

On the Development of Motor Control in Speech: Evidence from Studies of Temporal Coordination

SARAH HAWKINS

*Haskins Laboratories
New Haven, Connecticut*

- I. Introduction 318
- II. Prosody in Speech Perception and Production 318
- III. The Contribution of Studies of Temporal Coordination to Understanding the Development of Motor Control in Speech 320
- IV. Some Basic Assumptions about the Development of Speech..... 321
 - A. The Child's Perceptual Abilities 321
 - B. The Child's Productive Abilities 322
- V. Acoustic-Phonetic and Neuromotor Influences on the Development of Temporally Coordinated Speech 324
 - A. Temporal Distinctions That Serve as Primary Perceptual Cues.. 324
 - B. Temporal Regularities That Do Not Function as Primary Perceptual Cues, and the Concept of "Difficulty of Articulation" 327
 - C. Principles Common to the Development of All Motor Skills ... 330
 - D. Summary: General Principles of Motor Skill Learning 344
- VI. Two Strategies for Learning to Produce Clusters of Consonants Fluently..... 344
- VII. Toward a Theory of the Developing Motor Control of Speech 354
 - A. Defining a Motor Skill 354
 - B. The Unit of Analysis: "Top-Down" versus "Bottom-Up" Models 356
 - C. Schema Theory Applied to Speech Development 358
- VIII. Directions for Research 362
- IX. Some Unresolved Issues 365
- References 368

I. INTRODUCTION

The timing and rhythm of speech was, until recently, a neglected area of research, with a few notable but isolated exceptions (e.g., Lehiste, 1970; Allen, 1972, 1973; Klatt, 1973, 1974, 1975, 1976). The last decade or so has seen a change: segmental and syllabic duration has been recognized as being of basic importance to speech intelligibility and as an indicant of the central organization of speech. This change has taken place within the context of increased interest in all prosodic aspects of speech (see Cohen & Nootboom, 1975, and *Phonetica*, 1981, Vols. 1-3). Studies of timing in children's speech have followed a similar pattern: during the last 10 years there has been a slow but steady increase of interest in the area. There is now a relatively large body of data, and we are faced with the question of what to do with it.

One of the most important aspects of research into the timing of children's speech is that it provides a potential field for testing theories of the development of speech as a motor skill. Increasingly mature temporal organization of speech units is assumed to reflect increasing coordination of the motor gestures for speech. An initial aim of this article is to specify some assumptions basic to a theory from which useful models of the developing motor organization of speech can be derived. Some existing data on speech timing will then be examined in terms of these assumptions and a possible theoretical framework will be discussed. Questions raised by this examination will, it is hoped, help to focus future inquiry on those issues most urgently in need of answers.

II. PROSODY IN SPEECH PERCEPTION AND PRODUCTION

Because data on the role of prosody, and especially segmental duration, are not always familiar to specialists in children's speech and language, I summarize here some of the more important points. It has been shown that adult listeners can derive considerable information about the syntax and stress pattern of sentences even when segmental cues are either distorted by spectral rotation (Blessner, 1969) or absent because the sentence is hummed (Svensson, 1974). Other experiments have demonstrated the role of prosody in defining syntactic boundaries by using stylized synthetic intonation contours (e.g., Collier & 't Hart, 1975) or by pitting prosody against syntax in cross-spliced sentences (e.g., Wingfield & Klein, 1971; Darwin, 1975). Durational factors alone can also be crucial to speech intelligibility, as has been shown for both phonemic segments and larger units in real and synthetic speech (cf. Lehiste, 1970; Klatt, 1976; Nootboom, Brox, & De Rooij, 1976; Huggins, 1978). Adults appear to be especially aware of the rhythmic onset of stressed syllables, both when listening to speech (e.g., Huggins, 1972; Cutler & Foss, 1973, 1977; Shields, McHugh, & Martin, 1974;

Cutler, 1976) and when tapping to the rhythm of their own speech (Allen, 1972). These investigators have suggested that the listener appears to anticipate when stresses will occur and may attend more to the signal at these times. This integrative and predictive role of prosodic cues figures prominently in several models of speech perception. For example, Martin (1972) has elaborated on the notion of the predictive role of rhythm in speech perception, pointing out that efficient perceptual strategies such as attention-cycling between processing the signal and other activities could be facilitated when the signal need not be monitored continuously. Pisoni and Sawusch (1975) suggest that prosodic cues may form an interface between low-level segmental information and higher levels of syntax and meaning.

Segmental duration is more central to this article than are intonation and syllabic rhythm. The durations of speech segments corresponding to phones vary with phonetic context according to regular principles. One effect is that just as each syllable tends to become shorter as the number of syllables in a word increases, so we find that the more phonemes in a word, the shorter a given phonetic segment tends to be. Such changes in duration are not uniform across all segments, however, and indeed the durations of some segments do not change in certain phonetic contexts. Which segments are changed in duration, and by how much, depends on the manner and place of articulation of the segment itself and on its surrounding context. For example, /s/ is usually about 15% shorter in prevocalic clusters with /p, t, k/ as in *spin, sting, and skin*, than when it occurs without the stop, as in *sin* or *sing*. When preceded by /s/, the closure interval for the /t/ is also generally shorter than when alone, as in *a sting* versus *a tin*. In contrast, a bilabial stop preceded by /s/ is generally longer than when unclustered (*a spin* versus *a pin*), whereas there is little change in duration for /k/ in /sk/ compared with its unclustered duration (*a skin* versus *a kin*). Such changes have been described for a number of languages (e.g., Lindblom & Rapp, 1971; Haggard, 1973; Nootboom, 1973; Klatt, 1973, 1974, 1975; O'Shaughnessy, 1974; Suen & Beddoes, 1974) and have been summarized for English by Klatt (1976).

We do not know to what extent changes in segmental duration affect speech perception. This is especially true for consonants in languages such as English, which do not make phonemic distinctions based on consonantal length. Many of the rule-governed durational modifications that we observe for consonants are not perceptible by themselves because they are less than the 25 msec estimated as generally necessary for detection (Klatt, 1976), but their effects may be important for perception in interaction with each other or with other prosodic characteristics. Correct segmental duration is essential for naturalness and intelligibility in synthesis-by-rule systems, for example, and studying durational characteristics will presumably shed light on the motor organization of speech. For these reasons, segmental changes must be accounted for as part of any more global explanation of timing.

III. THE CONTRIBUTION OF STUDIES OF TEMPORAL COORDINATION TO UNDERSTANDING THE DEVELOPMENT OF MOTOR CONTROL IN SPEECH

Learning to speak is a task of immense complexity that involves far more than discovering which articulatory movements result in the percept of each phoneme of the language being learned. For example, the child must learn to control her breathing sufficiently to produce the more or less constant subglottal pressure necessary for speech as well as learning the necessary laryngeal adjustments for acceptable phonation. Additionally, the child must learn that consonants and vowels tend to alternate with one another although (in many languages) consonants can occur in clusters. She will learn to produce the suprasegmental constraints of intonation and timing that govern phrases, and if she is learning a language like English, she will learn to intersperse stressed syllables with unstressed syllables. Within words the child must learn not only the order of syllables and segments, but also the details of their temporal control, including coarticulation. The details of how this knowledge is acquired are only partly understood. Some recent approaches to the question have been offered by Kent (1980, 1982) and Fowler, Turvey, and their colleagues (e.g., Fowler, Rubin, Remez, & Turvey, 1980), some of whose ideas are discussed later in this article.

Much of this basic knowledge about speech production is acquired by most children quite early (e.g., Crystal, 1973, 1978; Oller & Smith, 1977; Delack & Fowlow, 1978; Oller, 1980; Stark, 1980) but some is not mastered until relatively late. Among those aspects that appear later in development—after 3 to 4 years of age—are the finer details of phasing or coarticulation of segments as their production becomes refined into the highly skilled routines of the adult speaker (e.g., Hawkins, 1979a; Kent, 1982). One approach to studying these latterly acquired aspects is to regard developing coarticulatory skills as the increasingly fluent coordination in time of the gestures required to produce specific segments, and this constitutes the main focus of the present article. That is, using temporal relationships as an index of increasing motor skill, we are interested in how children learn to coarticulate the gestures for articulatory units in speech.

There are several ways to study developing temporal control, all of which have in common that the criterion of maturity is the adult's speech behavior. For example, children's words, syllables, or segments may be compared with those of adults for evidence of greater temporal variability, longer durations, or differences in the pattern of temporal relationships found between measured units in different phonetic contexts. The changes that we observe in these parameters as the child grows older can indicate the course of the maturational process: whether maturation is gradual and steady, or proceeds in discrete "stages"; whether children's and adults' speech appears to be organized temporally in the same way, or whether there is a qualitative difference between the two, and so on.

Current research into adults' speech has led to some controversy about the role of timing in speech production. Briefly, there are those who believe that timing is an intrinsic aspect of speech production, with the occurrence of at least some events explicitly specified in time by an articulatory "program" (cf. Martin, 1972; Allen, 1973). Others maintain that the temporal relationships that we observe are not programmed as such by the central nervous system. Rather, the brain specifies and commands the execution of articulatory movements without explicit duration; temporal regularities occur simply as a result of the physical and physiological characteristics of the speech apparatus and only exist in time because speech cannot exist independently of time (cf. Fowler, 1980).

This controversy raises important theoretical issues for the study of adults' speech, but the study of developing temporal coordination need not await its resolution. As long as we allow that temporal reorganization must at least **reflect** changes in underlying organization, then duration can be used to study changes in articulatory skill. That is, although we do not fully understand the nature of the articulatory commands, we can still measure their presumed indices, and duration is one such index.

In summary, developmental studies of temporal coordination indicate the types of change that the speech production mechanism undergoes during its maturation toward the adult system. The specific nature of those changes must await our deeper understanding of the motor control of speech in the adult. In this article, increasingly mature temporal coordination is assumed to reflect increasingly fine motor skill in speaking and to imply greater coarticulatory skill.

IV. SOME BASIC ASSUMPTIONS ABOUT THE DEVELOPMENT OF SPEECH

Rather than turning immediately to the data on the development of temporal coordination, let us first step back a little to consider the background to any investigation of speech development. An account of the developing motor control of speech must consider the child's perceptual development as well as her motor skills. In this section of the article, I set forth assumptions concerning both perceptual and productive abilities of the child that lie behind my interpretation of the data presented in subsequent sections.

A. The Child's Perceptual Abilities

I make the following four assumptions about the child's perceptual abilities when she begins to speak. First, I assume that the adult listener need attend only to some of the properties in the acoustic signal and will often "fill in" the rest on the basis of knowledge shared with the speaker. But the young child lacks the linguistic and nonlinguistic experience that would allow her to fill in, i.e., to

minimize the attention paid to some acoustic properties. Although she may be as adept as are adults at filling in the message from many nonlinguistic contextual cues, her sensitivity to linguistic contextual cues must be relatively undeveloped as yet.

Second, I assume that the child's perception of speech neither is absolutely mature before she begins to speak (cf. Smith, 1973) nor develops in parallel with production (cf. Waterson 1970, 1971a,b). Rather, I assume that although the young speaker responds to some properties in the speech signal as the adult does, she fails to respond distinctively to other properties, i.e., she treats those portions containing these latter properties as only partly analyzed "noise." This view is similar to that of Ingram (1974), except that I assume that the position of the noise may not be heard in a single discrete region corresponding to a segment within the signal. It may extend over several segments which may vary in the detail of their specifications; or it may be loosely associated with the word's overall pattern without yet being tied down to a particular position. Wilbur (1981) supports this view. (The definition of a "word" for the young speaker/listener is itself problematical. Discussion of this topic falls outside the scope of this article, in which "word" is defined rather loosely as any consistent phonological pattern that the child uses with respect to a consistent referent or occurrence.)

Third, I assume that of those aspects of the speech signal that are maturely **discriminated**, only some will be **interpreted** in the adult way. That is, there will be some distinctions that the child can hear but does not yet identify as phonemically relevant. For example, although infants of less than 16 weeks can discriminate between stimuli differing in some durational aspects such as voice-onset time (VOT) (e.g., Eimas, Siquelund, Jusczyk & Vigorito, 1971) and syllable duration (Spring & Dale, 1977), children as old as 3-6 years do not necessarily use these durational cues in the same way as do adults (Zlatin & Koenigsknecht, 1975; Higgs & Hodson, 1978; Simon & Fourcin, 1978; Bailey & Haggard, 1980; Greenlee, 1980).

Last, I assume that at any point in time the young child's systems of speech, language, and cognition are best described as quasi-independent subsystems, any or all of which may be in a state of flux. Speculation about the interrelationships between these subsystems falls outside the scope of this article, but an awareness that the subsystems may influence one another must be incorporated in a theory of speech production.

B. The Child's Productive Abilities

I make three basic assumptions about processes underlying the child's production. The first of these is 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 children's speech. What distinguishes the two is the domain of influence of each process. For example, at any given speaking rate we might find evidence in adult speech that the duration of each syllable was independent of the durations of neighboring syllables, whereas the durations of phonemes within a single syllable were not independent in that an unusually long phoneme would be followed by a correspondingly shorter one, and vice versa. Such a situation could be interpreted as evidence for a hierarchical integration of phonemes into syllables, as in a "comb" or open-loop model, with the timing of the syllables themselves organized relatively independently of each other, as in a "chain" or closed-loop model. (For a discussion of these models, see Kozhevnikov & Chistovich, 1965; Bernstein, 1967; Ohala, 1975.) In the young child's speech, we might find that hierarchical integration indicative of a comb model may not yet be evident at the full syllabic level, but that similar integration occurred with less complex units, such as a vowel and following nasal stop. As a different example, consider pronunciation of the word /bæŋz/ (*bangs*). The adult produces this word with the segments coarticulated. We might find that the child, on the other hand, coarticulates /b/ with /æ/ and /æ/ with /ŋ/; but /ŋ/ itself might not be fronted in preparation for the following /z/, and the syllabic nucleus might not be shortened in preparation for the addition of /z/. In this hypothetical example, the child would be coarticulating some of the elements but failing to produce the whole word as an integrated unit in the way that the adult does. These two examples illustrate, then, 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 complex domains, thereby integrating the elements of those domains into functional units. These ideas are discussed further toward the end of this article.

My second assumption is that processes of production may interact with one another so that an observed parameter may have more than one underlying cause. Two or more processes (or rules) may have opposing influences on a given parameter. Klatt (1973) has reported such cases for factors affecting the duration of consonants in clusters, depending on the manner or place of articulation of the consonants involved. Presumably there may be occasions when such opposing influences exactly match each other so that there is no measured change in the parameter being observed. This, of course, can be true for the adult as well as for the child, but the implications of the point may be more important for the child, because at any given time she is likely to possess rules that are well established but relatively immature, and other rules that are part of a newer system that will eventually replace an older one. Many such rules may well be in conflict with one another. The implication is that, particularly for children's speech, small changes may be small because some factor is opposing the main direction of change rather than because the main factor has only a weak effect. Similarly, absence of change may not necessarily indicate complete absence of the processes required to produce that change.

Third, I assume that the role played by different processes may change during maturation. For example, delayed auditory feedback has a qualitatively different effect on children's than on adults' speech (Fry, 1966; MacKay, 1967); this suggests that normal auditory feedback may also play a different role for children.

Given these basic assumptions, what factors are likely to govern whether or not a child will observe a particular rule of timing in her speech? One might wish to argue that rules of speech timing appear simply as a consequence of increasing neuromuscular coordination and only gradually come to serve a perceptual function. On the other hand, one could argue that the child begins to produce mature temporal relationships between segments only after discovering that they aid perception. In the latter case, the age when adult timing relationships appear would depend partly on neuromuscular abilities and partly on the age when their perceptual cueing function is recognized. I suggest that both the neuromuscular and the perceptual point of view have merit but that they apply to different situations.

V. ACOUSTIC-PHONETIC AND NEUROMOTOR INFLUENCES ON THE DEVELOPMENT OF TEMPORALLY COORDINATED SPEECH

Bearing the above assumptions in mind, let us consider some of the different factors that are likely to affect the process of learning to speak, together with examples of evidence for their existence within speech timing. Three kinds of factors will be distinguished, of which only the first two are mutually exclusive. These are (1) temporal distinctions that serve as primary perceptual cues; (2) temporal regularities that do not function as primary perceptual cues; and (3) processes common to all motor development. This section attempts to identify in studies of the development of speech timing fairly "pure" examples of each of these three factors and discusses their implications. A primary perceptual cue is defined as an acoustic cue that is usually present and has a major influence in the perception of a particular phone or phonetic sequence. A secondary perceptual cue often accompanies a primary cue and can enhance perception, or under adverse listening conditions it could in fact become essential for correct perception.

A. Temporal Distinctions That Serve as Primary Perceptual Cues

Temporal distinctions that serve as primary perceptual cues are likely to be detected by the normal child relatively early as long as they do not signal semantic distinctions that are beyond the child's comprehension. Hence they

should appear in the child's speech in an order reflecting the degree of neuromotor coordination required, but those that require great neuromotor coordination may be marked in a nonadult way. A well-documented example is the development of VOT in stops.

