Journal of Phonetics (1977) 5, 115–133

Coarticulation in recent speech production models

R. D. Kent

Department of Communicative Disorders, University of Wisconsin, Madison, Wisconsin 53706, *U.S.A.*

and

F. D. Minifie

Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington 98195, *U.S.A.*

Received 17*th July* 1976

Abstract: Whereas speech usually is considered to be a sequence of discrete units or segments (such as phonemes), attempts to impose segment boundaries on articulatory events are quickly confounded by the fact that articulatory movements overlap one another in complex fashion. This difficulty gave rise to the concept of coarticulation, which is a conceptualization of speech behavior that implies (1) discrete and invariant units serving as input to the system of speech production, and (2) an eventual obscuration of the boundaries between units at the articulatory or acoustic levels. Models of speech production with discrete control units must reconcile the presumed discreteness of the input to the production model with the demonstrated existence of complex, overlapping, and somewhat variable articulations at the output of the model. This paper reviews recent speech production models with particular attention to the problem of coarticulatory variation. It is concluded that coarticulatory patterns described in the experimental literature are not explained fully by any of these models. The inadequacies of individual models or theories are discussed, and some general suggestions on the form of an hierarchical model of speech production are given.

Introduction

One of the most important but least understood aspects of motor behavior is that of serial ordering. In the study of speech production this problem is complicated by the fact that different levels of seriation may be explored. For example, one might inquire about the sequencing of syllables, or of phonemes, or of articulatory gestures, or of muscle contractions. Several papers have been written about each of these levels of seriation, but it remains a highly challenging task to consolidate them into a coherent picture of the serial ordering processes involved in speech production.

One level of the seriation problem is illustrated by the speech sequencing error, of which the spoonerism is one example. Sequencing errors in speech have been reported and analyzed by several authors (Cohen, 1967; Boomer & Laver, 1968; MacKay, 1970;

Fromkin, 1971), and most of these writers have concluded that the errors result from the interchange, omission, repetition or addition of features, phones, clusters of phones, syllables; words or even phrases. Such errors in speech production are detected simply by listening to live or recorded speech. It is not necessary to make precise measurements in time nor to use sophisticated equipment. With sequencing errors of this type it is sufficient to speak only of a "notional time" which has to do with the sequencing of segments (Tatham, 1970).

On the other hand, when one attempts to describe the timing and coordination of the submovements in an articulatory sequence, it is necessary to make observations in physical or clock time, and the intervals involved may be as brief as 10 ms (Kozhevnikov & Chistovich, 1965; Kent & Moll, 1975). Moreover, a problem arises as one attempts to reconcile the sequencing of articulatory movements with the seriation of more abstract units such as phonemes, syllables, or words. The difficulty is that the observed sequencing of articulatory gestures often is not at all clearly related to intuitive conceptions of phoneme, syllable, or word boundaries. This issue has been examined in many papers under the name of coarticulation. The current paper reviews these contributions and evaluates the adequacy of the solutions that have been proposed for the serial ordering problem at the articulatory level.

The concept of coarticulation

It has long been recognized that the articulatory movements of speech are not a transparency through which abstract segments such as phonemes are readily visible. In his famous analogy, Hockett (1955) imagined the process of speech production as being like a row of Easter eggs carried along a conveyer belt to a wringer that smashes them into a forbidding mess. The Easter eggs correspond to discrete phonemes, and the mess of broken eggs that comes out of the wringer corresponds to the speech output. This analogy is certainly vivid, but it has a negative tone, for it makes the process of speech production sound more like a pathological behavior than an orderly transformation of linguistic units into a sequence of intricately timed articulatory movements. But Hockett's analogy does capture some of the challenge that faces the investigator of speech movements. To meet this challenge, the investigator must deal with two worlds, one the world of abstract linguistic or intuitive units (notional time) and the other, a world of muscular contractions, structural movements, and acoustic signals (physical or clock time).

Attempts to impose rigid linguistic boundaries on articulatory events are quickly confounded by the fact that articulatory movements overlap one another in complex fashion. This difficulty gave rise to the concept of coarticulation, which in the barest sense means that the speech mechanism exhibits simultaneous adjustments for two or more of the presumed production units. To say that speech is characterized by coarticulation is to say that the peripheral events of speech behavior cannot be given a straightforward segmentation into intervals that correspond to linguistic units such as phonemes. Because of the complicated interaction of speech sounds, a given linguistic unit does not have invariant articulatory characteristics. Coarticulation is a conceptualization of speech behavior that implies (1) discrete and invariant units serving as input to the system of motor control, and (2) an eventual obscuration of the boundaries between units at the articulatory or acoustic levels. Models of speech production that assume discrete control units must reconcile the presumed discreteness of the input with the demonstrated existence of complex, overlapping, and somewhat variable articulatory gestures at the output.

The phenomenon of coarticulation applies to the acoustic as well as to the articulatory level. In writing about the former, Liberman, Cooper, Shankweiler & Studdert-Kennedy (1972, p. 14) commented as follows,

> . . . the acoustic cues for successive phonemes are intermixed in the sound stream to such an extent that definable segments of sound do not correspond to segments at the phoneme level. Moreover, the same phoneme is most commonly represented in different phonemic environments by sounds that are vastly different. There is, in short, a marked lack of correspondence between sound and perceived phoneme.

Liberman *et al.* continued their discussion by evaluating the amount of acoustic restructuring necessary for the realization of the "same" speech sounds in different phonetic contexts. They concluded with these remarks,

> . . . it is the usual case that the acoustic cues for a consonant are different when the consonant is paired with different vowels, when it is in different positions (initial, medial or final) with respect to the same vowels, and for all types of cues (manner or voicing, as well as place) . . . Vowels . . . articulated between consonants at rapid rates . . . also show substantial restructuring—that is, the acoustic signal at no point corresponds to the vowel alone, but rather shows, at any instant, the merged influences of the preceding or following consonant (Pp. 24–25).

Coarticulatory effects commonly are described as being of two types: forward (right-to-left or anticipatory) and backward (left-to-right or retentive). Forward coarticulation is said to occur when an articulatory adjustment for one phonetic segment is anticipated during an earlier segment in the phonetic string. A striking example of forward coarticulation of lip protrusion was provided by Benguerel & Cowan (1974), who reported that lip protrusion for a vowel in French may begin up to six segments in advance of its requisite appearance. The anticipatory lip gesture may be schematized as follows.

i s t r s t r y
←——

That is, in this phonetic sequence consisting of the unrounded vowel /i/, six consonants, and the rounded vowel /y/, the lips began to round with the first /s/.[1] In fact, Benguerel and Cowan sometimes observed the gesture of lip protrusion to begin during the unrounded vowel /i/. As an other example, Moll & Daniloff (1971) determined that anticipatory velar lowering for the nasal consonant /n/ in the phrase *free Ontario* may begin as early as with the articulatory movements toward the vowel /i/ in the word *free*. This coarticulatory pattern may be given the schematic representation below, where the symbol # designates a word boundary.

f r i # ɑ n t ɛ r ɪ o
←——

Backward coarticulation is said to occur when an articulatory adjustment for one segment appears to have been carried over to a later segment in the phonetic string. An example here would be the presence of lip protrusion during the /s/ segment of the word *boots*. As indicated in the schematic representation below, the lip protrusion associated with the vowel /u/ apparently has been retained in the following segments /t/ and /s/.

b u t s
——→

[1] Benguerel & Cowan (1974) note that the consonant /r/ is not rounded in French, so that anticipatory lip protrusion observed in the /strstr/ cluster is attributable only to the rounded vowel /y/.

