In J. Kingston and M.E. Beckman (eds.),
Papers in Laboratory Phonology I: Between
the Grammar and the Physics of Speech.
Cambridge, U.K.: Cambridge University Press.

752

19

Tiers in articulatory phonology, with some implications for casual speech

CATHERINE P. BROWMAN AND LOUIS GOLDSTEIN

19.1 Introduction

We have recently begun a research program with the goal of providing explicit, formal representations of articulatory organization appropriate for use as phonological representations (Browman and Goldstein 1986; Goldstein and Browman 1986). The basic assumption underlying this research program is that much phonological organization arises from constraints imposed by physical systems. This is of course a common assumption with respect to the *elements* – features – used in phonological description; it is not such a common assumption, at least in recent years, with respect to the *organization* of phonological structures.

In our view, phonological structure is an interaction of acoustic, articulatory, and other (e.g. psychological and/or purely linguistic) organizations. We are focusing on articulatory organization because we believe that the inherently multidimensional nature of articulation can explain a number of phonological phenomena, particularly those that involve overlapping articulatory gestures. Thus, we represent linguistic structures in terms of coordinated articulatory movements, called *gestures*, that are themselves organized into a *gestural score* that resembles an autosegmental representation.

In order to provide an explicit and testable formulation of these structures, we are developing a computational model in conjunction with our colleagues Elliot Saltzman and Phil Rubin at Haskins Laboratories (Browman, Goldstein, Kelso, Rubin and Saltzman 1984; Browman, Goldstein, Saltzman, and Smith 1986). Figure 19.1 displays a schematic outline of this model, which generates speech from symbolic input.

As can be seen from the number of submodels in the figure, gestures are relatively abstract. Even articulatory trajectories are one step more abstract than the output speech signal – they serve as input to the vocal tract model (Rubin, Baer, and Mermelstein 1981), which generates an acoustic signal. In addition, the actual articulatory trajectories associated with a gesture are generated from a dynamical description, which introduces another layer of abstraction. The

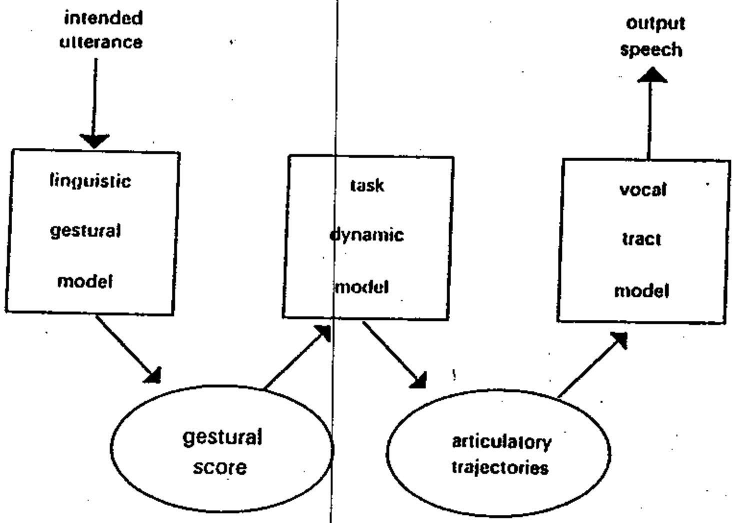


Figure 19.1 Computation modeling of gestures using articulatory dynamics

particular dynamic model we are using, the task dynamics of Saltzman and Kelso (1987), requires that gestures be discrete, a further abstraction. That is, we assume that continuous movement trajectories can be analyzed into a set of discrete, concurrently active underlying gestures. And finally, the discrete, abstract, dynamically defined gestures are further organized into gestural scores in the linguistic gestural model.

It is the qualities of discreteness and abstractness, when combined with the inherently spatiotemporal nature of the dynamically defined gestures, that give this system its power. As abstract, discrete, dynamic linguistic units, the gestures are invariant across different contents. Yet, because the gestures are also inherently spatio-temporal, it is possible for them to overlap in time. Such overlapping activation of several invariant gestures results in context-varying articulatory trajectories, when the gestures involve the same articulators, and in varying acoustic effects even when different articulators are involved. That is, much coarticulation and allophonic variation occurs as an automatic consequence of overlapping invariant underlying gestures (see Fowler 1980; Liberman, Cooper, Shankweiler, and Studdert-Kennedy 1967). And these qualities of the system also provide a relatively straightforward and invariant description of certain casual speech phenomena, as we shall see in section 19.3.

While the system is powerful, it is also highly constrained: there is a fair amount of structure inherent in the gestural framework. One important source of structure resides in the anatomy of the vocal tract, which provides a highly constrained, three-dimensional articulatory geometry; this structure will be outlined in section 1.1. A second important source of structure resides in the dynamical description, outlined in section 1.2. Both of these types of structure come together in the task dynamics model (Saltzman 1986; Saltzman and Kelso 1987), in which the basic assumptions are (1) that one primary task in speaking is to control the coordinated movement of sets of articulators (rather than the individual movements of individual articulators), and (2) that these coordinated movements can be characterized using dynamical equations.

19.1.1 *Articulatory organization*

Vocal tract variables. In order for the task dynamics model to control the movement of a set of articulators, those articulators needed to accomplish the desired speech task or goal must first be specified. For example, a lip closure gesture involves the jaw, the lower lip, and the upper lip. These articulators are harnessed in a functionally specific manner to accomplish the labial closure task. It is the movement characteristics of the task variables (called vocal tract variables) that are controlled in task dynamics. Thus, the lip closing gesture refers to a single goal for the tract variable of lip aperture, rather than to a set of individual goals for the jaw, lower lip, and upper lip. The current set of tract variables and their associated articulators can be seen in figure 19.2.

Gestures. In working with the tract variables, we group them into gestures. The oral tract variables are grouped in terms of horizontal–vertical pairs, where both members of a pair refer to the same set of articulators: LP-LA, TTCL-TTCD, TBCL-TBCD. ("Horizontal" and "vertical" refer to these dimensions in a straightened vocal tract, i.e. a tube model; thus, constriction degree is always considered to be orthogonal to, and hence "vertical" with respect to, the "horizontal" dimension of the stationary upper or back wall of the oral tract.) The oral gestures involving the lip, tongue tip, and tongue body thus consist of paired tract variables, where each tract variable associated with a gesture is modeled using a separate dynamical equation. That is, for oral gestures, two dynamical equations are used, one for constriction location and one for constriction degree. Since the glottal and velic aperture tract variables do not occur in pairs, they map directly onto glottal and velic gestures, respectively. We refer to the gestures symbolically, using the symbols displayed in table 19.1 for the gestures described in this paper.