The development of the voicing contrast in English has been studied longitudinally (e.g., Kewley-Port & Preston, 1974; Macken & Barton, 1979), and cross-sectionally (e.g., Menyuk & Klatt, 1975; Zlatin & Koenigsnecht, 1976; Gilbert, 1977; Barton & Macken, 1980). The contrast has also been studied in other languages, notably Spanish (Macken & Barton, 1980), Cantonese (Clum-eck, Barton, Macken, & Huntingdon, 1979), and Danish (Jørgensen, 1979). Except for Spanish, it has consistently been found that by about 2 years of age children are making a distinction between prevocalic short-lag and long-lag stops, but that only the short-lag distribution resembles the adult form. It is not until much later, about 6 years of age or more, that the long-lag VOT distribution reliably resembles that of the adults.

The late establishment of mature VOT for long-lag stops has been commonly accepted as resulting from differences in the neuromuscular coordination required: short-lag stops are thought to allow considerable variability in the coordination of laryngeal and oral activity, whereas long-lag stops demand rather precise coordination (cf. Kewley-Port & Preston, 1974). In a review of the literature on VOT for both adults and children, Cooper (1977) suggested two possible explanations for the late acquisition of the long-lag category of stops. The first, due originally to Klatt (1973), is that children learn to adduct their vocal folds ready for voicing upon receiving peripheral feedback from the drop in pressure that occurs when a stop is released. The time for the feedback loop to be completed is thought to be at least 40 msec, which would consequently produce a long-lag stop. If we assume that use of such closed-loop feedback is slow in children as well as a temporary stage prior to establishing central (or open-loop) control of VOT, then we also have a possible explanation for the overlong VOT categories observed by Menyuk and Klatt (1975) and by Barton and Macken (1980), after the short- versus long-lag distinction is produced but before the long-lag distribution is fully adult-like. The second explanation that Cooper (1977) put forward comes from his and his colleagues' work with adults, which led to the postulation of a central perceptuomotor processor controlling long-lag but not short-lag VOT in stops. The longer maturation period typically required for the development of perceptuomotor skills may explain the longer time required to establish a long-lag category of VOT compared with the short-lag category.

Jørgensen's (1979) Danish data differ from those of the other VOT studies in that his short-lag distribution covers as much as 50 msec VOT, whereas the other studies found much shorter times. Jørgensen argued convincingly that this finding casts doubt on the validity of the interpretation of the earlier data in terms of

differences in the degree of neuromuscular coordination required. It also casts some doubt on the pressure-feedback hypothesis, although Cooper's hypothesis of late development of perceptuomotor control remains unaffected. Whatever the explanation, for our purposes the point remains that children make this distinction in VOT early in their speech development, although the way they make it differs from the way adults make it.

This point is further underlined by Macken and Barton's (1980) study of children learning Spanish. They found that even children of almost 4 years of age were not consistently distinguishing between voiced and voiceless stops in their speech on the basis of VOT. They did, however, show that the children were distinguishing between /b, d, g/ and /p, t, k/ by spirantizing the "voiced" but not the "voiceless" stops, as indeed the adults in their environment tended to do. In addition to casting doubt on traditional analyses of Spanish in which spirantization of voiced stops is considered a secondary process overlaid on the basic voiced-voiceless distinction, this study raises crucial issues about the relationship between phonological and phonetic development.

The development of VOT control in stops nicely illustrates the following points: (1) perception and production do not always develop hand in hand; and (2) a phonetic distinction that may legitimately be regarded as lying along a single phonological dimension should not necessarily be treated as lying along a single dimension in terms of motor programming. That is, such a distinction may not necessarily represent two extremes of a single process in terms of the motor activity required to produce it, in which case the distinction should not necessarily be treated as a unitary process in a theory of speech development.

A second example of the early acquisition of a temporal distinction that serves as a primary perceptual cue is that of phonemically conditioned vowel duration in English. Vowel duration functions in English as a 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, including English (Lisker, 1974). Such evidence would suggest that the child might learn relatively early to produce vowel durations in the correct ratios according to voiced or voiceless consonantal context. Naeser (1970) found that these correct ratios were present by 21 months of age and in fact preceded control of the consonantal voicing feature that governs the distinction in adult speech.

These data illustrate the difficulties in trying to decide whether a child is perceiving a phonemic distinction as the adult does, and if so, whether the child can be considered as marking that phonemic distinction in her own speech. In Naeser's case, the problem is that for the child the phonemic contrast may rest upon differences in vowel length rather than in the voicing of the final stop. The child's discrimination and identification of this contrast would then be faultless, but the phonological basis of her underlying distinction would be quite different

from the adult's. In fact, a perceptual study by Greenlee (1980) suggests that (at least in the absence of all other cues) small children do not use differences in vowel length to distinguish between word-final voiced and voiceless stops, even though they themselves produce this distinction. In addition to answering the specific question raised by Naeser's study, Greenlee's data illustrate further the "complex and somewhat paradoxical relationship between developing production and perception" (p. 459).

B. Temporal Regularities That Do Not Function as Primary Perceptual Cues, and the Concept of "Difficulty of Articulation"

Temporal regularities that do not function as primary perceptual cues, especially those that appear to provide no perceptual information at all, would be expected to be acquired as the child's articulatory abilities become more sophisticated. We would expect those distinctions to appear in an order reflecting the neuromuscular complexity involved, and we would expect many of them to appear later than distinctions that reflect primary perceptual cues. An example is 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 typically produce these durational modifications varies according to the type of cluster, but most are not fully developed by 5 years of age, and the last ones are probably not mastered until as late as 9 to 11 years of age (Gilbert & Purves, 1977).

The question of whether these durational modifications appear in an order reflecting the degree of neuromuscular complexity is harder to answer. Even in the relatively objective area of segmental duration, the inference of complexity, or difficulty, tends to involve circular reasoning: a durational difference between adults and children that is greater than average is labeled as involving a more difficult articulation for the child. It seems reasonable to allow that where an especially large difference between adults and children is reliably found, **something** is more difficult than usual for the child. But this in itself does not tell us whether the difficulty is physiological, cognitive, or perceptual in origin. If it is perceptual, it does not tell us whether there is a perceptual deficit on the child's part or an effort by the child to increase a perceptual distinction for the benefit of the listener. In the latter case, the perceptual problem may arise from a deeper cognitive problem in that the child may not be aware of which properties can be omitted from or distorted in the speech signal and which must be maintained.

Moreover, in labeling some articulation as "more difficult than" another, it is not enough simply to assess the degree of difference between adults and children for the two articulations concerned. The quality of the difference must also be

considered. For example, although insufficient shortening of a consonant in a cluster may have any of several causes, the most likely is neuromotor inability to produce the consonant and its transitions fast enough. Lengthening of a clustered consonant, on the other hand, although it may have a physiological cause, might be just as likely to stem from the imposition of a different type of organization over that phonetic sequence and could be unrelated to the particular consonant in question.

What solution can be found to this problem of defining "difficulty of articulation"? One possibility is to use norms for age-of-acquisition of particular phonemes. This could certainly be helpful in some cases. MacNeilage, Hutchinson, and Lasater (1981) recently suggested that articulatory difficulty might best be defined in terms of linguistic markedness, which was itself originally defined largely in terms of sound preferences in infant babbling. But even if age-of-acquisition norms or a hierarchy of markedness failed to support some inference of differential complexity, the existence of that complexity would not necessarily be disproved. This point would be particularly true when considering fine details of control, as in temporal coordination of consonants in clusters. Control of such phonetic detail could involve quite different motor skills from those required to produce phonemes acceptable to even the most careful listener. In their discussion of difficulty, MacNeilage *et al.* (1981) suggested a possible approach to this last problem. They noted that one tenet of Fitts' law (Fitts, 1954) is that smaller targets are approached by slower movements. If we define "smaller" as requiring more precision, then we should be able to set up a hierarchy of phonemes in terms of necessary articulatory precision. There is some supportive evidence for the validity of this method in that movements toward fricatives require more precise positioning of the articulators than do stops (Kuehn & Moll, 1976).

In addition to the problem of defining and studying **motor** difficulty without circularity, it is not easy to justify excluding perceptual deficits as a reason for adult-child differences, at least in the absence of good perceptual experiments (and these are notoriously difficult to do with young children). For example, although infants and young children are sensitive to a wide range of basic prosodic contrasts (Crystal, 1973; Kuhl, 1979) it is not until they are at least 9 or 10 years old that they become aware of some of the more subtle aspects of intonation (cf. Cruttenden, 1974). Some of these later acquired skills demand a high degree of cognitive maturity. For example, the understanding and use of sarcasm, which is generally learned rather late, requires knowing that A can mean **not-A** and involves subtle situational as well as prosodic cues. We know little about the cognitive prerequisites for such sophisticated aspects of speech production, but it has been suggested that in production, increased syntactic and/or semantic complexity may disturb the prosody of a child's speech (Allen & Hawkins, 1980). It is inadvisable, then, to ignore the possible influence of perceptual or cognitive immaturity on production.

So far, this discussion of "difficulty" in learning how to talk has concentrated mainly on matters related to motor control and timing. The question has much more general relevance, however, and indeed raises issues central to all studies of the acquisition of speech and language. Two of these will be mentioned here. First, what is a simple unit for the child? It is reasonable to assume that the number of phonemes in a unit and its complexity are not in a one-to-one relationship, although some correspondence would be expected. For example, saying the syllable [ba] may be less effortful for the child (or for some children) than saying the single phone [s] or [ɹ]. So we have the question of how to identify a complex or a simple unit, and this takes us back to the circularity mentioned earlier. Second, running through any discussion of complexity is the issue of individual variability: what is difficult for one child may not be so for another. This is true even when the cause of "difficulty" can be attributed to physiological immaturity. Children mature at different rates and in slightly different sequences. Moreover, considering motor control again, there is often more than one way of achieving the same acoustic result. It is possible that, given two children with equal physiological maturity and ability, one child may discover a way that is easy for her, while the other may not have discovered this way, but may instead persist in trying with a way that is more difficult for her. Distinguishing group trends among the individual variation and identifying consistent individual differences from random fluctuation are problems in their own right, quite apart from being obstacles to the identification of "difficult" units. The problem is compounded in speech research, where numbers of subjects (and often numbers of tokens) tend to be small.

To conclude and to bring our focus back to the motor control of speech, we do not have a way of using the term difficulty in an explanatory way with respect to children's speech. Even assuming we can keep the cognitive and perceptual requirements of a task constant, final decisions as to the degree of neuromotor maturity involved in different situations must await our deeper understanding of the physiology of speech in the adult and in the developing child, as well as our increased ability to distinguish general tendencies amongst individual variation. Hence, if distinction *b* appears in a child's speech later than distinction *a*, we may speculate that *b* requires greater neuromuscular maturity than *a*, but we should not claim that we have unequivocal evidence that this is the case.

At the beginning of this section I suggested that distinctions that do not serve as major perceptual cues should show an orderly progression toward their mature adult forms reflecting only the degree of muscular coordination required to produce them. But this is a "long-term" view. In the short term, we may find that some temporal patterns resemble the adult form less than they did at an earlier stage of development. This problem of long-term versus short-term variation occurs in all types of learning. It has been directly addressed for phonology by Menn (1979). Similarly, Ingram (1979a) discusses variation in phonological

form due to specific characteristics of the language that the child is learning. One way to try to disentangle the various sources of variability may be to identify particular processes that are known to affect the course of learning, as discussed in the next section.

C. Principles Common to the Development of All Motor Skills

All aspects of speech development will be subject to general principles of motor skill learning, so we should see the influence of these general principles affecting development of all temporal distinctions, whether primary perceptual cues or not. The following five principles are some of the more common and easily identifiable ones for which we would expect to be able to find evidence: (1) slower and more variable performance would be expected from children compared with adults, as would the opposite, (2) more stereotyped behavior, which can result from a failure to differentiate between the various contexts in which an event may occur. (The apparent contradiction between 1 and 2 is resolved by distinguishing between tokens and types. For example, one would expect variability from the child in repeated tokens of the same phonemic sequence, but less differentiation between different types of sequence.) The third principle, (3) a gradual refinement of performance towards the adult norm, may be modified either by (4) overgeneralization of recently acquired rules, or by (5) periods of rapid change alternating with relatively quiescent phases.

Good cross-sectional studies can provide evidence for the third principle (gradual refinement), but the last two principles can only be unambiguously identified in longitudinal studies. Overgeneralization of rules can lead to stereotyped behavior but must be distinguished from an **initial** failure to differentiate between contexts; one way to make this distinction is to look for cases of regression in superficial maturity of performance. Finally, in order to distinguish unequivocally between individual differences in rate of maturation and the alternation of spurts of growth with periods of relative consolidation, each child's performance must be followed for some time, probably for at least 1 or 2 years.

Unfortunately, we lack the data that would allow us to identify some of these principles in the development of temporal coordination, especially those requiring detailed longitudinal study. The only published longitudinal study of which I know is my own (Hawkins, 1979a), but this took data from only two time periods separated by 14 months, and only six children were studied at both times. There are several cross-sectional studies, but usually the groups of children are separated in age by about 2 years, which is too great a time for many of these processes to show up (Eguchi & Hirsch, 1969; DiSimoni, 1974a,b; Gilbert & Purves, 1977; Smith, 1978; Kent & Forner, 1980). All these studies analyzed children's speech by either oscillograph or spectrograph, with some effort to

control for such variables as rate of speech and phonemic context. In most cases, the speech was elicited in fairly artificial situations. As a group, the studies cover the age range 3 to 12 years fairly well, but because they differ in type of material and in measurements made their information is not always comparable.

For example, several of the above studies measured durations in real words with constant stress patterns, whereas Kent and Forner (1980) used meaningful sentences but measured segments from words differing in stress, and Smith (1978) used nonsense words with stress as an independent variable. Again, some studies concentrated fairly exclusively on one aspect of temporal control while others examined several aspects, each in less detail. Thus, Gilbert and Purves (1977) and Hawkins (1979a) report data on duration of consonants in singleton and clustered word-initial contexts only, whereas Kent and Forner (1980) measured durations of selected stop closures, VOTs, other consonants, vowels, whole words, and phrases of varying length.

Each of these studies contributes to our understanding of the developing control of timing in different ways, but their differences often make direct comparison and synthesis of results quite difficult. In the following discussion, we will focus mainly on two studies, adding points from others where possible. These two studies, Gilbert and Purves (1977) and Hawkins (1979a), used more similar material than other studies and are complementary; one study (Hawkins) is longitudinal and comprehensive but with a small number of subjects spanning a relatively narrow age range (six children aged 4–8 years), and the other (Gilbert and Purves) involves less speech material but is a cross-sectional study with more subjects over a wider age range (five subjects per group, ages 5, 7, 9, and 11 years). These two studies are described in more detail before their results are discussed in terms of the above five general principles of motor skill learning.

Both investigations compared the durations of word-initial consonants as singletons and in clusters. [As noted earlier, in most but not all contexts adults reduce the durations of clustered consonants. Whether a consonant is shortened, and to what extent it is shortened, depends on the consonant itself and on the other consonants in the cluster (Haggard, 1973; Klatt, 1973; O'Shaughnessy, 1974). The intricate pattern of durational modifications that the adult produces provides a rich but complex base against which to evaluate the development of skilled articulation in children who are old enough to provide unambiguous and high-quality speech samples.] Both Hawkins (1979a) and Gilbert and Purves (1977) studied consonantal duration in meaningful monosyllables spoken by monolingual English-speaking children. (Gilbert and Purves' subjects were Canadian and Hawkins' were British.) The words made up sets such as /spn/, with /sun/ and /pʌn/, or /klin/, with /kin/ and /lin/. That is, for each word with an initial consonant cluster, there were others with the same vowel but only one element of the cluster. Gilbert and Purves studied /fl, sl, sw/. Hawkins included all initial consonant clusters of English except /θr, θw, tw, dw, kw/ (and /vr/

and /sf/ if these are counted as English clusters). In both investigations, segmental durations were measured oscillographically and compared with data collected under similar conditions from five adults. A possible source of difference between the two studies is the carrier phrase used: "*repeat* _____" by Gilbert and Purves, and "*a* _____" by Hawkins. This is unlikely to affect points made in the following discussion, although some measurement problems associated with the word *repeat* are mentioned in Hawkins (1979a).

Neither Gilbert and Purves' nor Hawkins' data can offer watertight proof for or against any hypothesis requiring longitudinal evidence, but they can provide some indication of whether or not such an hypothesis is worth pursuing. Nevertheless, most of the support for the existence of the three principles of motor skill learning that require longitudinal data comes from these two studies (i.e., gradual refinement, overgeneralization, and alternation of rapid change and quiescent phases); evidence for the first two principles (variability and stereotypy) is not restricted to these studies.

1. *Slower and More Variable Performance*

The first general principle, slower and more variable performance by children compared with adults, has been demonstrated for speech in many studies and at many levels of analysis, from the phrase to the segment. Both inter- and intrasubject variability are greater for children (e.g., Eguchi & Hirsh, 1969; DiSimoni, 1974a,b; Tingley & Allen, 1975; Smith, 1978; Hawkins, 1979a,b; Kent & Forner, 1979, 1980; Kubaska & Keating, 1981). It is worth considering the significance of both sources of variation. There is no question that neuromuscular constraints are an important cause of the child's failure to achieve adult values. There are probably additional causes, however. Consider the longer absolute durations of children's speech. Most studies agree that the child must learn to shorten rather than to lengthen articulatory units in order to produce phonological length distinctions (Oller & Smith, 1977; Smith, 1978; Allen & Hawkins, 1980). Even in such superficially simple cases as longer duration, disproportionate slowing sometimes occurs, strongly suggesting the involvement of additional causes in these instances. For example, measuring /b, d, t/ closure 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 on the basis of the estimated durational increase that he attributed to neuromotor immaturity. [The estimates came from the average difference in (nonsense) word durations between adults and each age group of children.] In addition to a general physiological factor, Smith suggested two possible causes for this disproportionate lengthening, either or both of which may have contributed to the observed effect: first, an effort to increase the perceptual difference between /t/ and /d/, and second, greater complexity of the laryngeal adjustments required for voiceless stops over voiced ones.