Some overlapping of articulatory movements is inevitable, given that the speech organs are not capable of infinite acceleration. The vocal tract cannot change instantaneously from one configuration to another, and the articulatory transitions between sounds will therefore reveal interactive influences. However, what makes actual coarticulatory patterns at once puzzling and theoretically pivotal is that their explanation apparently goes beyond simply inertial effects—although inertia must have some role. (See Daniloff & Hammarberg (1973) for a discussion of this issue). One of the major attractions of research on coarticulation is the possibility that coarticulatory patterns may reflect the structure of the programming units in speech production. That is, if the speech apparatus makes simultaneous adjustments for sounds that are up to six or seven segments apart, obviously the system of motor control has information concerning several segments at once. Perhaps the range of segments over which simultaneous information is available may define the organizational schema of speech articulation. The work of Kozhevnikov and Chistovich, to be discussed next, is a good example of research in this direction.

The Kozhevnikov-Chistovich articulatory syllable

Kozhevnikov & Chistovich (1965) assumed that the range of forward coarticulation mirrors in direct fashion the size of the programming unit of articulation. From recordings of speech movements in Russian subjects, they observed that gestures of lip protrusion for consonant clusters preceding rounded vowels began simultaneously with the first consonant in the sequence. Thus, using the schematic conventions introduced above, the forward coarticulation of lip protrusion assumed the following patterns, where C is a consonant and V_r is a rounded vowel.

$$\underleftarrow{C \quad V_r}$$
$$\underleftarrow{C \quad C \quad V_r}$$
$$\underleftarrow{C \quad C \quad C \quad V_r}$$

Kozhevnikov & Chistovich concluded that articulatory movements are organized in syllables in the form CV, CCV, CCCV, etc., that is, a syllable comprising any number of consonants followed by a vowel. Benguerel & Cowan (1974) expressed this idea with the syllabic formula $(C)^n V$, where n represents the number of prevocalic consonants. This suggestion of speech organization is reminiscent of that offered by Stetson (1951), who argued for CV-type syllables as the fundamental units of speech production.

The basic hypothesis generated by the Kozhevnikov-Chistovich theory is that the motor programming of speech is discontinuous at certain intervals, namely, following the production of any given vowel. That is, anticipatory adjustments (forward coarticulations) are bounded by vowel segments, because consonants are programmed with the following vowel. After a vowel is encountered, a new programming unit begins. Although the data presented in Kozhevnikov & Chistovich (1965) supported this notion, more recent data on American English are not compatible with the idea of a CV-type syllable as the basic unit of speech articulation. As an example of this conflict, Moll & Daniloff (1971), as mentioned earlier, reported that anticipatory velar lowering in the phrase *free Ontario* began with the articulatory movements toward the vowel /i/ in *free*. This result suggests that the vowel /i/ was programmed in a larger unit which extended up to at least the nasal consonant /n/ in the following word. Other data in conflict with the Kozhevnikov-Chistovich theory have been reported by Kent, Carney & Severeid (1974). They observed that in a phrase such as *I intend*, velar lowering in preparation for the nasal consonant /n/ may be present during the diphthong /aɪ/, in which case forward coarticulation of the velopharyngeal property has crossed a vowel segment, a word boundary, and a diphthong segment. Hence, data indicate that as far as velopharyngeal adjustments are concerned, a vowel segment is programmed with a following consonant (making a VC programming unit) or even with another vowel and a following consonant (making a VVC programming unit). If the extent of forward coarticulation is taken as evidence of the size and form of the programming units of articulation, then syllables of the form CV, or even $(C)^nV$ (that is, any number of consonants preceding a vowel) are not the only candidates revealed by articulatory data.

Furthermore, interactions between vowel segments are not restricted to velopharyngeal function. Öhman's (1966) spectrographic study of V_1CV_2 utterances revealed that the acoustic characteristics of the first vowel V_1 were influenced by the transconsonantal vowel V_2. Thus, the articulatory configuration for V_1 was dependent to some degree on the articulatory requirements for V_2. Using the conventions adopted above, this coarticulatory pattern may be schematized as:

$$\underleftarrow{V_1CV_2}$$

This result was corroborated in a cinefluorographic study reported by Kent & Moll (1972). Their X-ray data revealed that the tongue and mandible positions for V_1 in V_1V_2 and V_1CV_2 sequences varied in a way that apparently was governed by V_2. For example, when followed by a stressed back vowel, the tongue position for vowel /i/ was less front in the oral cavity than when vowel /i/ was followed by a stressed front vowel. If the coarticulatory criterion used by Kozhevnikov & Chistovich is applied to the data of Ohman and of Kent & Moll, then it would have to be concluded that speech articulation can be programmed in units of the form VV and VCV. Of course, the point of this discussion is not to argue for the validity of such units, but to question if coarticulatory effects are restricted to any particular canonical form. The data indicate that they are not. However, one might inquire about the relative cohesiveness of the different forms, which is essentially the question addressed by MacNeilage & DeClerk (1969), who compared coarticulatory influences in CV and VC sequences. They interpreted their data on C_1VC_2 monosyllables to mean that the articulatory interaction is stronger between C_1 and V than between V and C_2. Whether such a conclusion can be generalized to more complicated phonetic sequences is uncertain.

Bladon & Al-Bamerni (1976) discovered in a spectrographic investigation of English /l/ that both forward and backward coarticulation effects were constrained by a CV-type articulatory syllable, like that proposed by Kozhevnikov & Chistovich. However, Bladon and Al-Bamerni proposed a more complete explanation of coarticulatory variation by means of an articulatory control principle called "coarticulation resistance." They wrote of this principle as follows.

> Antagonistic vocal tract adjustments apart, coarticulation is inhibited only by coarticulation resistance (CR) at some point in the succession of speech events. Each extrinsic allophone (and indeed each boundary condition) is assigned a value for CR by rules which may in some instances be language particular and in others quasi-universal. The CR value could be represented as a numerical coefficient attaching to a phonetic feature, say (3 CR), along the lines proposed by Chomsky & Halle (1968) for all other phonetic specifications in the phonetic system. (P. 149).