Gestural scores. In order to apply task dynamics to speech in a linguistically interesting way, we must be able to model the articulatory structure of an entire utterance in terms of a set of gestures. This larger structure we refer to as a gestural score. Figure 19.3a shows a symbolic representation of a hypothetical gestural score (for the word "palm," pronounced [pʰam]). The articulatory trajectories associated with the gestures can be visualized with the aid of figure 19.3b, which shows the trajectories of four tract variables: velic aperture, tongue

tract variable		articulators involved
LP	lip protrusion	upper & lower lips, jaw
LA	lip aperture	upper & lower lips, jaw
TTCL	tongue tip constrict location	tongue tip, body, jaw
TTCD	tongue tip constrict degree	tongue tip, body, jaw
TBCL	tongue body constrict location	tongue body, jaw
TBCD	tongue body constrict degree	tongue body, jaw
VEL	velic aperture	velum
GLO	glottal aperture	glottis

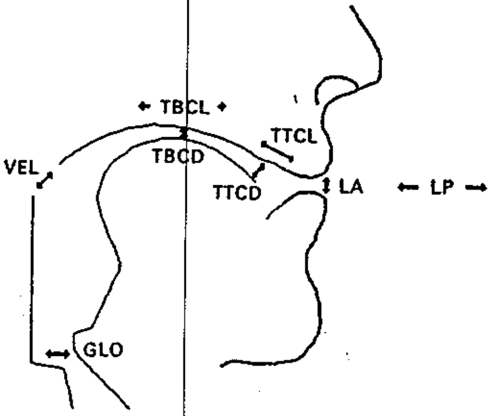


Figure 19.2 Tract variables

Table 19.1 *Gestural symbols*

symbol	referent	tract variable
i	palatal gesture (narrow)	TBCD, TBCL
a	pharyngeal gesture (narrow)	TBCD, TBCL
β	bilabial closing gesture	LA, LP
τ	alveolar closing gesture	TTCD, TTCL
σ	alveolar near-closing gesture (permits frication)	TTCD, TTCL
λ	alveolar lateral closing gesture	TTCD, TTCL
κ	velar closing gesture	TBCD, TBCL

body constriction degree, lip aperture, and glottal aperture. Each curve shows the changing size of a constriction over time, with a larger opening being represented by a higher value, and a smaller opening (or zero opening, such as for closure) being represented by lower values. Note that in some cases (e.g. velic aperture, lip aperture) this orientation produces a picture that is inverted with respect to the

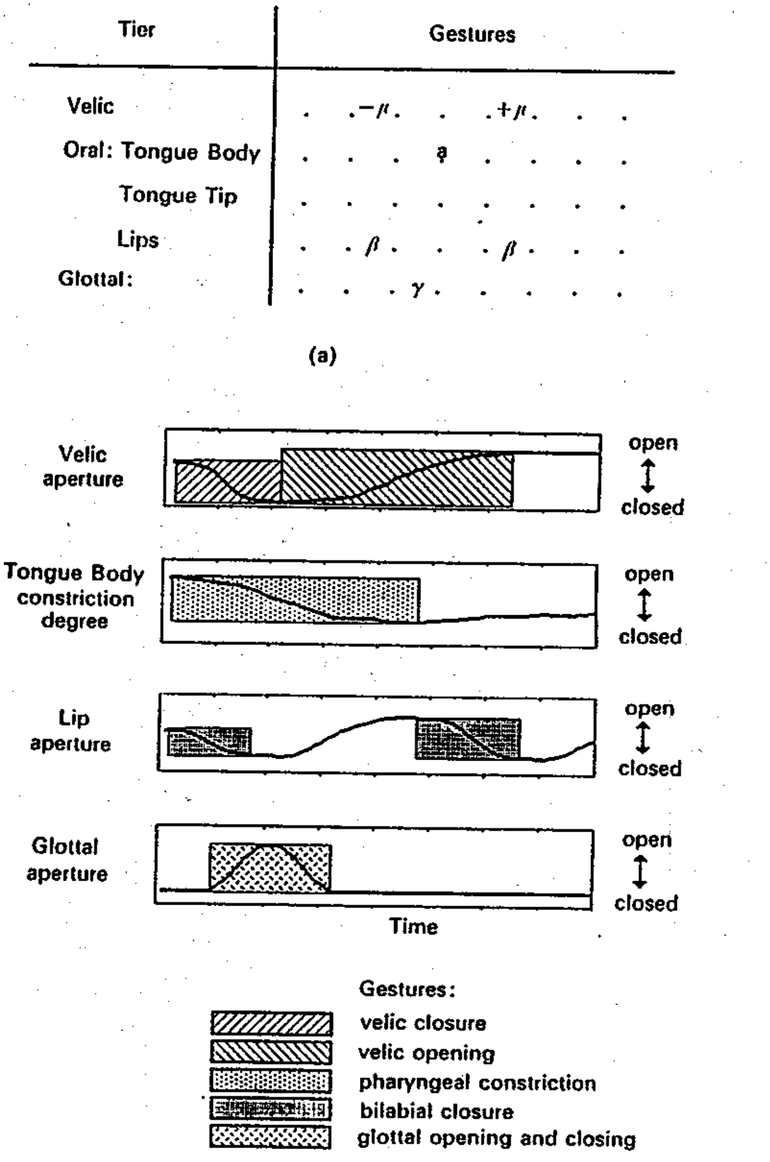


Figure 19.3 Hypothetical gestural representation for "palm." (a) Symbolic gestural score, (b) hypothetical trajectories (closure is indicated by lowering).

vertical movement of the major articulator (velum, lower lip) involved in changing the constriction size. For tongue body constriction degree, the constriction is in the pharynx, so that the relevant articulatory movement (of the rear and root of the tongue) is primarily horizontal. The trajectories of the tongue and lip tract variables are approximated from measured articulatory data (for the utterance "pop," rather than "palm"), while the trajectories for the velic and glottal variables are hand-drawn estimates, included for illustrative purposes. The shaded areas show the portions of these trajectories that would be generated by the dynamical systems representing the constriction degree tract variables for each of the six labeled gestures.

Articulatory tiers. The gestures in the gestural score are organized into articulatory tiers, where the tiers are defined using the notion of articulatory independence. Velic gestures are obviously the most independent, since they share no articulators with other gestures. In this limiting case, velic gestures constitute a wholly separate nasal (velic) subsystem and hence are represented on a separate velic tier. Glottal gestures also participate in an independent subsystem (although other laryngeal gestures, for example for larynx height, would also participate in this subsystem), and hence are represented on a separate glottal tier. The oral gestures form a third subsystem, with the jaw as a common articulator. Since the oral gestures are distinguished by different combinations of articulators, oral gestures are represented on three distinct oral tiers, one for the lips, one for the tongue body, and one for the tongue tip. Each of these is associated with a distinct pair of tract variables (see above). Note that within the oral subsystem, the tongue body and tongue tip tiers effectively form a further subclass, since they share two articulators, the jaw and the tongue body proper.

Note that these articulatory tiers correspond closely to organizations posited by both phoneticians and autosegmental phonologists. The three oral tiers of lips, tongue tip, and tongue body correspond to the traditional groupings of places of articulation into three major sets: labial, lingual, and dorsal (Vennemann and Ladefoged 1973; Halle 1982; Ladefoged and Maddieson 1986). And autosegmental phonologists have often proposed tiers that correspond to independent articulatory systems, e.g. the larynx (for tone and voicing), the velum (for nasality), and the oral articulators (Clements 1980, 1985; Goldsmith 1976; Thrainsson 1978).