Similar arguments for multiple causality may be raised for children's greater variability. Although greater variability in performance is such an integral part of developmental studies that failure to find it could render one's data suspect, implicit within this "fact of life" lies a theoretically significant question. That is, does the variability that we see in young children's speech stem solely from neuromuscular immaturity (i.e., motor **inability**) or does the child's **target area** (or range of acceptable articulations for each sound) differ from the adult's? We would naturally find greater variability in the child's speech if her physical target for a given sound covered a wider articulatory area than that of the adult. Such differences in target area might be caused, for example, by anatomical differences in the vocal tracts of adults and children, or by a different perceptual criterion of adequacy. Whatever the cause of the difference, it is easy to see how part of learning to speak could involve differentiating between and refining the articulatory targets for different phonemes. Extending this notion to the development of temporal control is similarly worth considering.

One way to see if children might have targets different from those of adults is to look at the variance distributions. The shapes of the distributions of relative variance (variance divided by the mean) for all consonants in all contexts differed dramatically between Hawkins' (1979a) first and second studies. In the first study, the distribution of relative variances of all clusters (with children pooled) was as much rectangular as it was normal, whereas 14 months later the distribution was much nearer to normal, although it was strongly positively skewed and the range of relative variance was almost as great as in the earlier study. Kent and Forner (1980) found similar trends for 4, 6, and 12 year olds, compared with adults; the distributions of the 12 year olds were frequently close to those of the adults. This presumably gradual progression from relatively random to much more ordered values strongly suggests that the main cause for the children's variability lies in motor immaturity. If the child's target simply had a larger area and there was no motor immaturity, we would still expect a normal distribution; a rectangular distribution implies inability to achieve the target. In the case of timing, however, there is another possible explanation whose plausibility depends partly on the status accorded to timing in a theory of the motor production of speech. That is, duration must be assumed to be a parameter of speech control (cf. Allen, 1975, versus Fowler, 1980, and Fowler *et al.*, 1980). The following discussion makes no strong claims about the units of temporal control, although the syllable and its components are assumed to be important.

Ohala (1970) suggested that the English-speaking adult has a timing-dominant rather than an articulation-dominant system of articulation. He proposed that the adult adheres to a relatively strictly timed articulatory program at the expense of precision of articulation. For example, speakers will produce vowels with "undershoot" to avoid disrupting the overall rhythm of the utterance. There must be some control on the limits of either articulation or timing dominance (articulation

must be clear enough to be understood and general rhythmic constraints must be observed); and it is possible that different types of utterance might particularly favor one or another system of dominance. Longer utterances, for example, might show timing dominance, whereas those spoken in a minimal and repeated context might favor articulation dominance (e.g., the *a-* and *repeat-* of Hawkins' and of Gilbert and Purves' studies).

Rather than a steady progression from articulation dominance to timing dominance, the relative dominance of articulation and timing may change several times as the child's speech production matures. There is some evidence that temporal regularity may predominate over articulatory targets in the earliest stages of learning to speak. Articulatory targets may have dominance in older children, with timing being reestablished as the dominant system by adulthood. Factors that appear to influence which system predominates are whether the utterance is meaningful or babbled and whether the child has productive control over the phonotactics of the language. Owing to the difficulty of defining and measuring speech rhythm (Allen & Hawkins, 1980), most of the evidence is unfortunately somewhat indirect. The strongest support for changes from timing to articulation dominance in the youngest speakers is an analysis by de Boysson-Bardies, Bacri, Sagart, and Poizat (1981) of the utterances of a French child between the ages of 1;6 and 1;8 who was in the "pivotal period between babbling and the production of meaningful utterances" (p. 525). The investigators found that the child's production of shorter utterances (two and three syllables) was divisible into two stages, with the later stage characterized by longer syllable duration and greater variability in duration from one syllable to another. No changes in syllable rate or variability were observed for longer utterances (four or more syllables) during this time period. These data were interpreted by de Boysson-Bardies *et al.* as indicating that the child's "pure" babbling was organized according to a rigid temporal structure, with a basic preprogrammed intonation contour rhythmically punctuated by syllables of relatively fixed duration. Segmental articulations within these babbled syllables were not related to meaning in the adult language and so could adhere to fairly simple standardized patterns that required a minimum of coarticulatory skills. In the later stage, the child was beginning to make meaningful utterances and therefore had to try to satisfy particular articulatory targets and follow the phonological constraints of French more closely. The result was a departure from strict temporal patterning in the interests of articulatory precision, that is, subordination of the temporal program to an articulatory one. Hence average syllable duration and variability both increased. These changes were observed only for shorter utterances because it was in these that the child began to express meaning; his longer utterances continued to be "pure" babble.

A study by Kirk (1973) also suggests that children's earliest meaningful utterances are produced in accordance with strict temporal constraints before articula-

tion is consistent. Kirk showed that in spontaneously imitated utterances spoken by four children aged between 2 and 3 years who were learning the Ghanaian language Gã, both the correct number of syllables and correct contrasts in vowel duration were controlled before segmental contrasts (and after tonal ones). In other words, not only were the children sensitive to temporal constraints, but they were able to produce them before they could produce the associated segments with complete accuracy.

Some studies of children learning English provide more circumstantial evidence than did Kirk's that basic rhythmic constraints in terms of duration are controlled from an early age. Naeser (1970) found that 2 and 3 year olds controlled distinctions in vowel length before they could produce the voiced-voiceless consonant distinction that conditions it. Kubaska and Keating (1981) showed that children shortened nonfinal word durations from the time they began to use two or more words together. In another study of word duration as a function of utterance length, Branigan (1979) measured word duration in three types of spontaneous utterance: single-word utterances, multiple-word utterances, and successive single-word utterances. Utterance types were distinguished in terms of duration of the pause between successive words. In single-word utterances the words were separated by 1 second or longer; multiple-word utterances had pauses of less than 400 msec between successive words; and words in successive single-word utterances were separated by pauses of between 400 and 1000 msec. The subjects were three children in whose spontaneous speech all three types of utterance cooccurred. Comparing single-word utterances with nonfinal words in the other two utterance types, Branigan found that single words were generally longest and words in multiple-word utterances were generally shortest. These observations are consistent with the usual adult pattern of word and syllable durations being successively shorter as utterance length increases. Especially interesting is the finding that words were also somewhat shorter in successive single-word utterances than in isolated words. That is, a temporal constraint on word duration seemed to be operating even when the words were separated by 400-1000 msec (but still appeared to be part of the same conceptual structure).

In addition to following general patterns of word duration, children reproduce in their speech many temporal characteristics of word and sentence stress. For example, Smith (1978) showed that 2 year olds reliably produce the appropriate relative durations of syllables and segments in two-syllable nonsense words varying in stress pattern. Finally, Hawkins and Allen (1978) showed that by the time children are 3 years old they control a complex system of durational and intonational contrasts that follows the adult pattern reasonably closely, while not being identical to it in all respects.

There are even fewer studies of older children that speak to the issue of articulation dominance versus timing dominance. On the basis of their studies of consonantal duration, however, both Hawkins (1973) and Gilbert and Purves

(1977) have suggested that the older child may possess an articulation-dominant system rather than the timing-dominant system of the adult. In other words, it is suggested that the child more than the adult may sacrifice timing demands in order to achieve "satisfactory" articulatory targets. Gilbert and Purves (1977) suggest that the change from an articulation-dominant to a timing-dominant system may happen sometime between 7 and 9 years of age.

These studies together indicate that we might expect several changes in emphasis between temporal and articulatory constraints as the child grows up. The study by de Boysson-Bardies *et al.* clearly suggests a change from temporal to articulatory dominance as babble is replaced by meaningful utterances. Additionally, it illustrates the coexistence of different systems in various stages of development (in this case meaningful speech and babble), each exerting its own influence on the same surface events (syllable duration). Other studies show that many constraints on word and syllable duration are observed when the child is between the two-word stage and about 3 years of age, when most phonemes are mastered but few complex sequences (such as consonant clusters) are produced. Studies of segmental duration in sequences of consonants, most of which are learned relatively late, suggest that at least in consonant clusters accuracy of articulation may again override temporal constraints as new skills are added to the articulatory repertoire. However, most of these data offer only circumstantial support to the hypothesis that an articulation-dominant system replaces an earlier timing-dominant one as the child begins to complete her phonemic and phonotactic inventory. In order to test the hypothesis directly, we need to make detailed studies of temporal and coarticulatory aspects of speech in interaction with changes in segmental phonetics in individual children's speech. One approach to this task might be to follow children learning languages with contrastive vowel or consonant length to see if there is any disruption in segmental or syllabic durations and variances at the times when the segmental inventory is added to or completed.

If children in the intermediate stages of learning to speak did prove to have a more articulation-dominant system than that of adults, neuromotor immaturity would still be indirectly responsible for differences in duration and variability: the child could not always obey both articulatory and temporal constraints. The emphasis towards articulatory accuracy (after a possibly greater dominance of temporal constraints) is presumably at least partly due to the child's increasing need to be understood by people outside her immediate circle, especially as her lexicon increases and she begins to comment on things other than concrete events and objects in her immediate environment. The interaction between temporal and articulatory systems during development would make a fascinating but highly complicated study. Following groups of children classified as "analytical" versus "Gestalt" (Peters, 1977) would be especially interesting and could shed light on the thorny problem of the status of individual variation, although we would

still have to account for the more common child who fits neither of these two extremes (Branigan & Stokes, 1981).

2. Stereotyped Behavior

The second general principle common to the development of all motor skills is stereotyped performance. The child's system may not allow her to differentiate between contexts, or she may overgeneralize a rule to inappropriate contexts. In Piagetian terms, such rigidity results from an imbalance between the complementary properties of assimilation and accommodation: there is too little accommodation to the environment and too much assimilation of new information to the established behavioral patterns of the organism. Although productive or perceptual factors may cause a failure to differentiate, we would not expect perceptual factors to be important for segmental timing with children aged 4 years or more.

Hawkins (1979a) showed evidence for failure to differentiate between contexts both for phonemic and subphonemic distinctions. An example of failure to differentiate between the duration of a segment in various contexts is provided by /s/ in different clusters. In adult speech, /s/ is shortest in clusters including /p/ (/spl, spr, sp/) and in /str/. It is longest in /sw, sl/; /st/ and clusters with /k/ (/skr, skw, sk/) are associated with intermediate durations for /s/ (Klatt, 1973; Hawkins, 1979a). These relationships are illustrated in Fig. 1. The clusters are arranged from left to right in order of degree of abbreviation of clustered /s/. Figure 1 also shows durations of /s/ in clusters spoken by children. In addition to not following the same pattern as adults between clusters, the children's clustered /s/ durations cover a smaller range of values across all contexts than that of the adults. Much of this failure to differentiate appeared to be due to a failure to reduce the duration of /s/ sufficiently when followed by /p/. A second influencing factor was a relative failure to abbreviate in homorganic clusters (i.e., when followed by /t/).

Similarly, at the subphonemic level, Hawkins (1979a,b) found that adults and children differed in the relative duration of VOT in stop-liquid clusters in a way suggesting somewhat stereotyped durations for the children. For example, Fig. 2 shows that for adults the range of mean VOT is greater for /p, t, k/ before /r/ than when these stops are not clustered, as would be expected because each stop's VOT is longer in the clustered context. (The ranges are 59 msec for clustered contexts and 35 msec for unclustered contexts.) Children, on the other hand, had a **smaller** range of VOT for /p, t, k/ before /r/ than when the stops were unclustered, even though, like the adults, they increased mean VOT for each clustered stop. (Their range of VOT was 36 msec for clustered stops and 41 msec for unclustered stops.) That is, a 6-msec difference between adults and children in the unclustered case rose to a 23-msec difference in the clustered case, with the children atypically producing the **smaller** range in clusters, despite

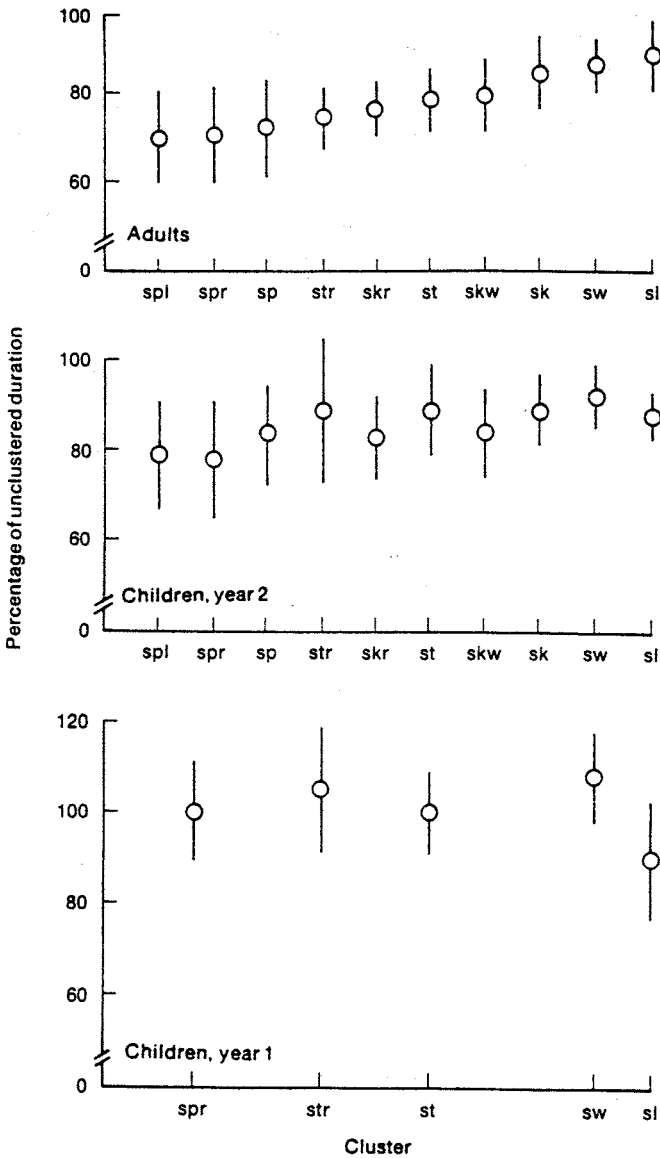


Figure 1. Means and standard deviations of /s/ in various clusters expressed as a percentage of the duration of unclustered /s/. Circles represent means and bars represent standard deviations. Clusters are ordered from left to right in terms of most to least shortened in the adults' speech. Children tended to abbreviate proportionately less than adults, and the children's pattern of abbreviation of /s/ across clusters differed from the adults'. (From Hawkins, 1979a.)

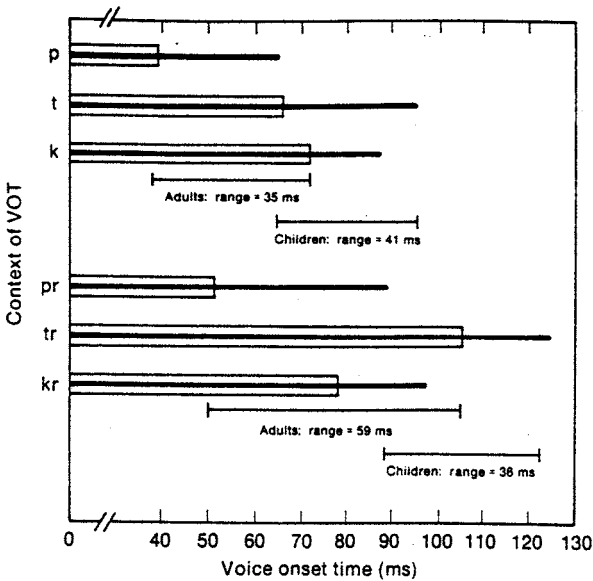


Figure 2. Voice-onset time for word-initial voiceless stops preceding a vowel or /r/. VOT for both adults and children increased when the stop was clustered with /r/. For adults, the range of VOT over place of articulation was greater in the clustered contexts, but for children the range was slightly greater in the unclustered contexts. The difference in ranges of VOT between the two groups of speakers was greater in the clusters, and the children typically produced the smaller of the two ranges in this context. □, Adults; ■, children, year 2. (From Hawkins, 1979a,b.)

their longer absolute durations. As can be seen in Fig. 2, this occurred because the children did not shorten VOT as much as adults do in clusters with /p/ and /k/ followed by /r/. VOT for the cluster /tr/ was more similar to that of the adults. This insufficient shortening is found whether we compare relative change in duration (the ratio of clustered to unclustered stop VOT for each speaker-group), or the absolute durations for clustered contexts only, between adults and children. The data suggested that the reason for this adult-child difference was probably because VOT for the homorganic /tr/ is long enough that the children approximated it relatively more closely than VOT for /p, k/. Presumably they had not yet learned to shorten VOT in /pr, kr/ to the adult standard. Their failure to shorten was especially evident for clusters with bilabial stops.