The valuation of CR for a speech segment appears to be an involved matter. For instance, Bladon & Al-Bamerni remark that, "... at least some of the rules for assignment of CR values must be context-sensitive" (p. 149). Furthermore, given that coarticulation may be sufficiently variable so as to be useful for the purpose of speaker identification (Su, Li & Fu, 1974), Bladon & Al-Bamerni allow that, "... there seem to be cases where CR is assigned on a highly idiolectal basis" (p. 150). Hence, the determination of CR is based on a combination of factors, including some that are universal, some that are language particular, some that are context sensitive, and some that are idiolectal. Perhaps the solution to coarticulation is as complex as this multiplicity of factors suggests, but even at that, the CR values seem to be little more than summary numbers that represent the contributions of many unknown, or poorly known, effects. The CR values do not in themselves generate predictions, but rather appear to be part of a numerical description of speech articulation.

Wickelgren's context-sensitive elementary motor responses

Wickelgren's (1969) theory of the motor control of speech is an interesting contrast to that of Kozhevnikov & Chistovich. Whereas the latter investigators proposed a $(C)^nV$ syllable as the basic programming unit of articulation, Wickelgren argued that speech units are coded allophonically as context-sensitive elementary motor responses. Wickelgren's notion of an allophone was that of "... a phoneme in a particular context of phonemes on either side" (p. 6). Thus, in a phonetic string XYZ, the element Y would be given the allophonic representation ${}_xY_z$, where the subscripted symbols designate the immediate bilateral context. Such context-sensitive allophones (or elementary motor responses) were assumed to be the basic unit of articulation, and Wickelgren (p. 11) explained their relevance to coarticulation in this way: "By assuming (context-sensitive) allophones to be the basic unit of articulation, rather than (context-free) phonemes, it is trivial to account for how the 'same phoneme' in different phonemic environments can be and must be different in some respects at all levels of the speech process, including the acoustic, vocal tract, and motor neuron levels". Later he adds that, "From the present point of view, coarticulation effects are not some strange complication to be disposed of by searching for a level of the speech process at which they do not exist, but are, instead, a basic feature of the speech code at all levels" (p. 6).

But Wickelgren's approach has a serious shortcoming in that only the immediate context of a segment is taken into account. The coarticulation research discussed earlier indicates that the interaction between segments is not limited to adjacent segments but may occur for segments that are as much as five to seven segments apart. One remedy to this problem is to specify additional context for each elementary motor response, but the required range is unknown and the ultimate number of coding units would seem to beg for Occam's razor. Wickelgren estimated that something on the order of 10^4 or 10^5 elementary motor responses would be required if only the immediate bilateral context were considered. Extending the context specification so as to include elementary motor responses such as ${}_{vw}X_{yz}$ (segment X produced in a VW__YZ context) would easily increase the number of basic articulatory units to over a million, perhaps even several millions. The brain might be capable of such a tremendous storage and retrieval task, but one hopes that more parsimonious and elegant explanations might be found for the process of speech production.

More detailed criticisms of Wickelgren's theory are available in other papers (MacNeilage, 1970; Whitaker, 1970; Halwes & Jenkins, 1971), and the many issues that have been raised will not be repeated here. For present purposes, it is sufficient to note that Wickelgren's theory attempts to explain coarticulation by an exhaustive listing of contextual possibilities, all of which presumably are learned and stored in the brain. Coarticulation thus becomes inextricably woven into the neural fabric of speech and exists in vast complexity even at the most central level of the speech production process. By handling the problem in this way, Wickelgren attributes all coarticulatory effects to central nervous control and ignores mechanoinertial factors completely. Wickelgren's monolithic explanation does not take advantage of the obvious physical limitations of the speech apparatus, and this failing greatly limits the appeal of his theory. Because mechanoinertial factors exist at the flesh and bone periphery, there is no need to make them part of the neural coding. Certainly, some backward coarticulation effects are inertial in origin, but Wickelgren's solution makes them indistinguishable from forward coarticulation effects that demand an explanation in terms of neuromotor strategy. Wickelgren further forsakes parsimony by proposing an extremely detailed allophonic representation. That is, he makes the motor control of speech dependent upon a large universe of articulatory units in which phonetic classes and similarities are unrecognized. For example, the generalization that vowels are nasalized when they precede nasal consonants is not conceived as a *modus operandi* for speech motor control; instead, Wickelgren's theory can only produce a tedious listing of all vowels in all nasal environments. With such an approach, all articulatory properties must be weighed equally to accomplish motor control, even though many properties would have slight effect on the motor response. Phonetic classes and hierarchies are denied operational significance because of the wholesale purchase of allophonic detail. At the central level of motor control, allophonic richness may only beget strategic poverty.

MacNeilage's target-based model

MacNeilage (1970) presented a conception of speech production that relies on a phoneme-sized input unit, an internalized space coordinate system for the specification of invariant articulatory targets, and a system of closed-loop control (probably the gamma motor system). In contrast to many writers on the subject of speech motor control, MacNeilage did not ask why phonemes fail to show articulatory invariance, but rather, "How do articulators always come as close to reaching the same position as they do?" (p. 184). In asking this question, MacNeilage proposed to solve the problem of motor equivalence in speech, which he defined as the ". . . achievement of relatively invariant motor goals from varying origins . . ." (p. 182).

MacNeilage argued that positional invariance is controlled by target specification. That is, he supposed that the control strategy generates a command which tells an articulator (or body part) to assume a certain position, but not to execute a certain movement nor to follow a particular pattern of muscle contraction. It is the task of the internalized space coordinate system to translate the phonological input into a series of target specifications. Then the motor system control mechanism generates movement command patterns based on the target information, and these patterns serve as instructions to the muscles. MacNeilage speculated on both open-loop and closed-loop control, suggesting that ". . . the open-loop component emits a *context-independent* command for an articulator to reach a certain position, and that closed-loop control circuits constantly sample the mechanical state of the articulator and adjust the command accordingly" (p. 190, emphasis in the original).

The notion of invariance of spatial targets has a rather long history in the study of speech. Hypotheses along this line have been proposed by Lindblom (1963), Stevens & House (1963), Halle & Stevens (1964), Ladefoged (1967), and Stevens (1972*b*). However, this approach is not without difficulties. Kozhevnikov & Chistovich (1965, pp. 81–82) cautioned against the assumption of phonemic invariance in spatial targeting as follows.

> In breaking down the flow of speech into sounds of speech, phoneticists attempt to find a compromise between two mutually exclusive requirements. One of them is to consider that a sound of speech is a segment of articulation in which there are all the characteristics required by the phoneme affiliation of the given sound . . . The second requirement is to divide the speech flow into segments which are immediately proximate to each other (one sound of speech should immediately follow another sound of speech). Since the various articulatory organs do not operate in strict unison and also have a certain amount of inertia, naturally these two requirements in principle are not compatible.

In MacNeilage's model, spatial targets are considered to be realized by variable forces applied to a plastic multi-channel system. MacNeilage's hypothesis of motor equivalence applied to speech production recognizes (1) the presence of multi-channel activity within the complex speech production apparatus, (2) potential articulatory constraints from both within and between articulatory channels, and (3) the dependence of such constraints upon the phonetic contexts of the utterance. MacNeilage speculates that the motor control system for speech involves a spatial coordinate system which senses the current locations of the individual articulatory structures involved in speech production and effects variable neuromotor commands to the muscles which act upon those structures in order to achieve internalized spatial targets.

