, 1966; MacKay, 1967).

Bearing the above assumptions in mind, let us consider some of the different factors that are likely to affect the speech learning process, together with examples of evidence for their existence within speech timing. Three such factors will be distinguished of which only the second and third are mutually exclusive: (1) processes common to all motor skill development; (2) temporal distinctions that serve as primary perceptual cues; and (3) temporal regularities that do not function as primary perceptual cues.

<u>Processes common to all motor skill development</u> will apply to all aspects of speech development; examples are slower and more variable performance. These phenomena have been demonstrated for speech in many studies and at many levels of analysis from the phrase to the segment (e.g. Eguchi and Hirsh, 1969; DiSimoni, 1974a,b; Tingley and Allen, 1975; Kent and Forner, 1977; Keating and Kubaska, 1978; Smith, 1978; Hawkins, in press). It could be argued that slower durations occur because the child possesses an articulation-dominant system, whereas the (English-speaking) adult uses a timing-dominant system (cf. Ohala, 1970). Nevertheless, the basis of such articulation-dominance is plausibly immature neuromuscular coordination, requiring more time to achieve adequate articulatory targets. Where a phonological length distinction occurs, the child seems to learn to shorten rather than lengthen articulatory units to produce it. This has been suggested for example for the effect of position-in-utterance on word duration (Keating and Kubaska, 1978), for the development of unstressed syllable production (Allen and Hawkins, in press), and in older children for the effect of clustering on consonant duration (Gilbert and Purves, 1977; Hawkins, in press).



Even in such apparently simple cases as longer duration, a single effect may have more than one underlying cause. For example, measuring /b, d, t/ durations in simple environments, Smith (1978) observed that /t/ was 40% longer in the speech of 2 and 4 year olds than might be expected just on the basis of estimates of the degree of durational increase due to neuromotor immaturity. Smith suggested 3 possible causes for this, any or all of which may have contributed to the observed effect: (i) an effort to increase perceptual differences between /t/ and /d/, (ii) greater complexity of the tongue tip innervation required for /t/ than /d/, and (iii) greater complexity of laryngeal adjustments for voiceless stops over voiced ones.

<u>Temporal distinctions that serve as primary perceptual cues</u> are likely to be detected by the child relatively early as long as they do not signal, for example, syntactic or semantic distinctions beyond the child's comprehension. Hence they should appear in the child's speech in an order reflecting the complexity of the neuromotor coordination required. I shall discuss two examples: phonemically conditioned vowel duration, and voice onset time (VOT) in stops.

Vowel duration functions in English as a primary perceptual cue to the voicing of following consonants, with longer vowels preceding voiced consonants. There is some evidence that this is a distinction that occurs naturally and has been exaggerated in some languages, such as English (Lisker, 1974). Such evidence would suggest that the child might learn to produce the mature voiced-voiceless ratios relatively early. Naeser (1970) found they were present by 21 months of age and in fact preceded control of the voicing feature that governs the distinction in adult speech.

The development of VOT control in stops nicely illustrates the following points: (i) perception and production do not always develop hand in hand, and (ii) a phonemic distinction that may be legitimately regarded as lying along a single phonological dimension should not necessarily be treated as two extremes of a unitary process in a theory of speech development. The development of the voicing contrast in English has been studied longitudinally (e.g. Kewley-Port and Preston, 1974; Macken and Barton, 1978) and cross-sectionally (e.g. Menyuk and Klatt, 1975; Zlatin and Koenigsknecht, 1976; Gilbert, 1977). It has been consistently found that children make a distinction between short-lag (voiced) and long-lag (voiceless) stops fairly early (around 2 years), but at this stage only the short-lag distribution resembles the adult form. It is not until much later (around 6 years or more) that the long-lag VOT distribution resembles that of the adults. The difference in age of mastery of the two VOT categories is commonly accepted as being due to differences in the neuromuscular coordination required: short-lag stops allow considerable variability in the coordination of laryngeal and oral activity, whereas long-lag stops require rather precise coordination.