A reasonable interpretation of this "failure to shorten" is that the children were not coarticulating (or overlapping) the /r/ with the preceding stop as much as the adults in /pr, kr/. There are stricter limits on the coarticulability of homorganic /tr/ than of nonhomorganic /pr/ and /kr/. Thus adult and child VOTs should be similar, as was found. Differences in degree of coarticulation could affect VOT by changing the rate of airflow and degree of pressure drop

which in turn could change the time when the vocal folds begin vibrating. The glottis must be wide open in order to maintain aspiration after stops in word-initial position. For any given degree of vocal fold adduction, a high airflow will cause the folds to vibrate earlier because the pressure drop across the glottis will be greater. The less constricted the upper vocal tract, the faster the airflow and hence the greater the pressure drop across the glottis; conversely, the more constricted the upper vocal tract, the smaller the pressure drop (Stevens, 1971). If then the tongue tip is close to the palate for /r/ when the preceding stop is released, glottal flow will be impeded and voicing will be delayed. This would explain the long VOT in /r/ in adults' and children's speech. In the case of /pr/ and /kr/, coarticulation of /r/ during the closure period could mean that the vocal tract was somewhat less constricted immediately after the stop release; the pressure drop across the glottis would be higher, and the folds would vibrate correspondingly earlier, thus shortening VOT (assuming a constant degree of vocal fold adduction). The suggestion is, then, that whereas adults coarticulate /r/ with /p/ and /k/, but not with /t/, children appear not to coarticulate the stop and liquid gestures in **any** of these clusters. Consequently, children's VOT for /r/ will differ from the adults' to a much smaller extent than will their VOT in clusters with /p/ or /k/. Similarly, children will produce much smaller differences in VOT dependent upon the place of articulation of the following stop. This explanation should be verifiable spectrographically. If correct, it suggests that neuromotor immaturity prevented the children from overlapping gestures in certain contexts so that they failed to differentiate between coarticulatable and noncoarticulatable contexts, thus producing "stereotyped" VOTs.

3. *Gradual Maturation*

Many aspects of children's temporal control seem to mature gradually to the adult norm. For example, in a cross-sectional study of adults and groups of 4-, 7-, and 10-year-old children, Flege, Brown, and White (1981) report that there is a gradual increase in the duration of voicing in word-final stops as a function of age. Certain aspects of this third point deserve closer consideration, however. First of all, we need to define clearly what **gradual** means: how much fluctuation in rate of change do we allow before we decide something is not changing gradually? This question obviously is related to the issue of variability discussed previously. Although statistics can help, most of us rely on intuition much of the time; however, a set of more objective criteria should be derived.

A related problem is the procedural question of how frequently a given phenomenon must be sampled before we can state confidently that it matures gradually. Cross-sectional studies pose particular problems in this respect, but longitudinal studies are not immune from it. To some extent the question revolves around whether one is concerned with long-term or short-term variability (Menn, 1979): a study of changes in speech patterns from 2 years to puberty is less likely

to be as concerned with small fluctuations in rate of change than is a study of the acquisition of a single contrast spanning only a few months. Even if it were possible, it is not necessarily desirable to try to develop explicit criteria for determining the optimal sampling rate to follow the development of a particular aspect of speech, especially as there are individual differences in rate of development. The best solution seems to be selection of a sampling frequency based on a combination of the probable rate of change and the overall context in which the results are to be placed.

Lastly, gradual change might not be expected all of the time. The individual often improves a skill at uneven rates, as discussed further below. Even group data might be expected to show some discontinuities. Puberty, for example, is associated with many abrupt changes in motor development as well as in other areas (Denckla, 1974; Kent, 1976). In speech timing, Gilbert and Purves (1977) found that 9- and 11-year-old children were similar to adults and different from 5 and 7 year olds in terms of the absolute durations of the consonants studied. It remains to be seen whether this change happens gradually or abruptly between the ages of 7 and 9 years. In any case, the change does not appear to be evenly distributed over the entire developmental period.

In summary, gradual maturation toward the adult norm appears to be the rule rather than the exception, and is generally relatively easy to identify. There are some interpretational problems even in this simple case, however. These problems include the statistical definition of "gradual" and the sampling frequency necessary in order to conclude with confidence that changes observed over time occurred gradually.

4. Overgeneralization

The fourth general principle, overgeneralization of rules, will contribute to the number of stereotyped patterns in a child's speech. As mentioned earlier, one way to distinguish overgeneralization from an initial failure to differentiate is to look in longitudinal studies for cases of previously correct or more mature performance "regressing" to a less advanced mode. Regression has been identified in most areas of language acquisition. In phonology, Smith (1973) includes it as one of the seven classes of data that any theory of language acquisition must account for. (Smith calls it "recidivism.") Bickley (1983) has shown that developing distinctions between high *versus* low and front *versus* back vowels may collapse again during periods of rapid lexical acquisition. In syntax, overgeneralization has been used to explain, for example, the replacement of previously correct past tenses of irregular verbs by regular verb endings after the latter have been learned, so that *went* becomes *goed* or *wented* (e.g., Brown, 1973; Slobin, 1973).

The lack of longitudinal data means that we have no really good example of regression from overgeneralization of rules in speech timing. However, there is

some suggestion that regression occurs in the timing of voiceless stops in initial stop-liquid clusters (Hawkins, 1979a,b). Contrary to the general rule for abbreviation of clustered consonants, adults do not shorten /p/ or /k/ closures in stop-liquid clusters, although /t/ is shortened. In contrast, the children generally shortened /p/ and /k/ as well as /t/ closures. The pattern for /p/ in /pɹ/ is especially germane in that the closure was **not** shortened in the first year of the study but was shortened when the children were 14 months older. In both years of study, /k/ in /kr/ was shortened, with most abbreviation when the children were younger. Thus /p/ but not /k/ showed the classical pattern for inappropriate generalization of a rule. The timing of clustered velar stops, however, was usually more advanced than that of bilabials in this study. It may be that the first study took place when clustered /k/ had already been inappropriately abbreviated, but before /p/ had, and by the time of the second study clustered /k/ was beginning to return to the adult nonabbreviated duration whereas clustered /p/ had become subject to the rule for abbreviation. Further support for overgeneralization of a common rule for abbreviation comes from the data for individual children in this study; these are described in Hawkins (1976).

If we accept this evidence for overgeneralization of rules, we should speculate on the nature of the underlying process. In syntax and phonology, overgeneralization is usually regarded as a manifestation of an active process of rule generation and hypothesis testing. These terms suggest conscious mediation, although the implication is not necessary. It is difficult to apply the same reasoning to changes in segmental duration, if only because one assumes that a pattern must be perceived at least partly consciously in order for it to be emulated. This does not seem to be likely in the case of temporal abbreviation of clustered consonants. Although the average change in consonant duration from all unclustered to clustered contexts is perceptible,¹ we do not know that such changes actually are perceived in natural situations. Certainly it is not something that the average adult or child is aware of in the same way as we are aware of the past tenses of verbs. Additionally, of course, the whole point of our argument concerning children's abbreviation of initial voiceless stops is that adults do not abbreviate them in clusters with /r, l/; so there should be no perceptible difference between these clustered and unclustered contexts. An alternative explanation is in terms of motoric maturation and/or longer practice at the task: As a motor task becomes more skilled, its components tend to become faster. Perhaps we are witnessing here just such an increasing motor skill, as distinct from a reorganization based on perceptual information of abbreviation of clustered consonants. If this is the case, however, why does the child become disproportionately faster for clustered stops, or indeed why does the **adult** not abbreviate in these con-

¹ Twenty-one percent over all consonants in Hawkins' (1979a) study; 33% if those segments that were either lengthened or unchanged in duration in the cluster are excluded from the calculation.

texts? One could speculate on the need to preserve a certain rhythm (cf. Huggins, 1972), on the fact that stop closure durations are already relatively short and must be longer than a certain minimum in order to be heard as silence, and so on. Answers to questions such as these must be found before we can expect to be able to interpret the children's data confidently in terms of processes such as "overgeneralization of rules."

5. *Alternation of Rapid Change with Quiescent Phases*

We do not have good data to support the fifth and final general principle, stages of rapid development alternating with "plateaux" or quiescent phases, but one study lends partial support and some of the reasons for our lack of data are worth mentioning. Some of the children that Hawkins studied appeared to mature faster than others between year 1 and year 2, as reported in Hawkins and Allen (1977). In this paired-comparison experiment, 48 students and 4 trained phoneticians rated the 6 children of Hawkins' (1979a) study in terms of the apparent maturity of their speech, with "fluency of articulation" playing a major role in the definition of maturity. Not only did the perceived maturity of speech fail to correlate with age within a given year of study, but the individual children were also perceived as maturing at very different rates between the 2 years in terms of the criterion used. However, although these judgments were reliable, they did not correspond with simple acoustic measurements of maturity in terms of degree of deviation in durational modification from the adult norm.

We need to investigate this idea of plateaux more closely in children's speech before we can take it as established fact. I do not know of any good evidence for the existence of plateaux in the acquisition of phonology, but this may be because we lack the appropriate investigative techniques rather than because there are no plateaux. There are at least two questions that must be resolved before we can obtain better data. One problem is that our units of measurement may not be fine enough, so that we miss small changes and interpret them as plateaux. Another question, more crucial theoretically, is whether we are looking at the right thing: there may be no change in the parameter being measured, but some other relevant parameter could be changing. For example, it is conceivable that in a given period variability in a child's VOT could be stable while variability in associated vowel duration could be decreasing. If we were only measuring VOT, we might conclude that the child is on a plateau with respect to the voiced-voiceless distinction. Measuring both VOT and vowel duration would be more likely to make us question what exactly we mean by "the development of the voicing contrast in stops." We would be correct in saying that the motor control of upper articulators and larynx was not changing, but whether or not this was a meaningful statement in terms of the child's developmental strategies would be open to question.

D. Summary: General Principles of Motor Skill Learning

Irrespective of whether we found strong support for it, each of the general principles discussed above raised theoretical or methodological issues of interpretation requiring further study. The issue of variability raises questions of multiple causes for measured phenomena, of adults and children differing in the target, and of trade-offs between articulation-dominant and timing-dominant systems of production. The discussion of stereotyped behavior, interpreted as an inability to differentiate between contexts, illustrates the influence of complex interactions between different aspects of speech, such as when a generalized failure to coarticulate has aerodynamic consequences that cause long voice-onset times in nonhomorganic stop-liquid clusters. Additionally, the discussion of stereotypy demonstrated the importance of distinguishing between constraints on production and constraints on perception. Productive versus perceptual constraints were also contrasted in the discussion of overgeneralization of rules, this time in the context of the development of increasingly sophisticated motoric routines versus active testing of rules for production derived from perceived differences in adult speech. Problems of measurement, especially of identifying appropriate units and of sampling frequency, were raised in evaluating evidence for gradual maturation as well as for alternation of periods of rapid development with more static phases. Evaluation of the latter principle also provided an example of the importance of considering and accounting for each of several phonetic attributes of a single phonological contrast before making claims about the way that contrast is acquired.

Although the framework we have been using is useful in a first attempt to systematize our current knowledge, there are at least two critical ways in which it should be refined. First of all, the general principles are descriptive only; they do not explain the data, although they can guide the search for explanations. Second, not all the data seen are easily described in terms of these general principles of learning. Some of these data that do not fit the general descriptive categories are especially interesting in that they suggest specific strategies that the child may use to refine her articulatory coordination; hence they take us a little nearer to explanatory rather than simply descriptive adequacy. The remainder of this article addresses these issues.

VI. TWO STRATEGIES FOR LEARNING TO PRODUCE CLUSTERS OF CONSONANTS FLUENTLY

In the early stages of learning to speak, children's attempts at consonant clusters differ from the adult forms in a number of different ways. The rela-

relationship between the child's attempt and the adult form is generally governed by strict rules, and the rules differ according to the manner class of the consonants involved. Examples of such modifications of adult clusters which are commonly heard in the speech of children less than 3 or 4 years of age are the following. An epenthetic vowel may be inserted, especially between a stop and a following approximant, implying a lack of coarticulation between the two; e.g., [gə^hlʌv] for /glʌv/, and [bə^hɪɛk] for /brɛk/. One or more elements of the cluster may be omitted. A frequently omitted sound is /s/; thus /stap/ becomes [tap] or [t^hap] and /spʌn/ becomes [pʌn] or [p^hʌn]. Another type of modification occurs when some features of each of two clustered consonants are combined to form a single new sound which does not exist in the adult language. Examples are [ʃ] for /sl/, and [m̩] and [ŋ] for /sm/ and /sn/, respectively. Hence we may find [ʃt] for /slt/, [ʃɪp] for /slɪp/, [mouk] for /smouk/, and [ŋou] for /snou/.

Between about 3 and 4 years of age, these early attempts at consonant clusters are usually replaced by forms in which all the correct phonemes are present; but, as we have seen, they are not yet timed maturely. Moreover, the details of the development of mature temporal integration of speech gestures differ according to the manner of articulation of the sounds involved. Most relevant to this discussion, stops in children's speech appear to be treated differently from continuants and to be integrated into phonetically complex clusters in a different way. This section discusses two strategies for production of different types of clustered consonants that apply to this later stage, that is, after all elements of the cluster are present, but before they are executed fluently as fully coarticulated units. One strategy, termed **reprogramming**, is applicable to clusters including a stop, and the other, **concurrent programming**, is applicable to clusters involving only continuants. The term **programming** denotes the detailed organization of articulatory gestures that allows their execution. It is presumably a relatively "low-level" process that takes place after such things as word selection and sequencing of syllables and phonemes. The strategies are described here in somewhat anthropomorphic terms, partly for clarity and partly because I do not know of an appropriate physical explanation. If future evidence is sufficiently supportive, the anthropomorphic connotations will have to be removed.

The two strategies were proposed by Hawkins (1976, 1979b) as explanations for her finding that the temporal integration of children's fricatives and liquids in clusters appeared to move gradually and consistently nearer to the adult patterns, whereas stops underwent changes away from the adult pattern at certain stages of development. The design of the study prevented unequivocal support for any explanation involving more than two stages and therefore the strategies are only tentatively proposed, but if the different ages and maturity levels of the individual children are considered, the evidence is quite supportive.

The explanation presupposes that the resources (or attention) that the child can apply to speaking are finite. Breakdowns in speed or fluency occur when re-

sources are tapped to capacity; attention is not necessarily distributed evenly over the entire range of an utterance that is planned. Hawkins (1979b) suggested that production of a stop closure requires a minimum of monitoring by the child, since articulatory overshoot is possible. (That is, as long as closure is maintained above a certain minimum of pressure, both the silent period of the stop and the burst following it will be successfully produced. The child need only ensure that the closure is sufficient to create these conditions. Within certain limits, it does not matter if there are variations in articulatory pressure above the necessary minimum, so attention can be directed to other things during the production of stop closures.) Producing a continuant, on the other hand, requires much more constant monitoring because the relative positions of the articulators involved may vary by only small amounts. An /s/ that is overshoot, for example, might become a /t/ or a voiceless tap [ɾ]; if an /s/ is undershot, it will produce a whispered vowel-like sound. Similarly, an overshoot /l/ might sound like a /d/, and an undershot /l/ would sound vowel-like.

Consider the articulatory strategy of reprogramming. The hypothesis is that at the time when the child is just beginning to integrate the gestures for clusters involving stops into single complex units, she divides the cluster into two parts. Let us use /spr/ as an example. The child executes the /s/ and its transition into the /p/ as a well-integrated complex gesture. She then executes the transition from the /p/ into the following /r/ as a second integrated complex gesture. During the early part of this stage, the child must concentrate most of her "programming resources," or attention, on the first gesture in order to execute it smoothly and rapidly, so that preparation for the second and later gestures is minimal. During the stop closure these later gestures are "reprogrammed" in more detail before execution. This is possible since maintenance of the closure should require relatively little attention because articulatory overshoot is possible. But reprogramming takes time and leads to a long closure period. In the later part of this stage, the segments surrounding the stop closure, /s/ and /r/ in this example, will be executed more maturely than the stop itself and more maturely than the same segments in clusters to which reprogramming does not apply.

Although any stop could potentially serve as a focus for a reprogramming pause, it is suggested that the stop closure will be used for reprogramming only when it is in an articulatorily complex environment. Thus, initial clusters of /s/-plus-stop (with or without a following liquid) will be subject to reprogramming more often than initial stop-plus-liquid clusters. Indeed, the latter clusters should only be subject to reprogramming when preceded by articulatorily complex sequences. Hence, in Hawkins' (1979a) data, in which the words were simply preceded by /ə/, we would expect to find reprogramming pauses only for clusters with initial /s/, and not for those beginning with stops, that is, only for /sp, st, sk, spr, spl, str, skr, skw/; this was in fact the case.

In contrast to reprogramming, concurrent programming was hypothesized to

be a strategy that the child might use for clusters of continuant consonants. It is suggested that clusters such as /sl/, for instance, require constant monitoring of articulatory position. Attempts at increasingly fluent transitions between elements of the cluster must occur concurrently with their production, since maintenance of each articulation requires so much attention that anticipatory planning of transitions is minimal. Consequently, in these clusters we find a general slowing of some or all segments and a gradual approximation to the adult norm, rather than the piecemeal and uneven changes of clusters subject to reprogramming. Although all segments would be expected to be produced slowly if the child used concurrent programming, they need not necessarily be produced equally slowly. Moreover, concurrent programming could logically involve anticipatory slowing of a relatively easy segment while preparing for a more difficult later one. Conversely, a relatively easy later segment may be produced with a slower onset if the child's "programming resources" were mainly occupied with the achievement and maintenance of a more complex preceding articulation. Thus, it is not suggested that concurrent programming always involves instant-by-instant monitoring. Rather, concurrent programming involves closer or more frequent monitoring than the execution of simpler gestures. In contrast with sequences subject to reprogramming, the monitoring required at any given instant in time is generally great enough to preclude much fine programming of later gestures.