Although MacNeilage's concept of motor equivalence in speech production has much to recommend it, there are certain modifications or additions that seem desirable. MacNeilage's concept emphasizes spatial targets that are related to phonemes in an invariant fashion. However, motor equivalence in speech probably involves more than spatial targets. That is, there are sounds which involve not only a spatial target, but rather dynamic or motion targets, such as a specified direction, velocity, or acceleration. It is necessary to accomplish a spatio-temporal targeting for many elements in speech, including the diphthongs, stops, and glides.

The ability of MacNeilage's model to explain coarticulation is difficult to gauge, because many important properties of the model are not defined. For example, it is essential to know the degree and nature of the target specifications. If the targets are concrete and positionally well-defined, then the model probably could not account for such coarticulatory effects as the anteroposterior variation in dorsal stop constriction (that is, the difference in the lingual closure for /k/ in the words *key* and *coo*). On the other hand, if the targets are abstract and only loosely specified as to position, the control system may allow too much variation and thereby jeopardize phonemic identity. Presumably, all necessary phonetic distinctions must be preserved in the target specifications, unless the motor control system also is constrained by phonetic principles (which would complicate the model immensely and destroy much of the appeal of the separate stages). To illustrate the complexity of target assignment, consider the fronto-palatal fricative /ʃ/, which becomes retroflex in the word *harsher*. The target specification either must automatically allow such a modification to occur by virtue of a tolerance of retroflexion or must be conditioned by *ad hoc* phonetic rules. Of course, the characteristic of retroflexion could be introduced as part of the phonological input, but then the model of speech production would become allophonically rather than phonemically based.

Forward coarticulation apparently occurs in MacNeilage's model because the mechanism generates movement command patterns for phonemes in advance of their issuance to the muscles. MacNeilage suggested that such advance preparation might look something like the following scheme for a seven-phoneme sequence A B C D E F G. For convenience, the status of each phoneme is considered within a five-part operational sequence, from the highest to the lowest level (central to peripheral).

(1) Phonological input to speech production model: phoneme G is being transferred from here to step 2.
(2) Target specification: phonemes E and F have been transformed to targets.
(3) Motor system control mechanism: phoneme C is formed and phoneme D is in the process of being formed.
(4) Motor system input: phoneme B (adapted to phonemes A and C) has been received.
(5) Realization as muscle contraction: phoneme A.

By way of a more concrete example, imagine that this pattern of operations is applied to a seven-phoneme sequence like that studied by Benguerel & Cowan (1974), specifically, /s t r s t r y/. Note that when the first /s/ in the sequence is being realized as muscle contractions, the rounded vowel /y/ is just being transferred from the phonological input to the target specification mechanism. But Benguerel and Cowan's data show that the first /s/ already is characterized by lip protrusion in anticipation of the vowel /y/. In other words, an articulatory target associated with vowel /y/ is being realized at the muscular level at the same time as the targets for /s/, which occurs six phonemes earlier in the string. Obviously, MacNeilage's imagined pattern of events cannot account for this result. This problem is illustrated not to dispute MacNeilage's example, but to indicate the implications that forward coarticulation has for a phoneme-based model. Models with a segmental (phoneme by phoneme) input often underestimate the range of coarticulatory effects that may occur in speech. To account for such effects, MacNeilage's model would have to be modified to allow target specifications to be simultaneously available for at least six to seven phonemes.

Chistovich and Kozhevnikov's goal-oriented model

Chistovich & Kozhevnikov (1969) introduced a model that is similar in some important respects to MacNeilage's model. They supposed that speech production is controlled by goals, with each goal being a "... description of motor acts abstracted from the concrete executive means ..." (p. 314). The speech control mechanism relies on the goals to select or calculate the particular combination of motor commands, given information on the momentary state of the articulatory system. That is, the commands actually issued by the motor executive depend both upon the instantaneous state of the system (presumably signaled by feedback information from the periphery) and the goals which apply to the speech input unit. The goals appear to be akin to the spatial target specifications described by MacNeilage, but Chistovich & Kozhevnikov did not elaborate the nature of the information provided by a goal. Because the model was not discussed in adequate detail, its predictions regarding coarticulatory effects are uncertain.

This model is representative of one major line of thinking in the development of speech production models. Models of this kind postulate that speech production is controlled by targets, goals, or features, that is, semi-independent parameters that are derived from the input units, whether they be phonemes, extrinsic allophones, or whatever. Such models have been introduced by Henke (1966), Mattingly & Liberman (1969), Liberman (1970), Tatham (1970) and Stevens (1972*a*). The most highly developed of these are the feature-based models (e.g. Henke, 1966), which recently have been tested in several studies of coarticulation. Given the refinement of these models and the availability of data to test them, feature-based models will be discussed in detail.

Feature-based models

Feature-based models assume a segmental input (phonemes, phonetic segments, or extrinsic allophones), with each segment being associated with a set of features. Typically, an input unit is conceptualized as a matrix or vector of component features, and the constituent features are thought to be more directly related to motor implementation than are the input units themselves.

Such a model was introduced by Henke (1966) in a computer simulation of speech articulation. Although Henke did not propose his model at the muscular level, subsequent work by others has applied essentially the same ideas to the muscular level of speech production. The input to Henke's model was a string of phonetic segments and the output was a description of the instantaneous states of the articulatory system. The model was founded on the idea that segments or portions of the articulatory apparatus continuously are seeking goals, which change abruptly in time. When taken together, the individual goals for portions of the vocal tract constitute a total or overall configurative goal of the speech mechanism. The subtargets for vocal tract segments are defined by the features associated with each input unit. Because the various features are not always specified for any given input unit, the model is equipped with a look-ahead operator that scans the forthcoming units to determine the next specified value of a feature. Hence, forward coarticulation comes about as the result of a forward scan based on a compatibility criterion for feature sets. The compatibility criterion allows a feature to be assumed (realized at the articulatory level) earlier than its parent segment so long as the articulation implied by that feature does not contradict the articulatory requirements of the segment currently being produced.

The application of feature specifications is illustrated in Table I, which shows how three hypothetical phonetic units X, Y and Z are classified with respect to four hypothetical features *a*, *b*, *c* and *d*. The features are assumed to be binary-valued, like distinctive features (Chomsky & Halle, 1968), but the binary base is by no means imperative (see Fant, 1971, for a discussion of this question). A given feature may be specified as "+" (positive) or "−" (negative) for each phonetic unit. As a more concrete example, a feature of VOICED may be either "+" (for a voiced segment) or "−" (for a voiceless segment). If a feature is unspecified for a given unit, then the value of "0" is indicated, meaning that neither a "+"

Table I Example of feature composition of phonemes: classification of three hypothetical phonetic units X, Y, and Z with respect to four hypothetical features *a*, *b*, *c* and *d*

Feature	Phonetic unit X	Y	Z
a	+	−	−
b	0	0	+
c	+	−	−
d	−	+	+

or "–" valence is required for that segment. The table shows that unit X is associated with valences of "+", "0", "+", and "–" for features *a*, *b*, *c* and *d*, respectively. Production of this phonetic unit therefore requires articulatory realization of features *a*, *c* and *d*. Because the value of feature *b* is unspecified the look-ahead operator may scan the upcoming units to determine the next specified value, which in this illustration occurs for unit Z, which carries a "+" valence for feature *b*. By the compatibility criterion, this "+" value of feature *b* for unit Z may be anticipated during the production of unit X, because no contradiction in articulatory requirements would occur.