<u>Temporal regularities that do not function as primary perceptual cues</u>, especially those that appear to provide no perceptual information, would be expected to be acquired as the child's action plans (articulatory programs) become more sophisticated. An example is the reduction of the duration of consonants in clusters. Although many of the durational differences between clustered and unclustered consonants are perceptible, they may not serve a perceptual function (Klatt, 1976). The age when children produce these durational modifications varies according to the type of cluster, but the last ones are probably not mastered until 9-11 years (Gilbert and Purves, 1977; Hawkins, in press). This is later than many of the temporal phenomena discussed above. Hawkins (in press, in preparation) discusses evidence from these data suggesting that many clusters undergo considerable reorganisation as complex units, rather than there being simply durational



reduction of each segment. Intermediate stages may involve uneven rates of development and durational changes in the opposite direction from that finally required. These observations, together with the late attainment of mature durational relationships, are consistent with the development of increasingly sophisticated motor action plans by the integration of subroutines.

This paper has discussed some of the considerations that should be included in a theory of the development of speech timing within a polysystemic, parallel processing approach. It suggests that adults and children will differ in speech production processes not so much in the nature of those processes as in their relative importance and their domain of influence. The child's 'system' cannot be regarded as static at any time, but rather as reflecting the effects of several continually changing systems that replace each other during development. Changes in one subsystem may affect others, producing either progression or temporary regression. Furthermore, a given phenomenon observed in development may have several causes, whose effects may all work in the same direction or in conflicting directions. Consistent with these points, the development of speech timing may be usefully considered in the contexts of maturing neuromuscular coordination and the perceptual cueing function of timing rules. Neuromuscular immaturity influences patterns of production both across-the-board and in specific contexts. It is reasonable to expect timing rules that are primary perceptual cues to be implemented earlier than those that are not, assuming both that the degree of neuromuscular complexity is constant and that the child perceives the distinction as linguistically relevant.

<u>Bibliography</u>


Allen, G. D. and S. Hawkins (in press): "Trochaic rhythm in children's speech", in <u>Current Issues in the Phonetic Sciences</u>, H. Hollien and P. Hollien (eds.).

Bernstein, N. A. (1967): <u>The Coordination and Regulation of Movements</u>, Oxford: Pergamon Press.

Cohen, A. and S. G. Nooteboom (1975): <u>Structure and Process in Speech Perception</u>, Berlin: Springer-Verlag.

DiSimoni, F. G. (1974a): "Effect of vowel environment on the duration of consonants in the speech of 3-, 6-, and 9-year-old children", <u>JASA</u> 55, 360-361.

DiSimoni, F. G. (1974b): "Influence of consonant environment on duration of vowels in the speech of 3-, 6-, and 9-year-old children", <u>JASA</u> 55, 362-363.

Eguchi, S. and I. J. Hirsh (1969): "Development of speech sounds in children", AcOtolaryng 257, 1-51.

Eimas, P., E. Siqueland, P. Jusczyk and J. Vigorito (1971): "Speech perception in infants", Science 171, 303-306.

Fry, D. B. (1966): "The development of the phonological system in the normal and the deaf child", in The Genesis of Language, F. Smith and G. A. Miller (eds.), 187-206, Cambridge, Mass.: MIT Press.

Gilbert, J. H. V. (1977): "A voice onset time analysis of apical stop production in three-year-olds", J. Ch. Lang. 4, 103-110.

Gilbert, J. H. V. and B. A. Purves (1977): "Temporal constraints on consonant clusters in child speech production", J. Ch. Lang. 4, 417-432.

Hawkins, S. (1973): "Temporal coordination of consonants in the speech of children: Preliminary data", JPh 1, 181-217.

Hawkins, S. (in press): "Temporal coordination of consonants in the speech of children: Further data", JPh.

Hawkins, S. (in preparation): "Processes in the development of speech timing control".

Hawkins, S. and G. D. Allen (1978): "Acoustic-phonetic features of stressed syllables in the speech of 3-year-olds", JASA 63, Suppl. 1, S56.