Let us examine the data that originally suggested the strategies of reprogramming and concurrent programming. Figures 3 and 4 show the duration of each segment in the clusters beginning with /s/ and with a stop as a second segment. In the two-segment clusters, adults lengthened /p/ but shortened /t/ relative to their durations in unclustered environments; /k/ remained unchanged in duration. In contrast, the children lengthened all three stops quite considerably. Similar differences occurred with the three-segment clusters; the children lengthened the stops in all five clusters, whereas the adults only lengthened /p/ and the /k/ of /skw/, with /t/ being shortened. (The adults' nonabbreviation of /k/ in /skr/ is meaningless since it represents the mean of two elongations, two abbreviations, and one unchanged duration in the individual data.) In the contexts in which the adults lengthened the stops, /sp, spr, spl, skw/, the children's lengthening was disproportionately large. This exaggerated lengthening of stops after /s/ demands an explanation involving phonetic context since it did not apply to stops in stop-liquid clusters. In these clusters, the children's stops were either abbreviated by a similar proportion to the adults' or overly abbreviated relative to their unclustered durations.

The relative durations of the segments in these clusters for individual children in each year provided evidence suggesting the following pattern for clusters of initial /s/-plus-stop, with or without a following liquid. When the cluster is first produced with all its consonants, the clustered segment durations are unchanged

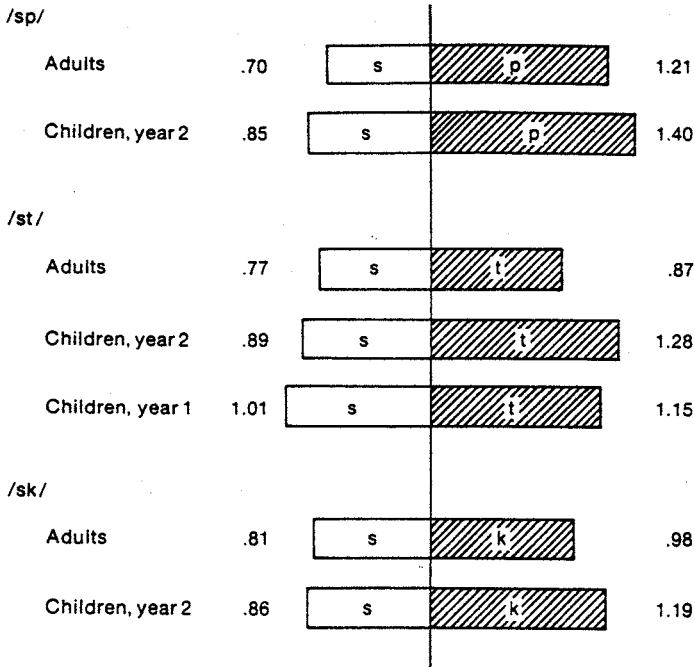


Figure 3. Durational changes in /sp/, /st/, and /sk/. Each rectangle represents the ratio of the mean duration of a clustered consonant to the same segment in unclustered context for a given group of speakers. Numbers to the left (open bar) of the figure are the ratios of mean durations of clustered to unclustered /s/ for the indicated group of speakers. Numbers to the right (shaded bar) are the ratios of mean durations of clustered to unclustered stops for the indicated group of speakers. A value that is less than 1.00 indicates that the consonant was shorter in the cluster than when it was unclustered, and a value that is greater than 1.00 indicates that the clustered consonant was longer than the unclustered one for that group of speakers. See text for further explanation.

from their respective unclustered durations, or possibly just slightly abbreviated. Later, the clustered stops are lengthened. Following this, the durations of /s/ and of the following liquid, if present, become quite adult-like relative to their unclustered equivalents, whereas the stop remains long. Finally, when the articulatory sequence is sufficiently practiced so that reprogramming is not necessary, the stop is no longer lengthened disproportionately and temporal relationships in the cluster are at or near their adult values. This reduction in stop duration would probably happen gradually, as the need for reprogramming decreased. Thus, the timing of the noninitial stop shows movement away from the adult norm and away from the usual pattern of increasing abbreviation of clustered segments

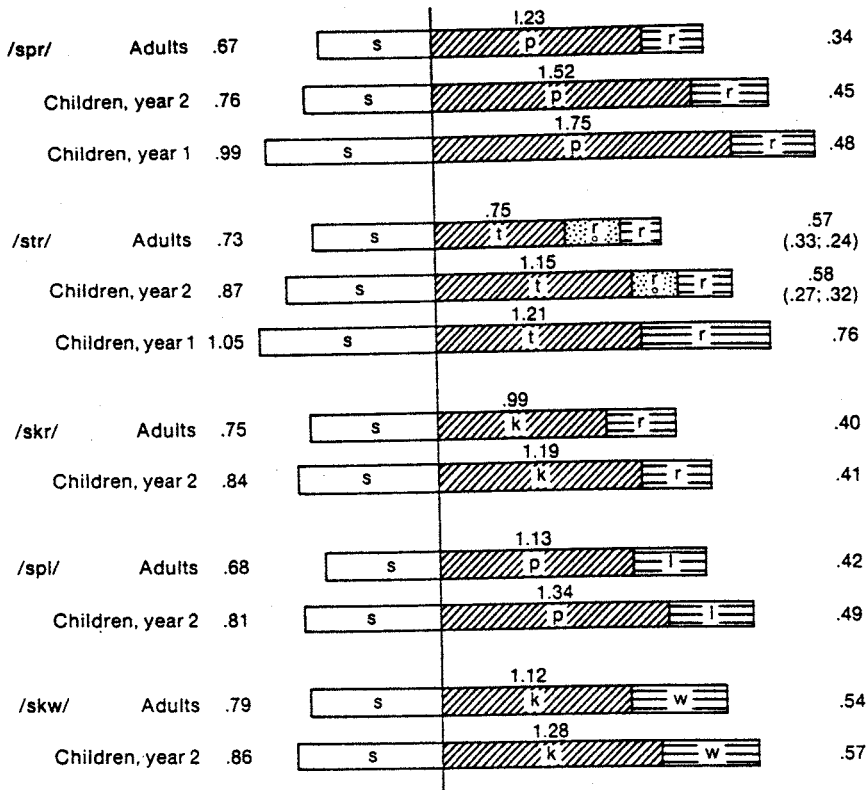


Figure 4. Durational changes in /spr/, /str/, /skr/, /spl/, and /skw/. Each rectangle represents the ratio of the mean duration of a consonant in a particular cluster to the same segment in unclustered context for a given group of speakers. Numbers to the left of the figure are the ratios of mean durations of clustered to unclustered /s/ for the indicated group of speakers. The numbers above the rectangles representing each stop are the ratios of clustered to unclustered mean duration for that stop, for each group of speakers. Numbers to the right are the ratios of mean durations of clustered to unclustered approximants, for the indicated group of speakers. Numbers in parentheses beneath the values for /r/ in /str/ are ratios for voiceless and voiced /r/ (or aspiration and voiced /r/). A value that is less than 1.00 indicates that the consonant was shorter in the cluster than when it was unclustered, and a value that is greater than 1.00 indicates that the clustered consonant was longer than the unclustered one for that group of speakers. See text for further explanation.

(relative to their unclustered durations) that characterizes most of the children's development. At the same time, the durations of segments surrounding the stop approach the adult norm more rapidly than the timing of the same segments in other contexts. For example, liquids in three-segment clusters were more maturely timed in Hawkins' (1979a) data than were liquids in most other clustered contexts (i.e., /fl, sl, sw/ and in initial voiceless stop-liquid clusters like /pl/ or /tr/.)

For reasons discussed below, it is difficult to adduce unambiguous evidence supporting the strategy of concurrent programming. However, the last point made, that liquids in three-segment clusters were timed more maturely than liquids in /sl, fl, sw/, supports concurrent programming as well as reprogramming. For example, /l/ would be expected to be longer in /sl/ than in /spl/ since concurrent programming is thought to induce slower execution of all elements of the cluster.

To summarize, the main evidence for reprogramming and concurrent programming as strategies lies in the uneven development of segmental timing in clusters with initial /s/ followed by a stop, compared with the apparently rather more steadily maturing durational modifications in fricative-liquid clusters. Hawkins' (1976) data included a substantial amount of more circumstantial evidence which, taken together, adds credibility to these ideas. This evidence is summarized briefly below and is followed by supportive evidence from other sources.

There were five additional aspects of the children's data that support the idea of reprogramming. First, the children's stops were considerably longer in three-segment clusters compared with stop-liquid clusters (e.g., in /spr/ compared with /pr/). The adults, in contrast, had very small differences in stop duration between the two types of cluster. The unusual shortness of the children's stops before liquids has already been discussed in the context of overgeneralization of a rule for abbreviation of clustered segments. Stops may be particularly subject to overgeneralization while the child is using reprogramming as an articulatory strategy: when a stop is the second element of a cluster and reprogramming occurs, fluent execution of the segments surrounding the stop apparently takes place at the expense of the "normal" duration of the stop. Consequently, the durational target for the stop might become less clearly specified than usual. If such a loss in specificity for the durational target generalized to all clustered voiceless stops, those stops not subject to reprogramming could become vulnerable to a general rule for abbreviation ("If clustered, abbreviate whenever possible"). The concept of reprogramming, then, can explain some findings that are not directly involved with the strategy itself.

Second, there is support for the idea that reprogramming applies selectively to more complex articulatory sequences. The children's stops were generally longer in three-segment clusters, whereas in the adults' speech, stops in three-segment

clusters were shorter than, or the same duration as, stops in two-segment clusters beginning with /s/. The difference discussed above in duration of stops in stop-liquid versus three-segment clusters also suggests that reprogramming may be a strategy reserved for complex clusters. Furthermore, the data suggested that the developmental pattern characteristic of reprogramming appears later in homorganic clusters, /st, str/, than in nonhomorganic clusters. Homorganic clusters are generally considered to pose more problems for reasonably fast production than nonhomorganics (e.g., Haggard, 1973), and this is supported in Hawkins' data for children in that temporal control of homorganic clusters generally lagged behind their control of the equivalent nonhomorganic clusters.

Third, reprogramming would be expected to appear relatively late in the developmental sequence. The data for homorganic /st/ confirm this in that the second time the children were recorded (14 months after the first time), /t/ was proportionately longer and /s/ was proportionately shorter, leaving the relative abbreviation of the whole cluster about the same in the 2 years. This difference between the 2 years does not hold for /spr, str/, which were the only other clusters with initial /s/ that were recorded in the first year. The discrepancy may be partly explained in that /spr/ is nonhomorganic and so might be expected to show the effects of reprogramming earlier, or for a shorter time, than homorganic clusters.

Fourth, reprogramming predicts that liquids in three-segment clusters should be more maturely timed than in other contexts not open to reprogramming since the reprogramming interval is thought to enhance the fluency of execution of the immediately succeeding articulations. This was the case. Differences in relative abbreviation between adults and children were generally smaller in the three-segment clusters than in other contexts. VOT was similarly adult-like in these clusters. Moreover, with only one exception, the children's /l/ was abbreviated less in /sl/ than in /spl/, whereas the adults' /l/s were abbreviated by the same amount in both clusters. (Three of the children's differences in /l/ duration were less than 10 msec and not statistically significant, but they are nevertheless in the right direction to support both reprogramming and concurrent programming.)

Finally, the data suggested not only that individual children were at different stages with respect to their use of reprogramming, but also that these differences corresponded with differences in perceived maturity. For example, the child who was judged as sounding most mature in the second year of recording (Hawkins & Allen, 1977) appeared to have used reprogramming in the first year but to have passed beyond that stage by the second year of recording. In the second year, he had an adult-like pattern of relative abbreviation for three-segment clusters, including very strong /s/ abbreviation. In the previous year, although his /s/ was short, /t/ in /str/ was very long. Because it is homorganic, /str/ is thought to be subject to reprogramming later than nonhomorganic clusters. In contrast, the child judged as speaking least maturely in both years showed little sign of

reprogramming. Most of her stops in both two- and three-segment clusters were either unchanged or abbreviated relative to their unclustered durations.

Further evidence supporting concurrent programming is often open to alternative explanations, as we shall discuss below. However, three points can be mentioned. The unusually adult-like abbreviation of liquids in three-segment clusters has already been discussed as supporting concurrent programming. Second, the children exaggerated the adult tendency for voiced [l] to be longer in /sl/ than in /kl/. Third, the children's [l] was longer in /fl/ than in /pl/ whereas the adults had no difference in liquid duration in this context. As with /sl/, this implies more slowing of the fricative-liquid sequence, which is probably a more complex articulation than a stop-liquid sequence and yet is not open to a reprogramming pause.

All the evidence used so far to support the notions of reprogramming and concurrent programming concerns details of segmental timing dependent upon phonetic context. In addition, certain aspects of reprogramming and concurrent programming accord well with observations on other aspects of children's developing linguistic skills. For example, Heusner and Hoffman (1980) studied the effects of contrastive stress on the durations of /C₁#sC₂/ clusters in the speech of adults and of normally speaking and /s/-misarticulating children of 5 and 6 years. C₁ and C₂ were /p/, /t/, or /k/ in all possible combinations and # was both a word boundary and an NP-VP boundary. (Examples are "The cop spilled the milk"; "The cook stained the rug.") When the word preceding the boundary (#) received contrastive stress, the adults lengthened the /s/ that followed this boundary, compared with the normal stress condition. In contrast, both groups of children left the duration of /s/ unchanged from the normally stressed condition but lengthened one or both stops, depending upon whether the preceding or following word was stressed. This suggests that children are more likely to change the durations of stop closures than of /s/.

Branigan and Stokes (1981) report that at about 2 years, when a child repeats a word in a phrase and produces a more phonemically correct version the second time, a fairly lengthy pause often precedes the correction. These interword pauses are much longer than the reprogramming pauses that we have suggested occur during stop closures, but they do provide good evidence that the child may pause before producing a particular articulatory sequence in a more sophisticated way than she usually does.

The two strategies seem to appear between the ages of about 5 and 7 years, which is the age at which children begin to be able to break words up into their component sounds (Fox & Routh, 1975; Wallach & Wallach, 1976; Barton, Miller, & Macken, 1980). Although it is possible to train younger children to "decode phonemically," this skill usually develops rapidly and relatively spontaneously in 5 to 7 year olds. Several investigators have suggested that the child first learns to speak with the whole word as the basic unit (Leonard, Newhoff, & Mesalem, 1980; Kent, 1982). Menyuk and Klatt (1968), Kornfeld (1971), Men-

yuk (1971), and Barton *et al.* (1980) maintained that consonant clusters are single phonological units for the child. Assuming these suggestions to be valid, then, reprogramming and concurrent programming may represent part of a process in which larger units like words or syllables are analyzed into their component segments, after which the temporal characteristics of the sequence can be refined. This possibility raises questions about the forms of reorganization that must occur as the speech production mechanism matures; this will be discussed further in Section VII.

There are several serious problems in assessing the validity of reprogramming and concurrent programming as articulatory strategies. We discussed earlier the problem of defining without circularity a difficult or articulatorily complex cluster. Although a sequence must satisfy certain structural requirements in order to be considered open to reprogramming or concurrent programming, the issue of difficulty is still important because the two strategies are thought to apply only to complex sequences. Sufficiently detailed longitudinal data should allow us to evaluate the validity of the notion of reprogramming. It is more difficult to avoid circularity for concurrent programming. One way to do so might be to look for the spread of coarticulatory effects across segments: there should be less anticipatory coarticulation in segments still subject to concurrent programming. For example, we might expect children to produce a normal degree of anticipatory liprounding for words like *to*, *soup*, and *loop*, but less liprounding for words like *sloop*, in which the juxtaposition of /s/ and /l/ could necessitate concurrent programming and hence less anticipation, at least in younger children. This approach has problems of its own, however, the most important of which is the difficulty of defining and measuring "degree of coarticulation." As it seems reasonable to expect gradual rather than abrupt changes in the development of coarticulated sequences, achieving an adequate definition of degree of coarticulation could be critical. A related issue concerns the linguistic domain to which the term "programming resources" refers. The similarity between Hawkins' data for segments and Heusner and Hoffman's (1980) for contrastive stress suggests that articulation of prosodic features as well as manner and place of articulation should be included. There seem to be no strong arguments for restricting "complexity" to articulation; a complex syntactic or semantic/cognitive structure may encourage a replanning pause or a general slowing consistent with concurrent planning, perhaps in addition to a pause between words, as reported by Branigan and Stokes (1981). To test this idea we could examine children's speech in contexts differing in type of complexity. In addition to expecting stronger evidence of reprogramming and concurrent programming, in more complex contexts we would predict less coarticulation, slower or more variable segment durations, disrupted prosody, and perhaps the reappearance of reprogramming pauses in the speech of children who were just moving out of the reprogramming stage.