Benguerel & Cowan (1974) used a feature-based model to explain their data on the forward coarticulation of lip protrusion in French. They proposed that in the case of the phonetic sequence /istrstry/ the feature specifications shown in line (2) below would apply at the systematic phonemic level. Line (3) gives the feature specifications at the articulatory goal level, at which point the action of the look-ahead operator is apparent.

(1) i	s	t	r	s	t	r	y
(2) –	0	0	0	0	0	0	+
(3) –	+	+	+	+	+	+	+

Benguerel & Cowan supposed that as soon as the unrounded vowel /i/ is articulated, the look-ahead operator scans forward until it reaches the next specified value of the lip protrusion feature, which occurs for the rounded vowel /y/. Because the articulation of each of the six nonlabial consonants is compatible with (noncontradictory to) lip protrusion, the gesture of lip protrusion may begin with the first consonant (/s/) in the sequence. Thus, the feature-based model predicts that the protrusion gesture will begin with the articulatory movements for the first /s/. Because their data generally agreed with this prediction, Benguerel & Cowan expressed support for a feature coding model of speech production.

However, there is reason to question the adequacy of Benguerel & Cowan's solution. The most important problem rests in their observation that "... more than 50% of the time the protrusion gesture starts on or even before the unrounded vowel V_u" (p. 51). The authors emphasized this aspect of the data because they saw it as evidence for the repudiation of the $C^{(n)}V$-type syllable proposed by Kozhevnikov & Chistovich (1965). Indeed, without this aspect of the data, the coarticulation effects could be explained as well by the Kozhevnikov & Chistovich model as by Benguerel & Cowan's feature-based model. Although the authors are correct in using this point to dispute Kozhevnikov & Chistovich's $C^{(n)}V$-type syllable, they overlook the fact that this same problem is sufficient grounds to reject their own feature-based model. If lip protrusion is observed to begin during the unrounded vowel, then a contradictory value (a value of +protrusion) has been assumed during a segment that presumably carries a specification of –protrusion. A forward scan or look-ahead would be impossible in this situation because the specified value of the lip protrusion feature for the /i/ segment would prevent such a process. Hence, the initiation of lip protrusion before completion of the unrounded vowel is at variance with the prediction of the feature-based model. If the model allows contradictory feature values to be anticipated, then indispensible phonetic contrasts would be lost and the features would be useless.

The possibility that lip protrusion may begin during a segment classified as unrounded merits further examination. Possibly, an error of segmentation was made in the collection of the data, such that protrusion was mistakenly assigned to the /i/ segment. But assuming that no errors were made, it might be argued that the feature of protrusion was redundant in this situation and that the factor of redundancy allowed the normally unrounded /i/ to assume the beginning of lip protrusion for the following /y/. This explanation might be satisfactory for English, which does not contain front rounded vowels, but it does not work for French, which has both rounded and unrounded front vowels.

Hence, at least this particular failing of a feature-based model cannot be explained by appeals to feature redundancy. But in a more general sense, such appeals are ill-considered unless the model is explicitly provided with a mechanism that takes redundancy into account. To do so probably would require an elaborate set of strategies, and to append such complexities on a model such as that proposed by Benguerel & Cowan might make it no more useful than any number of hierarchical solutions that could be imagined to solve the coarticulation problem. The basic attraction of the feature-based models is their parsimony, which results from the few and simple assumptions that they make about input units and recoding operations. To add considerations of feature redundancy would detract considerably from that attraction. Perhaps for that reason, proponents of feature-based models have not yet incorporated such a modification into their recoding schemes.

Another example of a feature-based model of coarticulation is that of Moll & Daniloff (1971), who were concerned with forward coarticulation of the velopharyngeal mechanism. They proposed the following velar feature specifications: "+", or open velum, for nasal consonants and utterance endings; "–", or closed velum, for nonnasal consonants; and "0", or unspecified, for vowels. Although this scheme of feature coding in conjunction with a look-ahead operation (like that described by Henke, 1966) appeared to fit Moll & Daniloff's data, several problems arise with this explanation of velar coarticulation. [These problems will be considered briefly here; for a more detailed discussion, see Kent, Carney & Severeid, (1974)].

The failing of Moll and Daniloff's model is apparent in comparing the model's predictions for the word *freon* (a word in their sample) and the word *trio* appearing at the end of a sentence. The segments in the word *freon* would be given feature specifications as follows.

f	r	i	a	n
–	–	0	0	+

Feature specifications for the word *trio* (with the number sign # used to indicate the utterance ending) would be these.

t	r	i	o	#
–	–	0	0	+

That is, because Daniloff & Moll assign to utterance endings the same value of the velar feature as they assign to nasal consonants, the two words *freon* and (utterance-final) *trio* would have the same sequence of velar feature values. In both cases, then, the look-ahead operator should allow anticipatory velar opening beginning with the first vowel /i/. Of course, the two vowels in the word *trio* are not nasalized in normal American speech, so this prediction of anticipatory velar opening cannot be correct. The problem cannot be solved by assigning a "–" value of the velar feature to utterance endings, because such a change would not explain the maintenance of a lowered velum in utterance-final vowels that follow nasal consonants (the very reason Moll & Daniloff made the "+" assignment in the first place). This example reveals the pitfalls that result from assigning feature values to entities such as utterance endings. (Note that the Moll-Daniloff model also predicts that all isolated vowels would be nasalized, because of the influence of the +nasal utterance ending.)[2]

As discussed by Kent, Carney & Severeid (1974), the feature-based model of Daniloff & Moll also fails to account for the fine details of articulatory timing. A particular example is the pattern of oral-velar articulatory timing for nasal-stop (NC) sequences. The data of Kent, Carney & Severeid (1974) as well as Moll & Daniloff's own data indicate that in these sequences, the beginning of the velar elevation required for C may occur simultaneously with the beginning of the oral closure movement for N. Thus, in the word *contract* (noun form), the velum begins to elevate at very nearly the same time that the tongue tip moves toward the alveolar ridge to make the oral closure for the /n/. A model that predicts coarticulation solely in terms of feature vectors for successive phonetic segments cannot explain this timing pattern. By the demands of its segmental input, such a model predicts that when the phonetic segment /n/ is received as the input unit, the lingual and velar goals for that unit must be satisfied concurrently. Hence, this model would not predict anticipation of the velopharyngeal target for /t/ in advance of the lingual closure needed to realize the /n/. A target of a closed velopharynx is contradictory to that for /n/, which must have dominance by virtue of its temporal hegemony in a feature-based model with segmental input. Even a look-ahead operator of the sort described by Henke (1966) would not command simultaneous gestures of velar elevation and lingual closure for /n/, because these subtargets are incompatible for this segment and therefore cannot be coincident.