19.1.2 *Dynamical description*

In addition to specifying the articulatory structure of an utterance by selecting the appropriate gestures, the gestural score also specifies the values of the dynamic parameters for use in the task-dynamic model. The task dynamic model uses these values as the coefficients of damped mass-spring equations (see appendix A), thereby generating characteristic movement patterns for the tract variables as well as coordinating the component articulators to achieve these movement patterns.

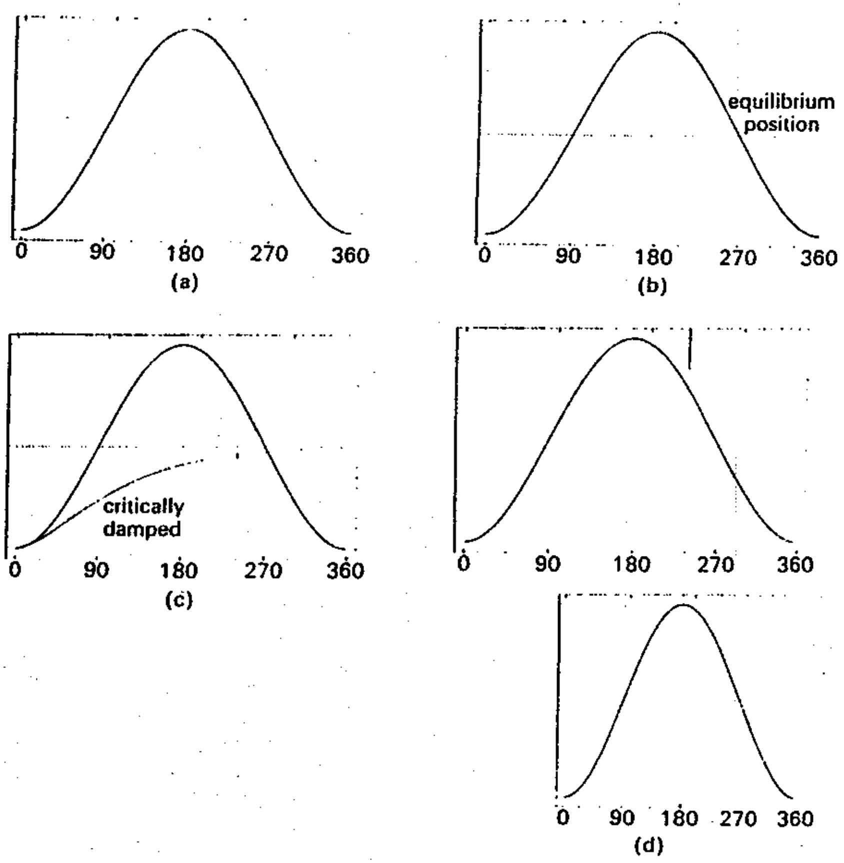


Figure 19.4 Abstract underlying gesture. (a) One cycle, (b) equilibrium position, (c) critical damping, (d) phasing between two gestures.

Since the current version of the model assumes unit mass and critical damping, the stiffness k and the equilibrium position x_0 are the parameters in the equation that can vary in order to convey linguistic information such as phonetic identity or stress (see appendix B). Figure 19.4 shows how these parameters are used to characterize an abstract underlying gesture.

We begin by assuming that a gesture consists of an abstract underlying 360 degree cycle, represented in figure 19.4a by the hump (in impressionistic terms), which is also a single cycle of an undamped cosine (in more quantitative terms). Figure 19.4b shows the equilibrium position for an arbitrary tract variable associated with this abstract gesture. For the undamped cosine in figure 19.4b, the trajectory generated by the abstract gesture oscillates around the equilibrium position, which is midway between the peak and valleys. The amount of time it takes for the gesture to complete this cycle is a reflection of its stiffness (given that we assume unit mass). The stiffer the gesture, the higher its frequency of oscillation and therefore the less time it takes for one cycle. Note this also means that, for a given equilibrium position, the stiffer the gesture, the faster the movement of the associated articulators will be.

However, the trajectory actually generated by our system is qualitatively different from the undamped "hump" seen in figure 19.4b since we assume critical damping rather than zero damping. As can be seen in figure 19.4c, the trajectory generated by a critically damped gesture approaches the equilibrium position increasingly slowly, rather than oscillating around it. In fact, the equilibrium position in a critically damped system approaches the peak displacement, or "target" (in effect, the target is the asymptote). Because it takes an infinite amount of time to actually reach the target in a critically damped system, we have specified that the effective achievement of the target is at 240 degrees with respect to the abstract underlying 360 degree cycle. This means that effectively only half the underlying abstract cycle is generated by a single gesture: the "to" portion of the underlying cycle. (This partial generation is exemplified in figure 19.3b; we are currently experimenting with generating the "fro" portion as well.)

Task dynamics serves to coordinate the articulators within a particular gesture; it also coordinates the effects, on a single articulator, of several different concurrent gestures. It does not, however, yet provide a mechanism for coordinating the gestures themselves. That must be explicitly specified in the gestural score. The abstract specification we adopt involves a relative phase description (Kelso and Tuller 1985). In such a description, gestures are synchronized with respect to one another's dynamic states, rather than timed by an external clock. In the current version of our system, gestures are phased with respect to each other's abstract underlying 360 degree cycles, as can be seen in figure 19.4d. In the figure, the two gestures are phased such that the point corresponding to 240 degrees for the top gesture is synchronized with the point corresponding to 180 degrees of the bottom gesture.

Given this gesture-to-gesture approach to phasing, a complete characterization of the gestural score for a particular utterance must specify which gestures are phased with respect to each other. In the next section, we explore this question further.

19.2 Gestural scores and tier redisplays

In the sections that follow, we will exemplify our explorations of the organization of gestures into gestural scores using X-ray data. These data come from the AT&T Bell Laboratories archive of X-ray microbeam data (Fujimura, Kiritani, and

Ishida 1973; Miller and Fujimura 1982), partially collected by researchers at Haskins Laboratories. In the data we examined, the X-ray microbeam system tracked the positions of (up to seven) small lead pellets placed on the lower lip, the jaw, the tongue blade, the tongue dorsum mid, the tongue dorsum rear, and/or the soft palate, in addition to two reference locations. We have examined a sample of utterances from three speakers of Standard American English, using the horizontal and vertical displacements of the pellets over time as the source data for deriving the dynamic parameter values and phasing for our gestural scores.