Another problem in assessing the validity of these articulatory strategies is that

of distinguishing between the effects of reprogramming and those of stereotyped responses to stops in clusters. For example, the extra lengthening of /t, k/ in three-segment clusters could be in conformity with the lengthening of /p/ after /s/, which even adults do. Although reprogramming is assumed to apply more strongly to more complex clusters, the existence of a reprogramming effect would not necessarily be disproved if the proportional increase in stop duration were no larger for three-segment than for two-segment clusters. Not only would such an argument have to assume that three-segment clusters are harder than two-segment ones, it would also have to assume a measurable increase in programming time with each additional phonemic segment, an untenable position. This question is, however, a potential topic for research; for example, how is "programming time" related to the number or type of segments to be produced?

Concurrent programming and reprogramming are by definition compatible with interpretations of adult-child differences in temporal organization based on motor difficulty, because the strategies are hypothesized to apply specifically to articulatorily complex sequences. However, the phonotactic constraints of English prevent us from discovering whether the strategies are in fact general to all clusters of the appropriate structure or whether they are specific to the particular clusters of English. For example, in English all syllable-initial clusters of three segments begin with /s/ so that the effects of larger clusters of prevocalic consonants cannot be separated from the influence of having to start the cluster with /s/. We need to study children learning to speak languages representing a wider variety of types of cluster before we can clearly dissociate the evidence for use of these strategies from that for a motor problem specific to sequences of /s/-plus-stop, for example.

Finally, the two strategies have been discussed in terms of differences in articulatory programming prior to execution. An alternative interpretation of the data is in terms of differences in the use of feedback according to the phonetic structure of the sequence. Feedback from stops, for example, may be used differently from feedback from continuants. This interpretation in terms of feedback is not necessarily very different from that in terms of programming strategy: the programming strategy may derive from the type of feedback. At this point there does not seem to be a good way of distinguishing between these two interpretations. The possibility that reprogramming pauses reflect the use of feedback is returned to below in connection with schema theory.

VII. TOWARD A THEORY OF THE DEVELOPING MOTOR CONTROL OF SPEECH

A. Defining a Motor Skill

The foregoing pages present data illustrating principles presumed to govern the developing control of speech. The present section draws on that discussion in

presenting a more unified account of what a theory of the developing motor control of speech might look like. We start from the basic premise that speech is a motor skill learned in interaction with developing cognitive and linguistic sophistication and subject to constraints on perception as well as on production. As a motor skill, speech is learned in accordance with laws governing the acquisition of any other motor skill, although the unique relationship between speech and other linguistic and nonlinguistic systems means that its acquisition may also have unique aspects.

What then is a motor skill? Bruner (1973) described a skilled movement as involving "the *construction* of serially ordered constituent acts whose performance is modified toward less variability, more anticipation, and greater economy by benefit of feedforward, feedback, and knowledge of results" (p. 5; original italics). He identified three crucial characteristics of skill learning: anticipation of subcomponents of the act; modularization, manifested by a reduction in variability of latency and execution time, together with more economical expenditure of energy; and reorganization of the act into a "higher order" pattern which may include the act mastered earlier. Modularization allows the attention necessary to control an act to be reduced, and this in turn allows the act's incorporation into higher order and longer sequences with only minimal disruption.

With an orientation similar to Bruner's, Connolly (1977) explicitly distinguished between movements, actions, and skills and stressed the importance of intention and flexibility in a skilled act. Movements make up actions designed to achieve some goal; a movement not part of a purposive action is meaningless to the organism (unless affected by some chance outside influence). Connolly (1977) defines the development of a skill as "the construction of a programme of action which is directed towards the attainment of a goal" (p. 133). Actions are made up of subroutines of movements (cf. Bruner's modules), and it is the fluent integration of the subroutines most appropriate for a particular context that constitutes a skilled action. An action is skilled if its consequences satisfy the original intention with maximal efficiency. Just as intention is not a defining characteristic of neuromuscular activity, so a particular pattern of muscular contractions is not a defining characteristic of a skilled action. The skill lies in selecting which muscular contractions will achieve the desired goal most economically in the particular context. Connolly illustrates this distinction through the skill of writing, either at a desk with a pencil held in the preferred hand, or vertically on a wall with a piece of chalk tied to a broomstick that must be held with both hands. The skill is the same in both cases; the intention may not differ appreciably, but the muscular contractions are quite different. The quality of the skilled action is judged by how efficiently the intention is satisfied, by the writer's flexibility in selecting, sequencing, and phasing the most economical set of subroutines to achieve the goal.

Speech presents a similar situation in that it consists of varied orderings of a

relatively small set of elements, and this characteristic necessitates being able to produce the same or similar acoustic effect from a variety of starting positions. There are external constraints on the acoustic variation allowed for the expression of any one phoneme, just as there are external constraints on writing posed by the necessity for others to read it, and on the movements of a ball in an organized sport posed by the rules of the game. So the child's task in becoming a **skilled** speaker is to discover the rules defining acceptable variation and how to satisfy them most economically in each possible context. This task involves identifying the most appropriate set of subroutines and discovering how to combine them.

B. The Unit of Analysis: "Top-Down" versus "Bottom-Up" Models

The immediate problem facing both the child and those who study her is defining the appropriate unit of analysis. The problem applies to studies of adults' as well as children's speech and to perception as well as production. That is, what are the units that the child uses in decoding the speech she hears and in producing speech herself, and what relationship do these units bear to those best used by people studying the child? An extreme way of expressing the question is by contrasting "top-down" with "bottom-up" approaches. A child using a "top-down" system would have a general schema for an overall prosodic frame into which segments are gradually fitted. A child using a "bottom-up" system would learn to integrate gestures for subsyllabic and syllabic units, and in so doing arrive at the overall prosodic frame. A third possibility is that the child employs both systems simultaneously.

Traditionally, studies of child phonology have tacitly adopted a bottom-up model in that they have generally documented phonemic change and paid little attention to prosody or prosodic context. (There have, of course, been exceptions to this emphasis, e.g., Waterson, 1970, 1971a,b; Menn, 1978; Macken, 1979.) In recent years there have been several reports of children who appear to function more with top-down models, at least for some structures (Peters, 1977; Priestley, 1977; Nelson, 1980). As with Waterson's earlier work (1970, 1971a,b), these reports have all been case studies and the authors have noted that they are not representative of the majority of children, a point that has been stressed by others as well [e.g., Branigan and Stokes (1981)]. Representative or not, the behavior of these children must still be accounted for, as indeed must the self-evident fact that children ultimately learn both the prosodic and the segmental structure of their language.

Our knowledge of the acquisition of segmental and prosodic structures in infancy and early childhood suggests the simultaneous use of top-down and bottom-up models. For example, from a few weeks or months of age infants can

discriminate between a large variety of segmental and nonsegmental speech contrasts (e.g., Eimas, 1974, 1975; Spring & Dale, 1977; Kuhl, 1979; Eilers, 1980). There are no grounds for believing that one class of contrasts is more salient for the child. When the child begins to speak, she usually shows some constraints on segmental production, many of which are contextually determined (e.g., Ingram, 1974, 1979b; Leonard *et al.*, 1980). Other constraints also seem to operate, although they are less well understood. Waterson (1970, 1971a,b), for example, analyzed her son's words into five classes according to their segmental and prosodic structure. Another child appeared to have a "whole-word schema" of [CVjVC] into which he fitted all attempts at polysyllabic words (Priestley, 1977). Allen and Hawkins (1979, 1980) found some support for the idea that 3 year olds tend to prefer a trochaic rhythm for polysyllabic words and will distort stress patterns to conform to this. Additionally, there are several reports of children who substitute a fixed "dummy syllable" (Menn, 1979) for any unstressed syllable in a polysyllabic word. Cases have been reported by Smith (1973), Menn (1979), and A. Bell (personal communication), and I have heard 2-year-old twins replace function words and other unstressed syllables in sentences with repetitions of [də] in approximately the correct number and rhythm for that sentence.

These data together suggest that the child is sensitive to phonetic contrasts at several different levels of analysis and attempts to reproduce these contrasts in her own speech. At each level she is subject to constraints on production and perhaps perception. I suggest then that the child operates simultaneously on several levels of overlapping units. As she becomes more skilled, both in decoding the speech-language she hears and in producing it herself, she is able to add more details at each level until finally she is refining subtle aspects such as consonantal duration and the use of fall-rises to imply negation or qualification of the meaning of the spoken words. At any time during development some children may concentrate more on one level than on another. Such differences in attention may have more obvious effects at earlier ages when speech skills are relatively rudimentary. For example, the occasional "pure" analytical or Gestalt child discussed earlier may reflect consistent differences in emphasis. The majority of children probably fluctuate between attending to larger and smaller units in their speech; this could account for some of the intrachild variability that we observed (cf. Branigan & Stokes, 1981). Similarly, the age-related changes between articulation dominance and timing dominance discussed in Section V,C,1 were hypothesized as reflecting changes in emphasis for these simultaneously developing systems.

In Section IV, I suggested that the child's overt speech reflects the interaction of several different processes in development. The points made in the present section substantiate this claim. We turn now to consider how such learning might take place.

C. Schema Theory Applied to Speech Development

One of the basic assumptions made in Section IV was that children and adults will use similar processes in speaking, although the domain of influence, or unit, to which the processes apply may differ. A reasonable extension of this assumption is that although there may be many units of speech production, their acquisition and realization as skilled movements are governed by the same general principles, perhaps with some variations to allow for differences between types of unit. For example, reprogramming pauses as defined above may apply specifically to particular consonantal sequences, but pauses may occur in other contexts such as between words or phrases while an unfamiliar articulatory strategy is being prepared. The question of how the child becomes skillful in speaking is thus reduced for any unit of speech to how a new articulatory sequence changes from being effortful to become automatic.

Phonologists have suggested that the child initially begins with a small set of canonical forms (e.g., CV) whose parameters vary in their specificity (e.g., Ingram, 1976; Menn, 1979). The phonological system matures as the set of canonical forms is enlarged and as the specifications for the parameters are changed. Menn (1979) compared setting up canonical forms to creating a basic program, with maturation occurring as the proportion of variable to fixed parameters in the program increased. She speculated that a further aspect of maturation could involve increasing the number of feature values for parameters that might have nonbinary features such as those specifying vowel height. A later stage of development involved integrating previously discrete programs into higher order units.

These principles derived for children's phonological development are also applicable to the developing motor control of speech. In basic concept they are compatible with recent models of motor performance such as the coordinative structures of Fowler, Kelso, Turvey, and their colleagues at Haskins Laboratories (e.g., Fowler *et al.*, 1980) and the schemata of Adams (1976) and Schmidt (1975, 1976).

Schmidt's schema theory will be briefly described here because one of its strengths is its emphasis on learning rather than simply performance of a motor skill, because it has excited a good deal of attention and because it has already been applied to the development of speech by Kent (1982). Schema theory postulates two separate states of memory. Recall memory generates commands for movement to the muscles, and recognition memory evaluates the response-produced feedback in order to derive information about errors. An additional assumption is that the CNS forms "generalized" motor programs that contain all the details of the muscle commands necessary to carry out a movement. The generalized motor program can be run off when it is provided with response

specifications as to how the program is to be executed (e.g., slowly or rapidly). The two types of memory play different roles depending on the duration of the movement. Rapid movements (of say less than 200 msec) are completely under the control of recall memory in conjunction with the motor program, which predetermines the details of the movement. With such a rapid movement, recognition memory operates only after the movement is completed, comparing expected sensory consequences with response-produced feedback and registering any resultant error. Slow movements allow ongoing adjustments for errors and so can be carried out using both recognition and recall memories. Recall memory initiates a series of programmed movements in the correct direction, and after each one the actual and expected sensory results are compared in recognition memory. Any discrepancy between the actual and expected results provides information for a corrective movement.

Four different types of information contribute to the generation of a skilled movement: the initial conditions for the movement (e.g., the current position of the articulators); the response specifications, or the motor program to realize the desired movement; the sensory consequences of the movement; and the actual result of the movement in relation to the desired outcome. Repeated movements generate several sets of these four types of information and allow a generalized specification of a given skilled movement to be formed, which may not be identical to any single instance of its performance and which is forgotten more slowly than any single instance. This general and long-lasting specification is a schema, which is thus an experience-based relationship between these different types of information. The schemata form the basis of recall and recognition memories, and so there is both a recall and a recognition schema for any given skilled movement. The recall schema is the relationship built up over trials among response specifications and actual outcomes, modified by initial conditions; the recognition schema is the relationship built up over trials among sensory consequences and actual outcomes, modified by initial conditions. So recall and recognition schemata for a given movement share its initial conditions and actual outcomes, but differ in that the recall schema is the relationship between these two variables and the response specifications, whereas the recognition schema is the relationship between these variables and their sensory consequences.

To make a response when the schema has been formed, the current initial conditions and desired outcome are input to the recall schema. The specifications necessary to achieve the desired result are determined from the relationship between past outcomes and response specifications. The particular combination of initial conditions and desired outcome need never have occurred before, since new values can be interpolated between values of actual instances. This accounts for the flexibility of skilled acts. The recognition schema enables the actors to generate the expected sensory consequences of the movement by comparing past

outcomes with past sensory consequences and initial conditions. During or just after the movement, the expected sensory consequences of the correct movement are computed and compared with actual internal and external feedback. Any mismatch is returned to the schema as an error and used in future computations. Thus the schema is a relationship derived from experience of initial conditions, past results, and response programs (the recall schema), and of initial conditions, sensory consequences, and actual results (the recognition schema). The schema allows the actor to predict outcomes, to refine actions on the basis of past errors (i.e., to learn) and to perform skillfully even when the actual motor program required is novel.

Kent (1982) has applied schema theory to speech, especially its development. He suggests that the schema is a useful abstraction describing the motor control of speech in the same way that the phoneme is a useful abstraction in phonology. Moreover, schemata may solve some of the problems in linking phonemes with articulation. In adducing support for schema theory in the development of speech motor control, Kent discusses some of the points touched upon earlier in the present article. For example, he suggests that the child's earliest words are functional units whose phonetic components are only vaguely specified and are not transferable to other contexts. Each word may have its own (probably primitive) schema. As the lexicon increases the whole-word schemata are replaced for reasons of economy by schemata for consonants, vowels, or perhaps syllables. As segmental production is increasingly controlled, motor schemata are refined so that we see increasing coarticulation, reduction in spectral and temporal variability, and increasing control of rate of articulation.

Although some of the predictions of schema theory have not been fully supported experimentally (Schmidt, 1976; Kent, 1982), the theory explains many specific characteristics of the developing motor control of speech as well as being compatible with the assumptions discussed in Sections IV and V concerning general influences on speech development. At the same time, the theory places articulation and its developing control within the broader context of movement. Since schemata can be formed for units of any size, the theory agrees with the suggestion that the production of speech involves parallel processing of several different types of unit. Additionally, it provides a possible explanation for variation in the degree to which each articulatory parameter is specified. Canonical forms postulated by phonologists can be interpreted in terms of primitive schemata, and the concept of modules making up a skilled movement is easily accommodated. A major strength of the theory is that it explicitly accounts for flexibility in the selection of specific movements. Because it assumes that current behavior is programmed from interpolations between memories of past initial conditions and past outcomes, schema theory accounts not only for intrasubject variability but also for intersubject variability. It also predicts more variability in

the child than in the older speaker, since an inexperienced speaker has fewer past movements to contribute to the schema and hence a less well-defined "field" from which to select the most efficient movement.

In Section IV, I suggested that the child's task involved applying similar routines to new and more complex domains rather than to learning new routines. As an abstraction of the relationships between input, output, and motor programs, the schema represents a mode of learning from past performance. It can apply to any unit of production from the simple ones of the unskilled speaker to the complex ones of the skilled speaker. Hence the schema makes the term "routine" more explicit. The schema is also compatible with my earlier assumption that the role played by different processes may change during maturation. The role of feedback in the recognition schema is such that it is possible that auditory feedback could be overtaken in importance by kinesthetic and proprioceptive feedback as the skill develops. Kent (1982) discusses some of the consequences of changes in the role of different types of feedback.

Section V argued that neuromotor maturation interacts with the development of cognitive-linguistic understanding of phonetic contrasts in determining whether or not a particular phonetic form will be observed in the child's speech. This too can be interpreted in terms of schema theory. To form a schema the child must discover relationships between initial conditions, movements, and outcomes (including feedback) and she must develop expectations concerning the results of their permutations. This must be an active process involving at some level awareness of cause and effect. With the exception of meaningless babble, it is unlikely that such a process could occur in the absence of an understanding that certain movements and their acoustic consequences have linguistic significance. Equally, an effective schema cannot form if the child is as yet incapable of producing the required movements; initial conditions would be randomly related to outcomes so that no motor program could be designed for future execution. (Sensory consequences would not be randomly related to movements and outcomes, so the recognition schema may be the starting point for input to what will ultimately be the recall schema, but only, of course, after the child can make the necessary movements at least some of the time.)

The notions of reprogramming and concurrent programming can also be interpreted within the framework of schema theory. One interpretation is that reprogramming pauses represent a shift from chaining successive consonants in a system of closed-loop feedback to producing them in an open-loop system as subcomponents of larger units. Assuming that maintaining the same articulation requires less attention than changing to a different one, the general segmental slowing of concurrent programming also fits this explanation. In terms of schema theory, schemata for these consonant clusters are evidently being developed. As we are assuming that feedback is involved, the first changes should affect the

recognition schema. One possibility is that during the stop closure the child is recording sensory input to develop an internal representation of the sensory effect of the preceding articulation (cf. Schmidt, 1975, 1976; Semjen, 1977). A related possibility is that the child may be taking the extra time to adjust a comparator between expected and actual sensory feedback. It will be recalled that in clusters subject to reprogramming, temporal control of the segments surrounding the stop progresses at a faster rate than control of the same segments in other contexts. These changes may reflect initial development of the recall schema for the cluster, or perhaps of two such schemata which have yet to be united into one higher order schema. As the recall schema is more firmly established, reliance on feedback decreases so subsequent executions of the gesture do not appreciably affect the recognition schema, and the segments are no longer lengthened.