Finally, a general caution to be exercised in the application of feature-based models has to do with the principle of compatibility, that is, the principle by which an articulatory feature value for one segment can be assumed earlier than the other feature values for that segment if the feature in question does not contradict the feature values for the preceding segment(s). This principle is used, either explicitly or implicitly, by Henke (1966), Moll & Daniloff (1971), and Benguerel & Cowan (1974). But data reported by Kent & Moll (1972) indicate that forward coarticulation may occur even for a physiological dimension for which the coarticulated segments assume contradictory values (for example, front *vs* back tongue position or close *vs* open jaw position). MacNeilage (1972) also has questioned the compatibility principle.

Proponents of feature-based models may attempt to answer these criticisms by postulating a "weakening" of features in certain contexts, by claiming that redundancy rules come into play, or by arguing that features are highly abstract and do not relate in a direct way to motor behavior in speech. The first two explanations may turn out to be useful, but formal statements of their operation are needed before they can be evaluated. No mention of such contingencies is offered by Henke (1966), Moll & Daniloff (1971), or by Benguerel & Cowan (1974). The third explanation, the appeal to the abstractness of the features, is difficult to accept, because if the advocates of feature-based models really believed this to be true, they would have no interest in proposing and testing models at the articulatory and muscular levels of the speech mechanism. If features are too abstract to be tested against articulatory data directly, then the feature-based models are incomplete and should be given the necessary mechanisms to make them testable.

[2]It remains an open question if values of phonetic features should be assigned at all to word boundaries, utterance terminations, and the like. Hence, we are not implying that either a "+" or "–" value of the velar feature necessarily has to be assigned to the end of an utterance. We only seek to demonstrate the problems that can arise once such an assignment has been made.

Hierarchical models

Models that might be described as hierarchical do not necessarily exclude all of the models discussed above, because some of them may be said to have an hierarchical character, albeit a rather simple one. However, the intent of the term here is to refer to models that incorporate many levels of speech organization, usually with complex interactions between and among the various levels. Hierarchical models may suppose that several units of different sizes are involved in the organization of speech articulation, but it is nonetheless pertinent to inquire about the smallest (and hence most basic) unit of the model. In fact, most of the hierarchical models posit units very much like those already discussed in conjunction with the previous models.

The model introduced by Liberman (1970) is one example of an hierarchical structure (see also Mattingly & Liberman, 1969). Liberman's production model is based on syllables, phones, features, and muscle gestures. As in the case of the feature-based models discussed above, Liberman assumed that each phone to be articulated consists of several features. Examples of features for the phone /b/ are "labial", "oral" and "voiced". Liberman (1970, p. 313) remarked that, ". . . each feature is represented by unitary nervous activity at a relatively high level of the production system". The features were supposed to have as their articulatory representation characteristic muscle gestures, and these gestures in turn were assumed to be largely independent and potentially coincident. Syllables serve as integrational units within which phones are converted to gestures, or in Liberman's words, ". . . the features that constitute the segments are organized in production into . . . syllabic bundles, each consisting of overlapped and largely independent articulatory components".

The subphonemic features are described further by Liberman *et al.* (1972), who propose that the features yield a phonemic invariance of muscle contraction insofar as these features are the result of "implicit instructions to separate and independent parts of the motor machinery" (p. 31). Given this view, ". . . the distinctive features of a phoneme are closely linked to specific muscles and the neural commands that activate them" (p. 31). The authors provide the following as examples of subphonemic features.

> The total gesture in the articulation of /b/ . . . can be broken down into several distinctive elements: (1) closing and opening of the upper vocal tract in such a way as to produce the manner feature characteristic of the stop consonants; (2) closing and opening of the vocal tract specifically at the lips thus producing the place feature termed bilabiality; (3) closing the velum to provide the feature of orality; and (4) starting vocal fold vibration simultaneously with the opening of the lips; appropriate to the feature of voicing. (P. 31).

From the foregoing description, it is easy to appreciate why Liberman and his colleagues expected to see invariance within the electromyographic (EMG) data collected from individual muscles involved in speech production. They assumed that a given muscle would be directly involved in the production of a given subphonemic feature. Ladefoged (1972) made a similar assumption in his discussion of prime and cover features. However, Ladefoged cautioned against the assumption that all the prime features will have simple articulatory or acoustic correlates. This admonition is warranted for at least two reasons: (1) many subphonemic features are produced through the concerted efforts of several muscle contractions, so that monitoring a single EMG channel might not demonstrate the expected invariance, and (2) the nature of the subphonemic features associated with particular phonemes varies substantially with the phonetic context, rate of utterance, stress, etc. Indeed, the relative importance of each muscle in accomplishing a desired movement pattern may be related to the instantaneous physiological strategy employed by the talker.

This principle of motor equivalence has been discussed at some length by MacNeilage (1970).

The model of Liberman and his colleagues may be distinguished from the feature-based models of the previous section by its requirement that features are organized within syllabic units or bundles. Henke's (1966) model, for example, made no suppositions regarding syllabic structure. In this respect, Liberman's model is based on a hierarchy involving interactions among syllabic units, phones, features, and muscle gestures.

Liberman's notion of syllabic bundles is by his own admission intuitive. He does not define a "syllable" but his remarks make it clear that the syllables must be discontinuous with one another, because the overlapping of articulatory components occurs within, but not between, syllables. Liberman's intuitive division of phones into syllables runs counter to Kozhevnikov & Chistovich's (1965) evidence that coarticulation is not limited to conventional syllabic units. They wrote on this issue that,

> It must be presumed that the syllable division accepted by linguists does not apply to the articulatory level. All combinations of CVC_1C_2V in an articulatory program represent two syllables where the first syllable ends with a vowel (CV/C_1C_2V) (P. 129).

Intuitive syllable divisions also do not correspond with data that show an interaction between adjacent or transconsonantal vowel segments (Öhman, 1966; Kent & Moll, 1972; Moll & Daniloff, 1971; Kent, Carney & Severeid, 1974). The usual notion of a syllable excludes any possibility of two vowels occurring within the same syllabic unit, so Liberman's hypothesis of articulatory overlapping within, but not between, syllables leaves much of coarticulation unexplained. Further evidence against conventional syllabic units has been reported by Kent (1972), who demonstrated that the particular pattern of tongue and mandible movement for the diphthong /aɪ/ depends upon the following phone, apparently without regard to conventional syllable boundaries.