Higgs, J. W. and B. W. Hodson (1978): "Phonological perception of word-final obstruent cognates", JPH 6, 25-35.

Ingram, D. (1974): "Phonological rules in young children", J. Ch. Lang. 1, 49-64.

Keating, P. and C. Kubaska (1978): "Variation in the duration of words", JASA 63, Suppl. 1, V10.

Kent, R. D. and L. L. Forner (1977): "A developmental study of speech production: Data on vowel imitation and sentence repetition", JASA 62, Suppl. 1, 0015.

Kewley-Port, D. and M. Preston (1974): "Early apical stop production: A voice onset time analysis", JPh 2, 195-210.

Klatt, D. H. (1976): "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence", JASA 59, 1208-1221.

Kozhevnikov, V. A. and L. A. Chistovich (1965): "Speech: Articulation and perception", Joint Publications Research Service, Washington, D. C.: 30, 543.

Lisker, L. (1974): "On "explaining" vowel duration variation", Haskins Labs.: Status Report on Speech Research SR-37/38, 225-232.

MacKay, D. G. (1967): "Metamorphosis of critical interval: Age-linked changes in the delay in auditory feedback that produces maximal disruption of speech", JASA 43, 811-821.

Macken, M. A. and D. Barton (1977): "A longitudinal study of the acquisition of the voicing contrast in American-English word-initial stops, as measured by voice-onset time", Stanford University: Papers and Reports on Child Language Development 14, 74-120.



Martin, J. G. (1972): "Rhythmic (hierarchial) versus serial structure in speech and other behavior", Psych. Rev. 79, 487-509.

Menyuk, P. and M. Klatt (1975): "Voice onset time in consonant cluster production by children and adults", J. Ch. Lang. 2, 223-231.

Naeser, M. A. (1970): "The American child's acquisition of differential vowel duration", Madison, Wisconsin: Technical Report 144.

Ohala, J. J. (1970): "Aspects of the control and production of speech", UCLA Working Papers in Phonetics, 15.

Ohala, J. J. (1975): "The temporal regulation of speech", in Auditory Analysis and Perception of Speech, G. Fant and M. A. A. Tatham (eds.), 431-453, London: Academic Press.

Pisoni, D. B. and J. R. Sawusch (1975): "Some stages of processing in speech perception", in Structure and Process in Speech Perception, A. Cohen and S. G. Nooteboom (eds.), 16-35, Berlin: Springer-Verlag.

Simon, C. and A. J. Fourcin (1978): "Cross-language study of speech-pattern learning", JASA 63, 925-935.

Smith, B. (1978): "Temporal aspects of English speech production: A developmental perspective", JPh. 6, 37-67.

Smith, N. V. (1973): The Acquisition of Phonology: A Case Study, Cambridge, England: The University Press.

Spring, D. R. and P. S. Dale (1977): "Discrimination of linguistic stress in early infancy", JSHR 20, 224-232.

Tingley, B. M. and G. D. Allen (1975): "Development of speech timing control in children", Ch. Dev. 46, 186-194.

Turvey, M. T. (1977): "Preliminaries to a theory of action with reference to vision", in Perceiving, Acting, and Knowing: Toward an Ecological Theory, R. Shaw and J. Bransford (eds.), Hillsdale, N. J.: Lawrence Erlbaum Associates.

Waterson, N. (1970): "Some speech forms of an English child: A phonological study", Transactions of the Philological Society, London: 1-24.

Waterson, N. (1971a): "Child phonology: A prosodic view", JL 7, 179-211.

Waterson, N. (1971b): "Child phonology: A comparative study", Transactions of the Philological Society, London: 34-50.

Zlatin, M. A. and R. A. Koenigsknecht (1975): "Development of the voicing contrast: Perception of stop consonants", JSHR 18, 541-553.

Zlatin, M. A. and R. A. Koenigsknecht (1976): "Development of the voicing contrast: A comparison of voice onset time in stop perception and production", JSHR 19, 93-111.