To summarize, schema theory seems to be a useful approach to considering how an articulatory gesture is first used to mediate changes in meaning, how the gesture changes from being effortful to becoming automatic, and why we find so much diversity in the routes that children take to becoming skilled speakers. As mentioned earlier, the predictions of schema theory have not always been supported by experimental results, so it is unlikely that the theory as presented here will stand up to thorough investigation without needing changes. However, it does provide a much needed systematic framework for the study of developing articulatory control, and it has the added advantage of having been developed for movement in general rather than specifically for articulation. Thus we can expect to see necessary modifications being made more rapidly as the theory should generate research in a number of disciplines. These advantages make the theory a promising avenue for future research.

VIII. DIRECTIONS FOR RESEARCH

One of the aims of this article has been to focus attention on aspects of the development of speech timing that are particularly in need of investigation. This section summarizes the research questions that have been raised in earlier sections.

One of the most important questions, which is also one of the most difficult to investigate, concerns the perceptual prerequisites for a child to begin to use sounds meaningfully and ultimately to use them as the adult does. Reasoning about relationships between perception and production currently tends to be circular. The circularity will not be broken until we discover more about the child's **linguistically interpreted** perception. This requires working with children aged between about 12 and 36 months in tasks that require not only discrimination but also identification of sound with meaning. Although some experiments of this type have been attempted (e.g., Barton, 1976; Eilers & Oller, 1976;

Strange & Broen, 1980; Broen, Strange, Doyle, & Heller, 1980), the formidable methodological problems involved have not yet been surmounted.

Related to the issue of perception is the need to establish the biomechanical and aerodynamic constraints under which the child is working, including the effects of speaking with a vocal tract whose components are changing their relative size as well as growing larger. A start has been made in this direction (e.g., Kent, 1976; Goldstein, 1980) but we need much more basic research of this type. With a deeper understanding of basic perceptual and productive constraints we will be able to approach more confidently questions such as to what extent the child's failure to match adult speech patterns is due to external constraints on her productive system versus the particular way she is setting up that system.

A third issue of perception concerns the status of the distinction made in this article between primary and secondary perceptual cues to a phonological contrast. Bernstein (1979) and Greenlee (1980) have shown that children identify obstruent consonants more accurately if more than one potential cue is present. Bernstein, who varied VOT and f_0 in stops independently, suggested from her results that it is incorrect to treat f_0 as a developmentally secondary cue to learning the voicing distinction. Greenlee varied vowel duration preceding a voiceless obstruent, as well as presenting natural productions of words with either voiced or voiceless obstruents. She found that 3 year olds could not use vowel duration alone as a cue to final voicing, although 6 year olds could. (Both 3 and 6 year olds, however, performed best on the naturally produced words.)

Interpretation of these types of data seems to rest largely upon one's orientation to the original issue. There can be no doubt that children (and most adults too) will be most consistent in identifying natural rather than synthetic or artificially modified speech. A model for perception that emphasizes multiplicity of levels of processing and parallel channels would predict this, assuming that synthetic speech generally lacks some of the properties of natural speech. The child's unwillingness or inability to perform optimally with reduced cues may reflect the fragility of her perceptual criteria for an identification. These criteria presumably become more robust and flexible with increasing exposure to speech, especially under poor listening conditions.

A second consideration in the issue of primary versus secondary perceptual cues is that it may be necessary to distinguish between a cue that optionally accompanies some other distinction and a cue that inevitably occurs with that distinction. Falling f_0 after a voiceless stop appears to be a consequence of the glottal configuration at release, for example (cf. Gandour, 1974). Similarly, there is some evidence that vowel lengthening before a voiced stop may be unavoidable, although at the same time it has been suggested that languages such as English may exaggerate the distinction (Lisker, 1974). So-called secondary cues that the child herself produces may have a different status for her than cues

she does not produce. The issue of secondary perceptual cues is an interesting one to investigate with adults; its status in development is no less important and in addition could help clarify an interpretational framework for children's production.

Turning from issues of perception to those of production, one crucial question for research concerns variability. There are two aspects to this: variability between and within speakers at any one time, and short-term versus long-term variability across time. Both types have been discussed earlier in this article, so little need be said here. With respect to the first type, we need to assess the significance of the variability we found and incorporate it into theories of production. One problem not mentioned earlier is that of deciding whether the child "intended" a particular form or whether it was an error. Error analyses can, of course, contribute to a theory of production, but it is crucial to know just what is an error. Unfortunately, we do not have a way of asking this of young children.

The second type of variation—long-term versus short-term—brings us up against the problem of identifying trends. This in turn raises the issue of when cross-sectional studies are appropriate and when longitudinal studies must be made. Our tendency to conduct cross-sectional experiments and to use time-free statistics encourages us to search for "stages" of development when in fact the stages may not really exist; in any case, progression between stages may provide more insight into developmental processes. Longitudinal studies and time-based trend analyses may furnish richer insights in many cases.

Longitudinal analyses are, in any case, necessary in order to study many aspects of developing motor skills in speech. Reprogramming and concurrent programming, for example, require longitudinal studies before they can be accepted as strategies that the child uses. Cross-sectional studies also play a role, especially if the ages of the children studied are selected nonarbitrarily. For example, the discussion in Section V,C,1 of age-related changes in dominance of timing-based and articulation-based systems of production would be appropriate for cross-sectional experimentation. Closely connected with studies of articulation- versus timing-dominant systems would be developmental studies of feedback in speech. Although some developmental work has been done (e.g., MacKay, 1967; Garber & Speidel, 1978; Siegel, Fehst, Garber, & Pick, 1980; Yanez, Siegel, Garber, & Wellen, 1982), it has not generally been related to temporal issues. Such studies are central to mapping possible changes between articulation and timing dominance as well as to more general issues of the development of schemata for other preprogrammed sequences, and organization of systems from closed- to open-loop control. Cross-sectional analyses might be appropriate here as long as they were sensitive to the possibility of variation between children in the relative importance of different modalities of feedback and also in the particular stage of articulatory development the child has reached, regardless of chronological age.

Certain other aspects of production could usually be safely studied cross-sectionally even if the chosen ages of groups were not ideal. One such area is the interaction of word stress with phonotactically determined differences in segmental duration. Kent and Forner (1980) have pointed out the need for such studies, and the work of Heusner and Hoffman (1980) suggests that this type of study could enrich our understanding of the interconnections between the development of articulation and of the semantic-syntactic system. Cross-sectional studies could in fact be preferable to longitudinal ones, at least in early investigations, because parameters of both segmental duration and stress are continuous and are influenced in complex ways by other parameters. Using discrete age groups could therefore help clarify the main developmental sequence.

Finally, we need to study children learning languages with different rhythmic and phonotactic constraints from English (cf. Allen, 1983) so that we can separate general characteristics of articulatory development from language-specific ones. Studies of children growing up bilingually could also provide important information about the developing motor control of speech.

IX. SOME UNRESOLVED ISSUES

The preceding two sections discussed theoretical and practical issues specific to the temporal control of speech and its development. This section addresses broader issues: most concern problems in the study of all aspects of children's speech and language development, and some concern problems in studying the motor control of speech whether in development, maturity, or disintegration. As the section title implies, no answers will be offered, but despite this I believe that the questions can usefully be stated.

A major and very pervasive problem is that of distinguishing between phonological and phonetic processes in children's speech. This issue was touched upon in discussing the development of the voicing distinction (Section V,A). It is formally similar to the issue of predicting the order of development for contrasts which function as primary versus secondary perceptual cues. Whereas we would expect phonetic contrasts to show up relatively early when they are mediated by automatic aerodynamic or biomechanical consequences of particular movements, nonobligatory properties should appear in an order reflecting phonological as well as neuromuscular development. It is commonplace for phonologists to distinguish between a phonological contrast involving single phonemes and its phonetic expression in the child's speech, but the issue is not necessarily restricted to such simple situations. An example of a more complex situation is that of the phonological status of consonant clusters. Clements and Keyser (1981) have argued that three-segment consonant clusters in English are derived from the amalgamation of two clusters, each having two segments. So,

for example, /spr/ is derived from /sp/ plus /pr/. What then do we make of the developmental evidence? Menyuk and Klatt (1968), Kornfeld (1971), Menyuk (1971), and Barton *et al.* (1980) have argued that consonant clusters are single **phonological** units for the very young child, and I have based much of my discussion of reprogramming and concurrent programming on the assumption that in the early stages of producing all segments of a cluster, each segment is a relatively discrete articulatory unit in terms of preprogramming, and that the two hypothesized strategies represent processes in their unification into single **articulatory** units. No claims were made about their phonological status. None of these analyses (Clements and Keyser's; Menyuk's, Kornfeld's, and Barton's; or mine) are necessarily incompatible with each other, but they do indicate a need to address issues of the relationship between phonology, phonetics, and motor control.

Related to this issue is the necessity of distinguishing the level of phonetics or phonology to which suggested processes apply. Reprogramming, for example, has been suggested for the integration of segments into syllables. Do pauses that occur between other units reflect a similar process or should they be analyzed in a different way? A pause preceding a modification in phrase structure may reflect some process of reprogramming, but one of quite a different nature from the way reprogramming was defined as an articulatory strategy in Section VI. On the other hand, a pause preceding an unusually successful attempt at a single word may reflect a process similar to that suggested for consonant clusters.

All studies of the motor control of speech, regardless of their immediate field of inquiry, must make some assumptions about whether the same neural mechanisms subserve both speech and nonspeech movements. There are two aspects to this issue, concerning whether speech and nonspeech movements of the same or different musculature are involved. That is, first, what are the similarities and differences between control of nonspeech activities such as chewing, swallowing, and spitting, and control of the same organs for speech; and second, are neuromuscular principles derived for limb movements (for example) also applicable to speech articulation, or is the speech musculature a special case? The implicit assumption of our theoretical discussion and application of schema theory to development (Section VII) was that there are general principles of control applying to all voluntary movement. Even if it is true, this assumption may apply only at a very abstract level of analysis; there may be significant differences in the details of control either of different "levels" of the nervous system (e.g., motor cortex versus cranial or spinal nerves) or of different motor systems such as the respiratory or laryngeal systems and the upper articulators, as well as nonspeech musculature such as limbs or fingers. This clearly must be a point of concern in any future theoretical development.

Any interpretation of children's behavior that relies on comparisons with that of adults can be valid only insofar as we understand the laws governing adults'

behavior. Unfortunately, this is not always the case in speech and several of the interpretations of data presented in this article suffer from this problem. If we do not know why in adults' speech the duration of /s/, for example, differs depending on what initial cluster it is in, then it is hard to discover why children fail to produce such differences. On the other hand, a reasonable first step in understanding underlying processes is to describe what is observed, and this has been done for much of adult speech. The data presented here come in large part from similar descriptions of children's speech. We can take description one step nearer to explanation by describing the data in terms of general traits, such as (for children) failure to differentiate between contexts. It may also be argued that tracing the development of a behavior can contribute towards an understanding of its underlying structure. Far from being dependent on the construction of adequate theories for the adult, then, developmental studies can provide crucial evidence for such theories. Thus, in the long term, concurrent study of behavior in development (and in dissolution) as well as in its mature state can only enrich our understanding. In the short term, however, the absence of an adequate theory for the adult makes the study of development more difficult and more speculative. For speech timing, our lack of understanding of adult processes is particularly evident in the areas of rhythm (and prosody in general), and aerodynamic and biomechanical constraints on segmental duration. Developmental studies are similarly hampered by our current inability to specify even for the adult what it means to be a "unit" of speech, and how one would identify such a unit. There are similar problems in identifying indices of different systems of production, such as articulation- or timing-dominant systems or hybrid systems. Cross-fertilization between disciplines helps answer questions such as these, but as yet we are a long way from solutions.

The last point is philosophical: the way that we organize the data we observe may not correspond to the way the brain organizes it. For example, we might construct a hierarchy of units of production, perhaps of words composed of syllables, which themselves are composed of phonemes, each of which is translated into a number of articulatory gestures. This hierarchy might describe our data well, but we would not thereby have proved that the brain works with such a hierarchy. The brain could in fact function in quite a different way with the same units, or the units themselves might differ. This problem merits more attention than space allows us to devote to it here. It is mentioned because it is an important issue in the philosophy of science and one which is often ignored in practice.

Acknowledgments

A number of individuals, particularly colleagues in the Speech Communication Group at the Massachusetts Institute of Technology, contributed to this article by discussion, reading earlier

drafts, and/or helping to prepare the manuscript. I would especially like to thank John Gilbert, Edith Maxwell, Lise Menn, Joseph Perzell, Marie Southwick, and Kenneth Stevens for their help.

References

- Adams, J. A. Issues for a closed-loop theory of motor learning. In G. E. Stelmach (Ed.), *Motor control: Issues and trends*. New York: Academic Press, 1976. Pp. 87-107.
- Allen, G. D. The location of rhythmic stress beats in English: An experimental study. Parts I and II. *Language and Speech*, 1972, 15, 72-100 and 179-195.
- Allen, G. D. Segmental timing control in speech production. *Journal of Phonetics*, 1973, 1, 219-237.
- Allen, G. D. Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics*, 1975, 3, 75-86.
- Allen, G. D. Linguistic experience modifies lexical stress perception. *Journal of Child Language*, 1983, 10, 535-549.
- Allen, G. D., & Hawkins, S. Trochaic rhythm in children's speech. In H. Hollien & P. Hollien (Eds.), *Current issues in the phonetic sciences*. Amsterdam: Benjamins, 1979. Pp. 927-933.
- Allen, G. D., & Hawkins, S. Phonological rhythm: Definition and development. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 1). New York: Academic Press, 1980. Pp. 227-256.
- Bailey, P. J., & Haggard, M. P. Perception-production relations in the voicing contrast for initial stops in 3-year-olds. *Phonetica*, 1980, 37, 377-396.
- Barton, D. P. *The role of perception in the acquisition of phonology*. Unpublished doctoral dissertation, University of London, 1976 (also published by Indiana University Linguistics Club, 1978).
- Barton, D., & Macken, M. A. An instrumental analysis of the voicing contrast in word-initial stops in the speech of four-year-old English-speaking children. *Language and Speech*, 1980, 23, 159-169.
- Barton, D., Miller, R., & Macken, M. A. Do children treat clusters as one unit or two? *Stanford University Papers and Reports on Child Language Development*, 1980, 18, 105-137.
- Bernstein, L. E. Developmental differences in labeling VOT continua with varied fundamental frequency. *Journal of the Acoustical Society of America*, 1979, 65, Suppl. 1 (Abstract).
- Bernstein, N. A. *The coordination and regulation of movements*. Oxford, England: Pergamon, 1967.
- Bickley, C. A. Acoustic evidence for the phonological development of vowels in young children. In A. Cohen and M. P. R. van den Broecke (Eds.), *Abstracts of the tenth international congress of phonetic sciences*. Dordrecht: Foris, 1983. P. 624.
- Blessner, B. A. *Perception of spectrally rotated speech*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1969.
- Branigan, G. Some reasons why successive single word utterances are not. *Journal of Child Language*, 1979, 6, 411-421.
- Branigan, G., & Stokes, W. *An integrated account of utterance variability in early language development*. Paper presented at the 2nd International Congress for the Study of Child Language, Vancouver, Canada, 1981.
- Broen, P. A., Strange, W., Doyle, S. S., & Heller, J. H. *Perception of approximant consonants by normal and articulation-delayed preschool children*. Paper presented at the annual convention of the American Speech-Language-Hearing Association, Detroit, 1980.
- Brown, R. *A first language: The early stages*. London: Allen & Unwin, 1973.
- Bruner, J. S. Organization of early skilled action. *Child Development*, 1973, 44, 1-11.
- Clements, G. N., & Keyser, S. J. A three-tiered theory of the syllable. *Center for Cognitive Science, MIT, Occasional Paper*, 1981. 19.