Another hierarchical model was proposed by Tatham (1970), who assumed as input to the speech production model a string of extrinsic allophones. Tatham attributed a special motor cohesion to CV units, basing this decision largely on MacNeilage & DeClerk's (1969) physiological study of articulatory movements and Lehiste's (1972) acoustic investigation of temporal compensation. Extrinsic allophones are linked within CV units, so that a C_1VC_2 syllable may be formalized as

$$\#C_1V\text{-}C_2\#,$$

where # is a syllable boundary and uninterrupted symbols (C_1V) are linked by motor cohesion and temporal compensation. The assignment of a special or unique status to CV units is a notion of long standing in experimental phonetics (Stetson, 1951; Kozhevnikov & Chistovich, 1965; MacKay, 1974; Kent, 1976), linguistics (Greenberg, 1966), and speech pathology (Panagos, 1974). However, as already noted in this paper, motor cohesion or articulatory overlapping has been demonstrated for units other than CV. Perhaps CV units are relatively more cohesive than VC units, but articulatory data that might bear on this question are difficult to evaluate (see the comments of MacNeilage, 1972) and even if a difference exists, the cohesiveness of VC, VV, and VCV units demonstrated in previously cited reports cannot be ignored. The special status accorded to CV sequences notwithstanding, such units do not explain the occurrence of coarticulation.

If it is assumed that coarticulation occurs within syllable units but not between such units, then it is doubtful that any basic syllable form can explain the diversity of coarticulatory patterns in speech. But perhaps this very line of thought is misleading, because it is contrary to principles of fluent motor behavior. Rather than assume that syllables impose disruptive boundaries on the patterning of speech movements, it may be better to assume that syllables are adjusted to one another to facilitate fluent motor control. Syllables may be the basis of a flexible, adaptive motor control because they serve both a time-keeping function, and thus govern temporal compensation (Kozhevnikov & Chistovich, 1965; Lehiste, 1972), and a high-order seriation function, and thus are reflected in seriation errors (Boomer & Laver, 1968; Fromkin, 1971; MacKay, 1970).

Conclusions and summary

The major conclusion of this paper is that coarticulatory patterns are not explained adequately by any of the theories or models discussed herein. Although it is possible that small refinements here and there might render one of the models adequate, it is our suspicion that the refinements would have to be of larger scope. Of the various contenders, we prefer a model of the hierarchical variety. However, the current hierarchical solutions leave much to be desired. All too often, the kind of organization that is required rarely is described in more than a very sketchy way, by merely listing the variety of units that might by involved (such as phonemes, syllables, words and phrases). A second frequent shortcoming is that some hierarchical models are in a sense too powerful, allowing any number of predictions about the structure of motor control in speech. These models may be endowed with phonological, morphological, syntactic, and even semantic sensitivity, with no indication as to priorities or directions of operation. Hence, they are all but impossible to test.

A summary conception of the hierarchical organization of speech motor control is shown in Table II. The highest level of the hierarchy is a rhythmic grouping of syllables, or a combination of syllabic forms organized into a suprasegmental pattern, with specifications of speaking rate (hence overall duration), stress contrasts, and phrase level intonation. The number of syllables within a rhythmic grouping is variable, depending upon the utterance to be produced, but probably has an upper limit of about seven (Kozhevnikov & Chistovich, 1965; Fromkin, 1971; Martin, 1972). This level of the hierarchy is particularly important to establish the respiratory requirements of the utterance, e.g. the volume of air that must be inspired, the location and magnitude of pulsatile pressures for

Table II Basic components of an hierarchical model of speech production

RHYTHMIC GROUPING OF SYLLABLES (multi-syllable unit or phrase)
SYLLABLE FORM (individual syllable shapes and associated intrasyllabic rules)
PHONEME OR SEGMENT (basic phonetic constituents of syllable)
SET OF PHONETIC FEATURES (goals or targets defined for each phoneme or segment)
PATTERN OF ARTICULATORY TRANSITIONS (articulatory realizations of phonetic features)
SEQUENCE OF NEUROMOTOR COMMANDS (neural instructions to the muscles, derived from transitional requirements)

special stress effects, and the timing of pauses for inspiration. Some aspects of laryngeal function also would be dictated at this level, especially with regard to the intonation contour of the utterance. Beyond these issues, the rhythmic grouping governs the seriation and relative timing of the component syllables.

Syllabic forms are determined at the second level of the hierarchy. It appears that a variety of syllable structures can be selected, including CV, CCV, VC, VCC and CCVC, to mention a few. Perhaps the CV-syllable shape is an optimal form (Kozhevnikov & Chistovich, 1965; MacKay, 1974; Kent, 1976), but it is by no means the only shape. Indeed, flexibility in syllabic structuring may be a central factor in the fluent execution of motor control. The failure of adaptive syllable organization may be a primary reason for the aberrant nature of deaf speech, which may depart markedly from the temporal patterns of normal speech.

The third level of the hierarchy is that of phonemic organization, which is necessary to establish the phonetic requirements of the utterance. However, phonemes are not directly interpreted as neuromotor commands for speech articulation. Rather, phonemes are interpreted as features, which satisfy the requirements of phonetic distinctions and also provide the basis for the determination of the requisite articulatory transitions. Each phoneme is defined by a minimal set of features, which are relatively abstract descriptions of articulatory goals or states. A given feature may be given more than one motor interpretation, so that considerable flexibility exists in the realization of features at the articulatory level. For example, in the articulation of the diphthong /aɪ/, raising of the tongue may be accomplished by several different combinations of tongue and jaw movement, all of which are realizations of a feature that might be called "tongue raising".

The conversion of features to articulatory transitions occurs at the next level of the model, and it is at this level that features are given a concrete articulatory expression. The translation of features into articulatory transitions specifies not only the articulatory positions to be attained but also the relative timing among the various transitions in an articulatory sequence. Thus, a set of transitions, with a prescribed temporal pattern, contains the information finally required for the generation of neuromotor commands to the vocal tract. The basic temporal plan for motor control of articulation is determined by the transitions which have an immediate successional impact; other transitions are then adapted to this program.

That is, in order that the minimal requirements of phonetic seriation can be satisfied, some movements must follow one another in a prescribed sequence. On the other hand, some submovements do not have an immediate successional impact. The distinction between the two kinds of submovements can be illustrated with the word *snooze* (phonetically, /s n u z/). In this word, the lip protrusion required for the vowel /u/ does not have immediate successional impact because the segments /s/, /n/ and /z/ can be either rounded or unrounded. But velopharyngeal function does have an immediate successional impact for the segments /s/ and /n/. For the former, the velopharynx must be closed whereas for the latter, it must be open to ensure nasal resonance. The essential motor pattern depends upon the identification of critical submovements—those submovements that mark transitions between adjacent phonetic segments. The initial consonant cluster can be produced correctly only if the /s/ to /n/ transition is marked by a tongue tip adjustment (from partial constriction to complete closure), an opening of the velopharynx, and the onset of voicing. These articulatory changes have an immediate successional impact within the /sn/ cluster. There is some evidence that articulations without immediate successional impact, such as lip rounding in this example, are accommodated within the sequential pattern defined by the locally critical articulatory transitions. Kent, Carney & Severeid (1974) suggested that articulatory movements are "... programmed as coordinated structures, so that movements of the tongue, lips, velum and jaw often occur in highly synchronous patterns" (p. 487). The important point here is that articulatory motions without immediate successional impact are not initiated haphazardly, but rather in accord with a successional pattern that tends to (1) maximize the anticipation of articulatory goals and (2) maximize the number of simultaneous movements within an articulatory sequence.