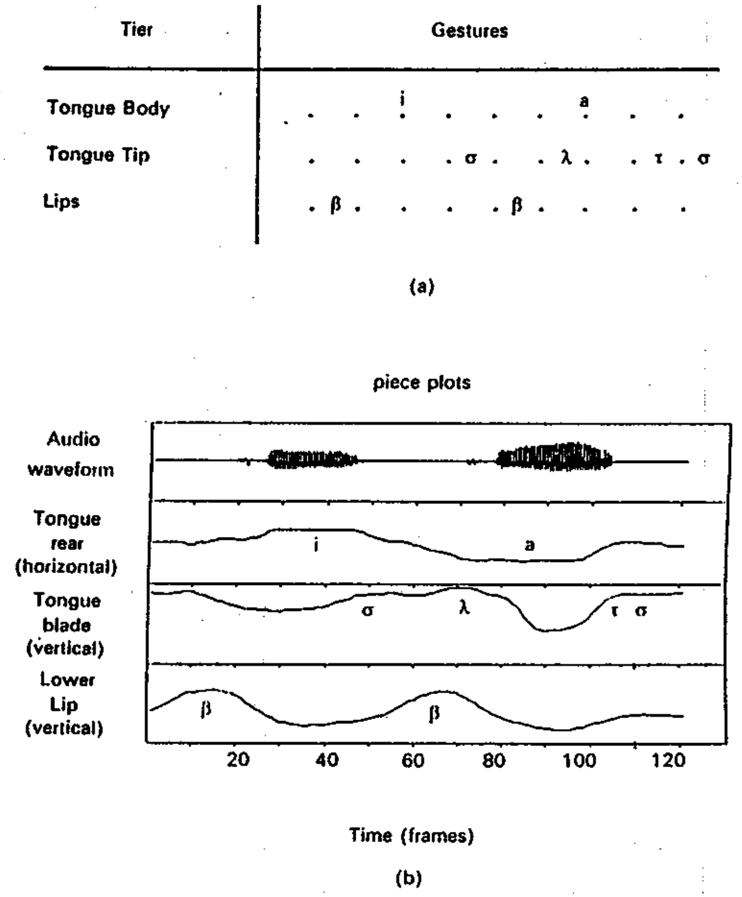


Figure 19.5 "piece plots" ([pis#ˈplats]). (a) Symbolic gestural score (oral tiers only). (b) X-ray pellet trajectories (closure is indicated by raising).

We begin, in this section, by exploring gestural scores for canonical forms; in section 19.3, we will look at how these canonical forms are modified in casual speech. In this section, then, we focus on data in which the syllable structure differs among the utterances. So far, we have examined paired utterances of the form [...iCa...], [...iCCa...], and [...iCCCa...], where the second syllable is always stressed, and the pairs are distinguished by a word boundary occurring before the first consonant in one member of the pair, and after the first consonant in the other member of the pair. The single consonants were [s], [p], and [l]; the triplet was [spl]; the doublets were [sp], [sl], and [pl]. Thus, the paired utterances differ in terms of how the initial consonant is syllabified, e.g. [...i#'spa...] vs. [...is#'pa...].

Figure 19.5a displays a symbolic gestural score (oral tiers only) for one of these utterances, [pis#'plats]; we will be working with this score and variants thereof throughout this section. Figure 19.5b shows the articulatory trajectories from the X-ray data that this score is intended to represent, with the gestural symbols added at approximately the target. For our preliminary investigations, we have assumed that the horizontal displacement of the rear of the tongue corresponds most closely to the trajectory generated by the vocalic gestures we are using ({i} and {a}), while the lip and blade vertical displacements correspond to the trajectories generated by the consonantal gestures we are considering ({β}, {σ}, and {λ}). (We are using the curly braces { and } to denote gestures.) The measured pellet trajectories can only be an approximation to the trajectories that would be generated by individual gestures, however, partly because of uncertainty as to pellet placement (especially on the tongue), but mostly because of the overlap of gestures, particularly when they are on the same articulatory tier. In the case of gestural overlap, several gestures will contribute to the observed articulatory trajectories. That is, gestures are discrete, abstract entities that combine to describe/generate the observed continuous articulatory trajectories.

19.2.1 *Rhythmic and functional tiers*

Gestural scores are subject to the same display problem as any multitiered approach, namely, associations among tier nodes can only be easily specified for immediately contiguous tiers in a display. Moreover, the gestural score as described so far only contains articulatory information. Stress information, for example, is not included. Therefore, in this subsection we introduce two additional types of display, one that incorporates stress information, and one that facilitates the display of associations and phasing among the gestures. (Our use of "display" is based on that of Clements and Keyser 1983.)

Before moving further into discussions of formalism, however, a word of caution is in order. We do not intend our use of formalism and symbols to be introducing new or "deeper" realities into a gestural description. That is, for us symbols do not generate the gestures, rather, symbols are pointers to gestures, for descriptive convenience. Similarly, the various displays serve merely to emphasize one or another aspect of the gestural organization; displays are always projections that serve to decrease the dimensionality of the underlying reality, again for descriptive convenience.

Rhythmic tier. To incorporate stress information, we use an additional rhythmic tier. (We assume pitch accent is a separate but related phenomenon, following Selkirk [1984]; we do not attempt to account for pitch accent at present.) Nodes on the rhythmic tier consist of stress levels; each syllable-sized constellation of gestures will be associated with a stress node. Each stress node affects the stiffness and constriction degree of the gestures associated with it (see appendix B). Note that we do not call this a syllable tier. Our current hypothesis is that the rhythmic component is a separate and independent component, whereas syllabicity is a complex function of gestural organization. Since we are only barely beginning to get a handle on gestural organization, we prefer to be conservative in our postulated structures. That is, we continue to err on the side of under-structured representations. (In addition, it is possible that the rhythmic component may be associated with its own set of articulators. We are intrigued, for example, by the notion that the jaw may be heavily implicated in the rhythmic component, as suggested by Macchi [1985] among others.) The first redisplay, then, seen in figure 19.6a, simply adds the rhythmic tier to the gestural score. All the gestures occurring under the curly bracket are assumed to be affected by the stress value of the relevant node on the rhythmic tier. That is, the curly bracket is a shorthand notation indicating that the value on the rhythmic tier is associated with every gesture on the articulatory tiers displayed beneath it.

Oral projection tier. An alternative display of the associations between the rhythmic tier and oral tiers is seen in figure 19.6b, where the dimensionality is reduced by projecting the gestures from the lip, tongue body, and tongue tip onto a single oral tier. The sequence on the oral tier indicates the sequence of achievement of the constriction degree targets for the gestures. That is, since the gestures are inherently overlapping, a single point must be chosen to represent them in their projection onto a single tier. The achievement of their target values represents the sequencing that occurs in canonical form; it also captures the sequencing information that most directly maps onto acoustically defined phonetic segments.

Functional tiers. Another type of display serves to reorganize the gestures, using functional tiers. At the current stage of our explorations, we are positing two functional tiers, a vocalic one and a consonantal one. This distinction is functionally similar to that made in CV phonology (Clements and Keyser 1983; McCarthy 1981), especially in its formulation by Keating (1985). Like Keating, we are struck by the usefulness of separate C and V tiers for describing certain aspects of articulation, although we differ in our definition of the nodes on the tiers, which for us are dynamically-defined articulatory gestures. What C and V tiers can

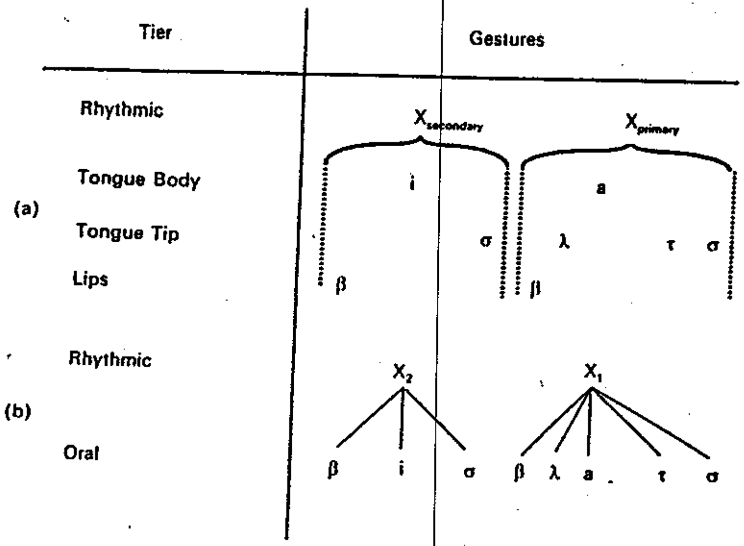


Figure 19.6 Tier displays for "piece plots" ([pis#' plats]). (a) Symbolic gestural score (oral tiers only) with rhythmic tier added. (b) Oral projection with rhythmic tier.

crucially capture is the fact of articulatory overlap between the vowels and consonants.