- Clumeck, H., Barton, D., Macken, M. A., & Huntington, D. The aspiration contrast in Cantonese word-initial stops: Data from children and adults. *Stanford University Papers and Reports on Child Language Development*, 1979, 18.
- Cohen, A., & Nootboom, S. G. (Eds.). *Structure and process in speech perception*. Berlin and New York: Springer-Verlag, 1975.
- Collier, R., & 't Hart, J. The role of intonation in speech perception. In A. Cohen & S. G. Nootboom (Eds.), *Structure and process in speech perception*. Berlin and New York: Springer-Verlag, 1975. Pp. 107-123.
- Connolly, K. The nature of motor skill development. *Journal of Human Movement Studies*, 1977, 3, 128-143.
- Cooper, W. E. The development of speech timing. In S. J. Segalowitz & F. A. Gruber (Eds.), *Language development and neurological theory*. New York: Academic Press, 1977. Pp. 357-373.
- Cruttenden, A. An experiment involving comprehension of intonation in children from 7 to 10. *Journal of Child Language*, 1974, 1, 221-231.
- Crystal, D. Nonsegmental phonology in language acquisition: A review of the issues. *Lingua*, 1973, 32, 1-45.
- Crystal, D. Prosodic development. In P. Fletcher & M. Garman (Eds.), *Studies in language acquisition*. London and New York: Cambridge Univ. Press, 1978. Pp. 33-48.
- Cutler, A. Phoneme monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics*, 1976, 20, 55-60.
- Cutler, A., & Foss, D. J. *The importance of lexical item stress for lexical access*. Paper presented at the 44th annual meeting of the Midwestern Psychological Association, Chicago, 1973.
- Cutler, A., & Foss, D. J. On the role of sentence stress in sentence processing. *Language and Speech*, 1977, 20, 1-10.
- Darwin, C. J. On the dynamic use of prosody in speech perception. In A. Cohen and S. G. Nootboom (Eds.), *Structure and process in speech perception*. Berlin and New York: Springer-Verlag, 1975. Pp. 178-194.
- de Boysson-Bardies, B., Bacri, N., Sagart, L., & Poizat, M. Timing in late babbling. *Journal of Child Language*, 1981, 8, 525-539.
- Delack, J. B., & Fowlow, P. J. The ontogenesis of differential vocalization: Development of prosodic contrastivity during the first year of life. In N. Waterson & C. Snow (Eds.), *The development of communication*. New York: Wiley, 1978. Pp. 93-110.
- Denckla, M. B. Development of motor coordination in normal children. *Developmental Medicine and Child Neurology*, 1974, 16, 729-741.
- DiSimoni, F. G. Effect of vowel environment on the duration of consonants in the speech of 3-, 6- and 9-year-old children. *Journal of the Acoustical Society of America*, 1974, 55, 360-361. (a)
- DiSimoni, F. G. Influence of consonant environment on duration of vowels in the speech of 3-, 6- and 9-year-old children. *Journal of the Acoustical Society of America*, 1974, 55, 362-363. (b)
- Eguchi, S., & Hirsh, I. J. Development of speech sounds in children. *Acta Otolaryngologica*, 1969, Suppl., 257, 1-51.
- Eilers, R. E. Infant speech perception: History and mystery. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 2). New York: Academic Press, 1980. Pp. 23-39.
- Eilers, R. E., & Oller, D. K. The role of speech discrimination in developmental sound substitutions. *Journal of Child Language*, 1976, 3, 319-329.
- Eimas, P. D. Auditory and linguistic processing of cues for place of articulation by infants. *Perception and Psychophysics*, 1974, 16, 513-521.
- Eimas, P. D. Auditory and phonetic coding of the cues for speech: Discrimination of the r-l distinction by young infants. *Perception and Psychophysics*, 1975, 18, 341-347.

- Eimas, P. D., Siquelund, E. R., Jusczyk, P., & Vigorito, J. Speech perception in early infancy. *Science*, 1971, 171, 303-306.
- Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 1954, 47, 381-391.
- Flege, J. E., Brown, W. S., & White, K. Voicing control in speech. In T. Bennett-Kastor (Ed.), *Mid-America linguistics conference proceedings*. English Department, Wichita State University, Wichita, Kansas, 1981. Pp. 11-30.
- Fowler, C. A. Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 1980, 8, 113-133.
- Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production*. New York: Academic Press, 1980. Pp. 373-420.
- Fox, B., & Routh, D. K. Analyzing spoken language into words, syllables, and phonemes: A developmental study. *Journal of Psycholinguistic Research*, 1975, 4, 331-341.
- Fry, D. B. The development of the phonological system in the normal and the deaf child. In F. Smith & G. Miller (Eds.), *The genesis of language*. Cambridge, Massachusetts: MIT Press, 1966. Pp. 187-206.
- Gandour, J. Consonant types and tones in Siamese. *Journal of Phonetics*, 1974, 2, 337-350.
- Garber, S. R., & Speidel, T. M. The effects of palatal appliances on the speech of five-year-old children. *Journal of the Acoustical Society of America*, 1978, 63, Suppl. 1 (Abstract).
- Gilbert, J. H. V. A voice onset time analysis of apical stop production in three year olds. *Journal of Child Language*, 1977, 4, 103-110.
- Gilbert, J. H. V., & Purves, B. A. Temporal constraints in consonant clusters in child speech production. *Journal of Child Language*, 1977, 4, 417-432.
- Goldstein, U. G. An articulatory model for the vocal tracts of growing children. Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1980.
- Greenlee, M. Learning the phonetic cues to the voiced-voiceless distinction: A comparison of child and adult speech perception. *Journal of Child Language*, 1980, 7, 459-468.
- Haggard, M. P. Abbreviation of consonants in English pre- and post-vocalic clusters. *Journal of Phonetics*, 1973, 1, 9-24.
- Hawkins, S. Temporal coordination of consonants in the speech of children: Preliminary data. *Journal of Phonetics*, 1973, 1, 181-217.
- Hawkins, S. *The developing organization of speech production in children: Evidence from consonant clusters*. Unpublished doctoral dissertation, University of Cambridge, 1976.
- Hawkins, S. Temporal coordination of consonants in the speech of children: Further data. *Journal of Phonetics*, 1979, 7, 235-267. (a)
- Hawkins, S. Processes in the development of speech timing control. In H. Hollien & P. Hollien (Eds.), *Current issues in the phonetic sciences*. Amsterdam: Benjamins, 1979. Pp. 927-933. (b)
- Hawkins, S., & Allen, G. D. Judgements of speech maturity: Correlations with chronological age and acoustic measurements. *Journal of the Acoustical Society of America*, 1977, 62, Suppl. 1 (Abstract).
- Hawkins, S., & Allen, G. D. Acoustic-phonetic features of stressed syllables in the speech of 3-year-olds. *Journal of the Acoustical Society of America*, 1978, 63, Suppl. 1 (Abstract).
- Heusner, D. M., & Hoffman, P. *Acoustic characteristics of children's articulation*. Paper presented at the annual convention of the American Speech-Language-Hearing Association, Detroit, 1980.
- Higgs, J. W., & Hodson, B. W. Phonological perception of word-final obstruent consonants. *Journal of Phonetics*, 1978, 6, 25-35.

- Huggins, A. W. F. On the perception of temporal phenomena in speech. *Journal of the Acoustical Society of America*. 1972, **51**, 1279-1290.
- Huggins, A. W. F. Speech timing and intelligibility. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, New Jersey: Erlbaum, 1978. Pp. 279-297.
- Ingram, D. Phonological rules in young children. *Journal of Child Language*. 1974, **1**, 49-64.
- Ingram, D. *Phonological disability in children*. Amsterdam: Elsevier, 1976.
- Ingram, D. Cross-linguistic evidence on the extent and limit of individual variation in phonological development. *Proceedings of the Ninth International Congress of Phonetic Sciences, Copenhagen, 1979*, **II**, 150-154. (a)
- Ingram, D. Phonological patterns in the speech of young children. In P. Fletcher & M. Garman (Eds.), *Language acquisition*. London and New York: Cambridge Univ. Press, 1979. Pp. 133-148. (b)
- Jørgensen, J. The acquisition of stop consonants by Danish two-year-old children. *Annual Report of the Institute of Phonetics, University of Copenhagen*, 1979, **13**, 9-25.
- Kent, R. D. Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, 1976, **19**, 421-447.
- Kent, R. D. *Motor skill component of speech development*. Paper presented at the annual convention of the American Speech-Language-Hearing Association, Detroit, 1980.
- Kent, R. D. Sensorimotor aspects of speech development. In R. N. Aslin, J. R. Alberts, & M. R. Peterson (Eds.), *Development of perception: Psychological factors*. New York: Academic Press, 1982. Pp. 161-189.
- Kent, R. D., & Forner, L. L. Developmental study of vowel formant frequencies in an imitation task. *Journal of the Acoustical Society of America*, 1979, **65**, 208-217.
- Kent, R. D., & Forner, L. L. Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics*, 1980, **8**, 157-168.
- Kewley-Port, D., & Preston, M. Early apical stop productions: A voice onset time analysis. *Journal of Phonetics*, 1974, **2**, 195-210.
- Kirk, L. An analysis of speech imitations by Gã children. *Anthropological Linguistics*, 1973, **15**, 267-275.
- Klatt, D. H. Durational characteristics of prestressed word-initial consonant clusters in English. *MIT Research Laboratory of Electronics Quarterly Progress Report*. 1973, **108**, 253-260.
- Klatt, D. H. The duration of /s/ in English words. *Journal of Speech and Hearing Research*, 1974, **17**, 51-63.
- Klatt, D. H. Voice-onset time, frication and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 1975, **18**, 686-706.
- Klatt, D. H. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 1976, **59**, 1208-1221.
- Kornfeld, J. R. What initial consonant clusters tell us about a child's speech code. *MIT Research Laboratory of Electronics Quarterly Progress Report*, 1971, **101**, 218-221.
- Kozhevnikov, V. A., & Chistovich, L. A. Speech: Articulation and perception. *Joint Publications Research Service, Washington, D.C.*, 1965, **30**, 543.
- Kubaska, C. A., & Keating, P. A. Word duration in early child speech. *Journal of Speech and Hearing Research*, 1981, **24**, 615-621.
- Kuehn, D. P., & Moll, K. L. A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics*. 1976, **4**, 303-320.
- Kuhl, P. K. The perception of speech in early infancy. In N. Lass (Ed.), *Speech and language: Advances in basic research and practice*. (Vol. 1). New York: Academic Press, 1979. Pp. 1-47.
- Lehiste, I. *Suprasegmentals*. Cambridge, Massachusetts: MIT Press, 1970.

- Leonard, L. B., Newhoff, M., & Mesalem, L. Individual differences in early child phonology. *Applied Psycholinguistics*, 1980, 1, 7-30.
- Lindblom, B., & Rapp, K. Reexamining the compensatory adjustment of vowel duration in Swedish words. *Stockholm, KTH. Speech Transmission Laboratory Quarterly Progress and Status Report*, 1971, 4, 19-25.
- Lisker, L. On explaining vowel duration variation. *Haskins Laboratories Status Report on Speech Research*, 1974, 37/38, 225-232.
- MacKay, D. G. Metamorphosis of critical interval: Age-linked changes in the delay in auditory feedback that produces maximal disruption of speech. *Journal of the Acoustical Society of America*, 1967, 43, 811-821.
- Macken, M. A. Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua*, 1979, 49, 11-49.
- Macken, M. A., & Barton, D. The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, 1979, 7, 41-74.
- Macken, M. A., & Barton, D. The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language*, 1980, 7, 433-458.
- MacNeilage, P. F. Speech production. *Language and Speech*, 1980, 23, 3-22.
- MacNeilage, P. F., Hutchinson, J. A., & Lasater, S. A. The production of speech: Development and dissolution of motoric and premotoric processes. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX*. Hillsdale, New Jersey: Erlbaum, 1981. Pp. 503-519.
- Martin, J. G. Rhythmic (hierarchical) vs. serial structure in speech and other behavior. *Psychological Review*, 1972, 79, 487-509.
- Menn, L. Phonological units in beginning speech. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments*. Amsterdam: North-Holland Publ., 1978. Pp. 157-171.
- Menn, L. Transition and variation in child phonology: Modeling a developing system. *Proceedings of the 9th International Congress of Phonetic Sciences (Copenhagen)*, 1979, II, 169-175.
- Menyuk, P. Clusters as single underlying consonants: Evidence from children's productions. In A. Rigault & R. Charbonneau (Eds.), *Proceedings of the 7th international congress of phonetic sciences*. The Hague: Mouton, 1971. Pp. 1161-1165.
- Menyuk, P., & Klatt, D. H. Child's production of initial consonant clusters. *MIT Research Laboratory of Electronics Quarterly Progress Report*, 1968, 91, 205-213.
- Menyuk, P., & Klatt, M. Voice onset time in consonant cluster production by children and adults. *Journal of Child Language*, 1975, 2, 223-231.
- Naeser, M. A. The American child's acquisition of differential vowel duration. *University of Wisconsin Research and Development Center for Cognitive Learning, Madison, Technical Report* 1970, 144.
- Nelson, K. *Individual differences in language development: Implications for development and language*. Paper presented at the fifth annual Boston University Conference on Language Development, 1980.
- Nooteboom, S. G. Temporal patterns in Dutch. In A. Rigault & R. Charbonneau (Eds.), *Proceedings of the 7th international congress of phonetic sciences*. The Hague: Mouton, 1973, Pp. 984-989.
- Nooteboom, S. G., Brokx, J. P., & de Rooij, J. J. Contributions of prosody to speech perception. *Eindhoven, Holland: IPO Annual Progress Report*, 1976, 11, 34-54.
- Ohala, J. J. Aspects of the control and production of speech. *UCLA Working Papers in Phonetics*, 1970, 15.
- Ohala, J. J. The temporal regulation of speech. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech*. New York: Academic Press, 1975. Pp. 431-453.

- Oller, D. K. The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 1). New York: Academic Press, 1980. Pp. 93-111.
- Oller, D. K., & Smith, B. L. The effect of final-syllable position on vowel duration in infant babbling. *Journal of the Acoustical Society of America*, 1977, **62**, 994-997.
- O'Shaughnessy, D. Consonant durations in clusters. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1974, **22**, 282-295.
- Peters, A. Language learning strategies: Does the whole equal the sum of the parts? *Language*, 1977, **53**, 560-573.
- Pisoni, D. B., & Sawusch, J. R. Some stages of processing in speech perception. In A. Cohen and S. G. Neebom (Eds.), *Structure and process in speech perception*. New York, Springer-Verlag, 1975. Pp. 16-35.
- Priestley, T. M. S. One idiosyncratic strategy in the acquisition of phonology. *Journal of Child Language*, 1977, **4**, 45-66.
- Schmidt, R. A. A schema theory of discrete motor skill learning. *Psychological Review*, 1975, **82**, 225-260.
- Schmidt, R. A. The schema as a solution to some persistent problems in motor learning theory. In G. E. Stelmach (Ed.), *Motor control: Issues and trends*. New York: Academic Press, 1976. Pp. 41-65.
- Semjen, A. From motor learning to sensorimotor skill acquisition. *Journal of Human Movement Studies*, 1977, **3**, 182-191.
- Shields, J. L., McHugh, A., & Martin, J. G. Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, 1974, **102**, 250-255.
- Siegel, G. M., Fehst, C. A., Garber, S. R., & Pick, H. L. Delayed auditory feedback with children. *Journal of Speech and Hearing Research*, 1980, **23**, 802-813.
- Simon, C., & Fourcin, A. J. Cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America*, 1978, **63**, 925-935.
- Slobin, D. I. Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development*. Holt, 1973. Pp. 175-208.
- Smith, B. L. Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics*, 1978, **6**, 37-67.
- Smith, N. V. *The acquisition of phonology: A case study*. London and New York: Cambridge Univ. Press, 1973.
- Spring, D. R., & Dale, P. S. Discrimination of linguistic stress in early infancy. *Journal of Speech and Hearing Research*, 1977, **20**, 224-232.
- Stark, R. E. Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 1). New York: Academic Press, 1980. Pp. 73-92.
- Stevens, K. N. Airflow and turbulent noise for fricative and stop consonants. *Journal of the Acoustical Society of America*, 1971, **50**, 1180-1192.
- Strange, W., & Broen, P. A. Perception and production of approximant consonants by 3-year-olds: A first study. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 2). New York: Academic Press, 1980. Pp. 117-154.
- Strange, W., & Broen, P. A. The relationship between perception and production of /w/, /t/, and /l/ by 3-year-old children. *Journal of Experimental Child Psychology*, 1981, **31**, 81-102.
- Suen, C. Y., & Beddoes, M. P. The silent interval of stop consonants. *Language and Speech*, 1974, **17**, 126-134.
- Svensson, S. G. Prosody and grammar in speech perception. *Monographs from the Institute of Linguistics, University of Stockholm*, 1974, **2**.

- Tingley, B., & Allen, G. A. Development of speech timing control in children. *Child Development*, 1975, 46, 186-194.
- Wallach, M. A., & Wallach, L. *Teaching all children to read*. Chicago, Illinois: Univ. of Chicago Press, 1976.
- Waterson, N. Some speech forms of an English child: A phonological study. *Transactions of the Philological Society, London*: 1970, 1-24.
- Waterson, N. Child phonology: A prosodic view. *Journal of Linguistics*, 1971, 7, 179-211. (a)
- Waterson, N. Child phonology: A comparative study. *Transactions of the Philological Society, London*, 1979, 34-50. (b)
- Wilbur, R. Theoretical phonology and child phonology: Argumentation and implications. In D. Goyvaerts (Ed.), *Phonology in the 1980's*. Ghent: Story-Scientia, 1981. Pp. 403-429.
- Wingfield, A., & Klein, J. F. Syntactic structure and acoustic pattern in speech perception. *Perception and Psychophysics*, 1971, 9, 23-25.
- Yanez, E. A., Siegel, G. M., Garber, S. R., & Wellen, C. J. The effects of different masking noises on children with /s/ or /t/ errors. *Journal of Speech and Hearing Disorders*, 1982, 47, 150-153.
- Zlatin, M. A., & Koenigsnecht, R. A. Development of the voicing contrast: Perception of stop consonants. *Journal of Speech and Hearing Research*, 1975, 18, 541-553.
- Zlatin, M. A., & Koenigsnecht, R. A. Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research*, 1975, 19, 93-111.