Given this perspective, the strategy of motor control would assume that a string of phonetic segments is translated into a pattern of transitional instructions, some of which may be coincident. The basic elements of this pattern are the articulatory transitions that have immediate successional impact. Once these transitions have been specified, other transitions (those without such impact) are fitted into the pattern.

This work was supported by NIH grants NS-11022 and HD-32793.

References

Benguerel, A.-P. & Cowan, H. A. (1974). Coarticulation of upper lip protrusion in French. *Phonetica* **30**, 41–55.

Bladon, R. A. W. & Al-Bamerni, A. (1976). Coarticulation resistance in English /l/. *Journal of Phonetics* **4**, 135–50.

Boomer, D. S. & Laver, J. D. M. (1968). Slips of the tongue. *British Journal of Disorders of Communication* **3**, 2–12.

Chistovich, L. A. & Kozhevnikov, V. A. (1969). Some aspects of the physiological study of speech. In *Biocybernetics of the Central Nervous System* (Proctor, L. D., Ed.). Boston: Little, Brown.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Cohen, A. (1967), Errors of speech and their implications for understanding the strategy of language users. *Zeitung für Phonetik und allgemeine Sprachwissenschaft* **20**, 177–81.

Daniloff, R. G. & Hammarberg, R. E. (1973). On defining coarticulation. *Journal of Phonetics* **1**, 239–48.

Fant, G. (1971). Distinctive features and phonetic dimensions. In *Applications of Linguistics. Selected Papers of the Second International Congress of Applied Linguistics, Cambridge, 1969* (Perren, G. E. & Trim, J. L. N., Eds). London: Cambridge University Press.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language* **47**, 27–52.

Greenberg, J. H. (Ed.) (1966). *Universals of Language*. Cambridge, Mass.: M.I.T. Press.

Halle, M. & Stevens, K. N. (1964). Speech recognition: a model and a program for research. In *The Structure of Language* (Fodor, J. A. & Katz, J. J., Eds). Englewood Cliffs, N.J.: Prentice-Hall.

Halwes, T. & Jenkins, J. J. (1971). Problem of serial order in behavior is not resolved by context-sensitive associative memory models. *Psychological Review* **78**, 122–9.

Henke, W. L. (1966). Dynamic articulatory model of speech production using computer simulation. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, Mass.

Hockett, C. F. (1955). *A Manual of Phonology*. In *International Journal of American Linguistics* (*Memoir 11*). Baltimore, Maryland: Waverly Press.

Kent, R. D. (1972). Some considerations in the cinefluorographic analysis of tongue movements during speech. *Phonetica* **26**, 16–32.

Kent, R. D. (1976). Syllabic complexity and syntagmatic rules for syllable production—A comment on "Aspects of the syntax of behavior" by D. G. MacKay). *Quarterly Journal of Experimental Psychology* **28**, 483–9.

Kent, R. D., Carney, P. J. & Severeid, L. R. (1974). Velar movement and timing: evaluation of a model for binary control. *Journal of Speech and Hearing Research* **17**, 470–88.

Kent, R. D. & Moll, K. L. (1972). Tongue body articulation during vowel and diphthong gestures. *Folia Phoniatrica* **24**, 286–300.

Kent, R. D. & Moll, K. L. (1975). Articulatory timing in selected consonant sequences. *Brain and Language* **2**, 304–23.

Kozhevnikov, V. A. & Chistovich, L. A. (1965). *Rech: Artikulyatsiya i Vospriyatiye* (Moscow-Leningrad, 1965). Trans. *Speech: Articulation and Perception*. Washington, D.C.: Joint Publication Research Service, No. 30, 543.

Ladefoged, P. (1972). Phonological features and their phonetic correlates. *Journal of International Phonetics Association* **2**, 2–12.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America* **51**, 2018–24.

Liberman, A. M. (1970). The grammars of speech and language. *Cognitive Psychology* **1**, 301–23.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1972). Perception of the speech code. In *Human Communication: A Unified View* (David, E. E., Jr. & Denes, P. B., Eds). New York: McGraw-Hill.

Lindblom, B. E. F. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* **35**, 1773–81.

MacKay, D. G. (1970). Spoonerisms: the structure of errors in the serial order of speech. *Neuropsychologica*, **8**, 323–50.

MacKay, D. G. (1974). Aspects of the syntax of behavior: syllable structure and speech rate. *Quarterly Journal of Experimental Psychology* **26**, 642–57.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review* **77**, 182–96.

MacNeilage, P. F. (1972). Speech physiology. In *Speech and Cortical Functioning* (Gilbert, J. H., Ed.) New York: Academic Press.

MacNeilage, P. F. & Declerk, J. L. (1969). On the motor control of coarticulation in CVC monosyllables. *Journal of the Acoustical Society of America* **45**, 1217–33.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, **79**, 487–509,

Mattingly, I. G. & Liberman, A. M. (1969). The speech code and the psychology of language. In *Information Processing in the Nervous System* (Liebovic, K. N., Ed.). New York: Springer-Verlag.

Moll, K. L. & Daniloff, R. G. (1971). Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America* **50**, 678–84.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* **39**, 151–68.

Panagos, J. M. (1974). Persistence of the open syllable reinterpreted as a symptom of language disorder. *Journal of Speech and Hearing Disorders* **39**, 23–31.

Stetson, R. H. (1951). *Motor Phonetics*. Amsterdam: North Holland.

Stevens, K. N. (1972*a*). Segments, features, and analysis by synthesis. In *Langauge by Ear and by Eye* (Kavanagh, J. F. & Mattingly, I. G., Eds). Cambridge, Mass.: M.I.T. Press.

Stevens, K. N. (1972*b*). The quantal nature of speech: evidence from articulatory-acoustic data. In *Human Communication: A Unified View* (David, E. E. Jr. & Denes, P. B., Eds). New York: McGraw-Hill.

Stevens, K. N. & House, A. S. (1962). Perturbations of vowel articulation by consonant context: an acoustical study. *Journal of Speech and Hearing Research* **6**, 111–128.

Su, L. S. Li, K. P. & Fu, K. S. (1974). Identification of speakers by use of nasal coarticulation. *Journal of the Acoustical Society of America* **56**, 1876–82.

Tatham, M. A. A. (1970). A speech production model for synthesis-by-rule. *Working Papers in Linguistics*, No. 6, Computer and Information Sciences Research Center, Ohio State University.

Whitaker, H. A. (1970). Some constraints on speech production models. *Language Centre Occasional Paper*, No. 9, 1–13, University of Essex, England,

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review* **76**, 1–15.