The X-ray data we have analyzed thus far (see, for example, figure 19.5b) have consistently supported the contention (Öhman 1966; Fowler 1983) that consonant articulations are superimposed on continuous vowel articulations, which themselves minimally overlap. As can be seen in figure 19.7a, this description is directly captured using functional C and V tiers. The association lines between the two tiers indicate that the associated consonantal gestures all co-occur with the vowel gesture. The adjacency of the gestures on the V tier indicates that the vowel articulations are effectively continuous, with the possibility as well of minimal overlap. We will discuss the C tier below. Here we simply note that the last consonantal gesture is not associated with a vocalic gesture. While we have not yet investigated syllable-final clusters in any detail, a cursory inspection suggests that the vocalic gesture in fact does not co-occur with this final consonantal gesture. It is intriguing to speculate how this might relate to extra-metrical consonants (cf. Hayes 1981) and/or phonetic affixes (Fujimura and Lovins 1978).

C and V tiers have several advantages, beyond the clear advantage of direct representation of articulatory overlap. As has been noted elsewhere (e.g. Fowler 1981; Lindblom 1963; Öhman 1967), the ability to deal with overlapping articulatory specifications makes it possible to unify the account of temporal and segmental variability. For example, increasing the amount of overlap between the articulatory movements associated with a vowel and a following consonant will simultaneously shorten the acoustic signal associated with the vowel and cause increasing amounts of coarticulation to be observed in the acoustic spectrogram and in the movements of individual articulators. In addition, positing separate C and V tiers as universal articulatory organizations leads to a new perspective on the role of both nonconcatenative (CV-based) morphology and vowel harmony. In the former case, McCarthy's (1981) analysis of Semitic morphology using C and V tiers can be seen simply as a morphologization of an already existing, universal articulatory organization. (A similar point was made by Fowler 1983.) In the latter case, vowel harmony simply becomes a natural extension of the already existing V tier.

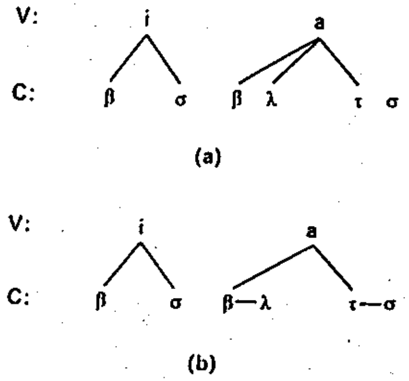


Figure 19.7 Consonant and vowel tier displays for "piece plots" ([pis#'plats]). (a) Associations (overlap). (b) phasing.

Our C and V tier display differs from related displays in other phonologies in the interpretation of sequencing, which acts like a combination of tier and (linear feature) matrix displays. That is, like autosegmental tier displays, associations among nodes on different tiers are indicated by association lines rather than by sequencing, thereby permitting many-to-one associations. Like both autosegmental and matrix displays, the sequence on each tier is meaningful. Unlike autosegmental displays, however, and like matrix displays, sequencing between tiers is also meaningful. In this sense, the C and V tier display is a two-dimensional plane, with sequencing proceeding along the horizontal dimension, and function type along the vertical dimension. The horizontal sequencing must capture exactly the same sequence of gestures as that displayed by the oral tier discussed above; that is, a constraint on all displays is that they must portray the canonical sequencing relations when projected onto a single tier. The sequencing between gestures on the V tier and the C tier is visually conveyed in figure 19.7 by the angle of the lines: a line slanting right (i.e. with its top to the right of its bottom) indicates that the consonant(s) precede the associated vowel, while a line slanting left indicates that the consonant(s) follow the associated vowel.

Contiguity operates differently on the C and V tiers. This captures a functional difference between vowels and consonants, where vowels act as a kind of background to the "figure" of the consonants. On the one hand, gestural contiguity on the V tier is completely independent of the C tier. This reflects the fact, to be discussed in detail in the next section, that vowel articulations are contiguous (or partially overlapping), regardless of the number of consonants intervening. On the other hand, contiguity on the C tier is sensitive to intervening vocalic gestures, in the sense that consonantal gestures overlap considerably less, if at all, when a vocalic gesture intervenes. A related effect of this functional difference has to do with the change in stiffness as more consonants are inserted between vowels: the vowel gestures decrease in stiffness while the consonantal gestures increase their stiffness.

19.2.2 *Using functional tiers with phasing*

In this subsection, we explore some details concerning the specification of phase relations among gestures on the C and V tiers. That is, given that gestures are spatio-temporal in nature, we need to be able to specify how they are coordinated – we cannot simply assume they are coordinated onset-to-onset, or onset-to-target. Figure 19.7b shows a schematic representation of the phase associations for our sample symbolic gestural score. Here, the only association lines from figure 19.7a that remain are those that phase gestures relative to each other. Statement (1) summarizes the phasing associations that hold between the V and C tiers:

(1) A vocalic gesture and the leftmost consonantal gesture of an associated consonant sequence are phased with respect to each other. An *associated consonant sequence* is defined as a sequence of gestures on the C tier, all of which are associated with the same vocalic gesture, and all of which are contiguous when projected onto the one-dimensional oral tier.

Notice that this phasing association statement follows the unmarked pattern of associations for tiers in autosegmental phonologies (left precedence).

Statement (2a) specifies the numerical values for phase relations between a vowel and the following leftmost consonant. An example of this phasing can be seen in the X-ray pellet trajectory data depicted in figure 19.8 for [pip#'ap], for the vocalic gesture {i} (tongue rear) and the following consonantal gesture {β} (lower lip).

(2a) A vocalic gesture and the leftmost consonantal gesture of a preceding associated sequence are phased so that the target of the consonantal gesture (240 degrees) coincides with a point after the target of the vowel (about 330 degrees).
This is abbreviated as follows:
$C(240) = = V(330)$

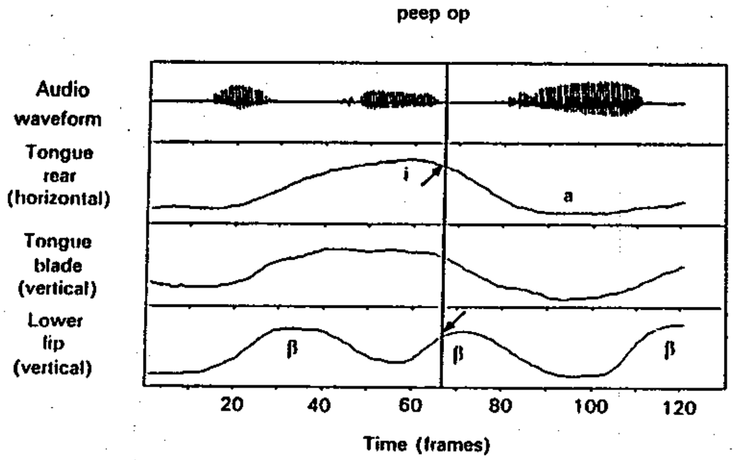


Figure 19.8 X-ray pellet trajectories for "peep op" ([pip#'ap]), showing phasing for vowel and leftmost following consonant ([ip]). The arrows indicate the points being phased.

Statement (2b) specifies the numerical values for phase relations between a vowel and the preceding leftmost consonant. Figure 19.9a exemplifies this statement in the utterance [pi # 'plats] for the vocalic gesture {a} (tongue rear) and the preceding consonantal gesture {β} (lower lip).

(2b) A vocalic gesture and the leftmost consonantal gesture of a preceding associated consonant sequence are phased so that the target of the consonantal gesture (240 degrees) coincides with the onset of the vocalic gesture (0 degrees).
This is abbreviated as follows:
$C(240) = = V(0)$

To complete our statements about phase relations for a single syllable-sized constellation, we need only specify how the remaining consonantal gestures are phased. Figure 19.9b exemplifies this statement, again using the utterance [pi # 'plats], but this time for two consonantal gestures, the {λ} gesture (tongue blade) and the immediately preceding {β} gesture (lower lip).

(3) Each consonantal gesture in a consonant cluster is phased so that its onset (0 degrees) coincides with the offset of the immediately preceding consonant (about 290 deg.):
$C_n(0) = = C_{n-1}(290)$
A consonant cluster is defined as a well-formed associated consonant sequence. A sequence is well-formed iff it conforms to the syllable structure constraints of the language.

indicates that the consonant(s) precede the associated vowel, while a line slanting left indicates that the consonant(s) follow the associated vowel.

Contiguity operates differently on the C and V tiers. This captures a functional difference between vowels and consonants, where vowels act as a kind of background to the "figure" of the consonants. On the one hand, gestural contiguity on the V tier is completely independent of the C tier. This reflects the fact, to be discussed in detail in the next section, that vowel articulations are contiguous (or partially overlapping), regardless of the number of consonants intervening. On the other hand, contiguity on the C tier is sensitive to intervening vocalic gestures, in the sense that consonantal gestures overlap considerably less, if at all, when a vocalic gesture intervenes. A related effect of this functional difference has to do with the change in stiffness as more consonants are inserted between vowels: the vowel gestures decrease in stiffness while the consonantal gestures increase their stiffness.

19.2.2 *Using functional tiers with phasing*

In this subsection, we explore some details concerning the specification of phase relations among gestures on the C and V tiers. That is, given that gestures are spatio-temporal in nature, we need to be able to specify how they are coordinated – we cannot simply assume they are coordinated onset-to-onset, or onset-to-target. Figure 19.7b shows a schematic representation of the phase associations for our sample symbolic gestural score. Here, the only association lines from figure 19.7a that remain are those that phase gestures relative to each other. Statement (1) summarizes the phasing associations that hold between the V and C tiers:

(1) A vocalic gesture and the leftmost consonantal gesture of an associated consonant sequence are phased with respect to each other. An *associated consonant sequence* is defined as a sequence of gestures on the C tier, all of which are associated with the same vocalic gesture, and all of which are contiguous when projected onto the one-dimensional oral tier.

Notice that this phasing association statement follows the unmarked pattern of associations for tiers in autosegmental phonologies (left precedence).

Statement (2a) specifies the numerical values for phase relations between a vowel and the following leftmost consonant. An example of this phasing can be seen in the X-ray pellet trajectory data depicted in figure 19.8 for [pip#ˈap], for the vocalic gesture {i} (tongue rear) and the following consonantal gesture {β} (lower lip).

(2a) A vocalic gesture and the leftmost consonantal gesture of a preceding associated sequence are phased so that the target of the consonantal gesture (240 degrees) coincides with a point after the target of the vowel (about 330 degrees).
This is abbreviated as follows:
$C(240) == V(330)$

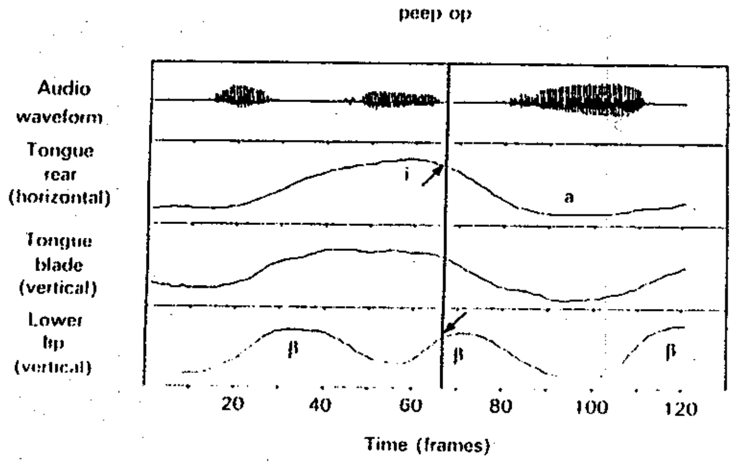


Figure 19.8 X-ray pellet trajectories for "peep op" ([pip#ˈap]), showing phasing for vowel and leftmost following consonant ({ip}). The arrows indicate the points being phased.

Statement (2b) specifies the numerical values for phase relations between a vowel and the preceding leftmost consonant. Figure 19.9a exemplifies this statement in the utterance [pi#ˈplats] for the vocalic gesture {a} (tongue rear) and the preceding consonantal gesture {β} (lower lip).

(2b) A vocalic gesture and the leftmost consonantal gesture of a preceding associated consonant sequence are phased so that the target of the consonantal gesture (240 degrees) coincides with the onset of the vocalic gesture (0 degrees).
This is abbreviated as follows:
$C(240) == V(0)$

To complete our statements about phase relations for a single syllable-sized constellation, we need only specify how the remaining consonantal gestures are phased. Figure 19.9b exemplifies this statement, again using the utterance [pi#ˈplats], but this time for two consonantal gestures, the {λ} gesture (tongue blade) and the immediately preceding {β} gesture (lower lip).

(3) Each consonantal gesture in a consonant cluster is phased so that its onset (0 degrees) coincides with the offset of the immediately preceding consonant (about 290 deg.):
$C_n(0) \quad C_{n-1}(290)$
A consonant cluster is defined as a well-formed associated consonant sequence. A sequence is well-formed iff it conforms to the syllable structure constraints of the language.

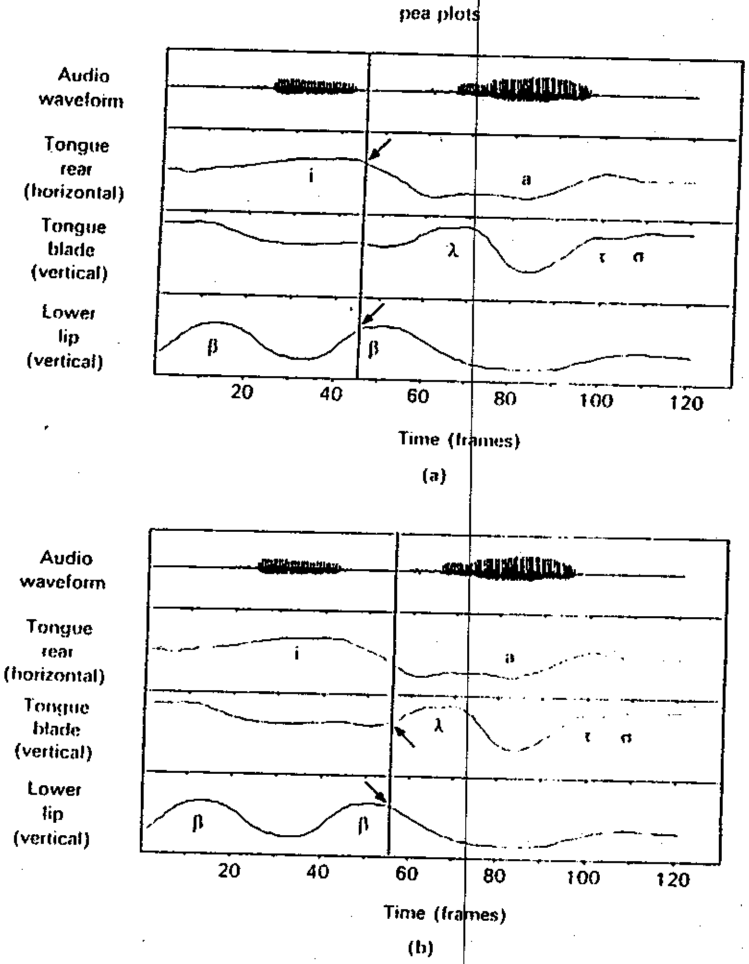


Figure 19.9 X-ray pellet trajectories for "Pea plots" [pi#ˈplats]. (a) Phasing for vowel and leftmost preceding consonant ([#ˈp...a]). (b) phasing for consonants ([pl]).

The above set of statements is sufficient to phase the vocalic and consonantal gestures in a single syllable-sized constellation of gestures; it does not, however, completely account for the phase relations required for the entire utterance [pis#ˈplats]. (Nor, of course, does it account for gestures on the other tiers, the glottal and velic. We will have nothing to say about the phasing of gestures on these tiers in this paper.) One more statement is needed, one that somehow

coordinates the two constellations we are dealing with. For the X-ray data we are exploring, this additional statement appears to associate the leftmost of a sequence of intervocalic consonants with both vowels (an analog of ambisyllabicity):

(4) The leftmost consonantal gesture of a consonant sequence intervening between two vocalic gestures is associated with both vocalic gestures.
A consonant sequence is defined as intervening iff the entire sequence lies between the two vocalic gestures when projected onto the one-dimensional oral tier.

Once statement (4) has applied to associate a consonantal gesture with the vocalic gesture in the neighboring constellation, the structural descriptions of statements (2) and (3) are satisfied for this new C-V association so that they automatically apply. A symbolic display of this process is seen in figures 19.10a and 19.10b for the utterance [pis#ˈplats], showing how the two constellations in figure 19.7 are associated and phased with respect to each other. Figure 19.11 exemplifies the process using X-ray trajectories for the utterance [pis#ˈplats]. This reapplication of statement (2) thus phases the onset of a vocalic gesture with respect to the target of the leftmost consonant of a preceding associated consonant sequence, regardless of the canonical syllable affiliation of that consonant. In the reapplication, the potential transitivity of statements (2a) and (2b) is activated, so that the two vocalic gestures are effectively phased to each other. That is, the onset of the second vocalic gesture in figures 19.7b and 19.10b will coincide with 330 degrees of the first vocalic gesture. (This will also be the default in the case of two vocalic gestures with no intervening consonantal gestures.)

The phasing statements (2)–(4) are a very sketchy beginning to the task of specifying the relations among gestures. How likely is it that they are a dependable beginning? We expect all of them, with the possible exception of (2a), to be confirmed, at least in essence, in future investigations. We are fairly confident about statements (2b) and (4) (although the precise numbers for the phases may need refining), since similar observations have been made by others. For example, Gay (1977, 1978) has shown that tongue movement toward the second vowel in a VCV sequence begins shortly after the onset of acoustic closure for the consonant. Similarly, Borden and Gay (1979) have shown such vowel movements beginning during the /s/ of /sC/ clusters. We also are struck by the convergence of these articulatory phasing statements with principles of syllable affiliation proposed by phonologists. For example, statement (2b) is the articulatory analog of the Principle of Maximal Syllable Onset (cf. Clements and Keyser 1983), while statement (4) is the articulatory analog (for oral gestures) of Resyllabification (Clements and Keyser 1983).

We are not particularly confident about the related statement (2a), however, because there is a complex interaction between phasing and stiffness, at least for vowels, about which we still understand very little. This is true in particular when

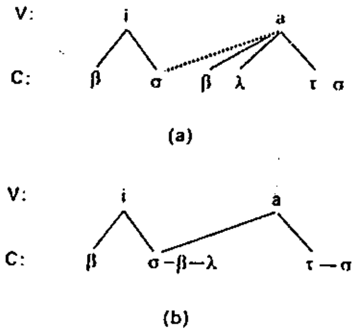


Figure 19.10 Resyllabified consonant and vowel tier displays for "piece plots" ([pis#ˈplats]). (a) Associations (overlap), (b) phasing.

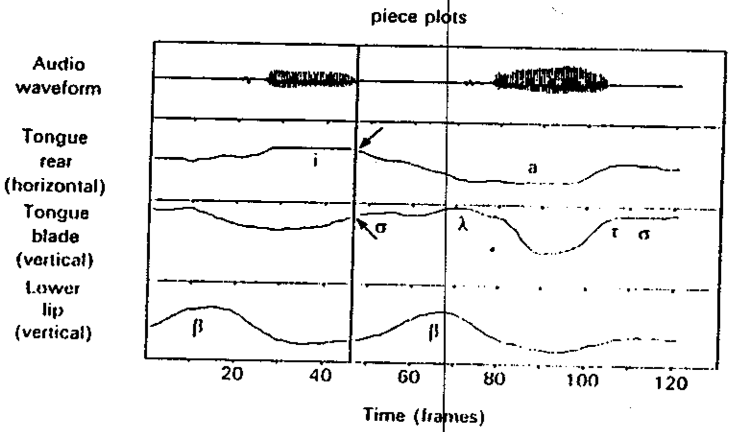


Figure 19.11 X-ray pellet trajectories for "piece plots" ([pis#ˈplats]), showing resyllabified phasing for vowel and leftmost preceding consonant ([s#ˈ...a]).

two constellations are concatenated, as in our example. Here, rephasing the second vowel to an earlier preceding consonant (as in the rephasing between figures 19.7b and 19.10b) will either cause the target of that vowel to be reached earlier, or necessitate a modification of the stiffness of the vocalic gesture so that its target is reached at about the same time relative to the immediately preceding consonant. If the stiffness of the vowel is changed, however, it will change the temporal relations between the vowel and the following consonantal gestures, assuming their phasing remains the same. While the current version of our phasing rules in the computational model uses the second approach, that of modifying the stiffness of the vocalic gesture, much more work is needed in this area.

Statement (3), on the timing of onsets to preceding releases for consonants, has been reported elsewhere (e.g. Kent and Moll 1975), and so we are fairly confident about this statement as well. Kent and Moll's data also show that the position of syllable boundaries is irrelevant, at least for sequences that are possible well-formed syllable onsets. The exact formulation of the statement, however, is a matter of considerable interest to us. In particular, to what extent is the timing of a consonant dependent on the well-formedness of the consonant sequence (in terms of not violating syllable structure constraints)? In the next section, we explore some implications of this and other aspects of our proposed gestural structures for describing casual speech.

19.3 Generalizations about casual speech

The gestural representations that we have outlined above form the basis of a simplified account of the phonological/phonetic alternations that occur in continuous, fluent speech. In particular, a number of superficially unrelated alternations (unrelated in that their description requires separate phonological rules of the conventional sort) can be shown to follow from a generalization about gestural organizations and how they may be modified in the act of talking. The power of such generalizations follows from the incorporation of the spatio-temporal nature of speech in the representation, both in terms of the definition of individual gestures as events occurring in space and time, and in the explicit specification of the spatio-temporal (phase) relations among gestures.

There have been a number of attempts by linguists to characterize the differences between the pronunciation of words in isolation and their realization in "casual" connected speech (e.g., Zwicky 1972; Shockey 1973; Oshika, Zue, Weeks, Neu, and Auerbach 1975; Kahn 1976; Brown 1977; Dalby 1984; Barry 1984). In this paper, we define "casual" speech as that subset of fast speech in which reductions typically occur. In casual speech, then, there are typically gross restructurings between the "ideal" phonetic representation of a word – its canonical form – and a narrow phonetic transcription of its form in context. Segments are routinely elided, inserted, and substituted for one another. The examples in (5) (taken from Brown 1977) show (a) consonant deletion, (b) consonant assimilation, and (c) simultaneous deletion and assimilation.

(5) (a) /ˈmʌst bi/ → [ˈmʌsbi] ("must be")
(b) /hʌndrəd ˈpaUndz/ → [hʌndrəbˈpaUndz] ("hundred pounds")
(c) /ˈgraUnd ˈprɛʃɚ/ → [ˈgraUmˈprɛʃɚ] ("ground pressure")

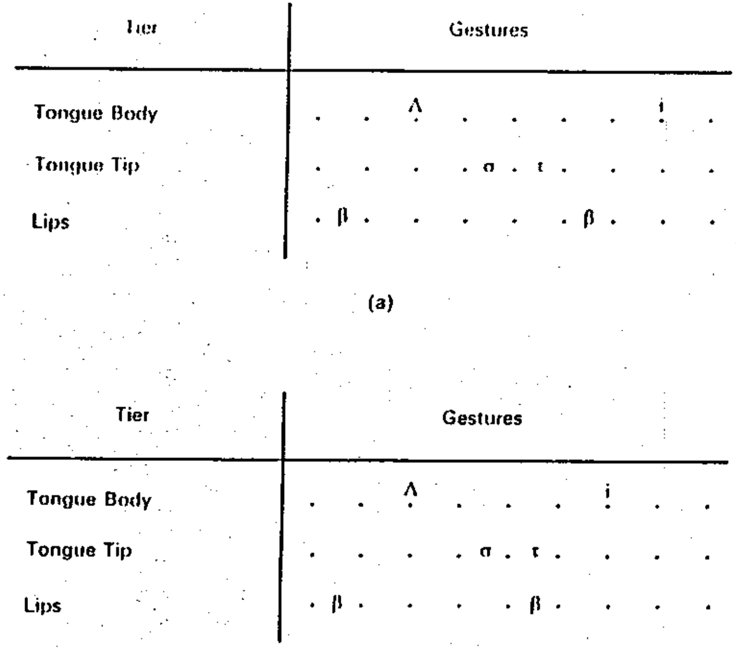
Thus, the narrow phonetic transcription of a word in context can be radically different from its systematic phonetic representation. While a number of the above authors have attempted to describe such changes with lists of phonological rules that apply in casual, or fluent speech, these lists fail to uncover generalizations about casual speech that underlie these particular changes. Such generalizations do emerge, however, from a description of these changes in terms of the variation in their gestural scores.

From the gestural point of view, the relationship between the lexical characterization of a word and its characterization in connected speech is much simpler and more highly constrained. We propose that most of the phonetic units (gestures) that characterize a word in careful pronunciation will turn out to be observable in connected speech, although they may be altered in magnitude and in their temporal relation to other gestures. In faster, casual speech, we expect gestures to show decreased magnitudes (in both space and time) and to show increasing temporal overlap. We hypothesize that the types of casual speech alternations observed (segment insertions, deletions, assimilations and weakenings) are consequences of these two kinds of variation in the gestural score.

19.3.1 *Gestural overlap within and across tiers*

When two gestures overlap in time, we expect to see different consequences (in actual articulator movement) depending on whether the two gestures are on the same or different articulatory tiers, that is, depending on the articulatory organization. Gestures on different tiers may overlap in time and yet proceed relatively independently of one another, without perturbing each other's trajectories, since they affect independent vocal tract variables. The possibility of such events on separate tiers "sliding" in time with respect to one another provides the basis for an analysis of the apparently diverse changes in (5).

Across tiers. Example (5a) is described as an example of segment deletion. However, looking at this change in terms of the gestures involved, we hypothesize that the alveolar closure gesture for the /t/ is still present in the fluent speech version, but that it has been completely overlapped, or "hidden," by the bilabial closure gesture. This means that the movement of the tongue tip towards the alveolar ridge and away again may occur entirely during the time that the lips are closed (or narrowed), so that there will be no local acoustic evidence of the alveolar closure gesture. Figure 19.12 shows the hypothesized variation in the symbolic gestural score for "must be." Only the oral subsystem is shown. In figure 19.12a, the alveolar closure precedes the bilabial closure. This implies that the gestural overlap is only partial. In figure 19.12b the gestures associated with "be" have slid to the left so that the bilabial closure is effectively synchronous with the alveolar gesture. This view contrasts sharply with the more traditional description that there is a fluent speech rule that deletes the /t/ in the appropriate environments. Under the latter hypothesis, one would not expect to find any articulatory movement associated with an alveolar closure. Articulatory evidence of such hidden closures is presented in the next section.



(b)

Figure 19.12 Hypothetical symbolic gestural scores (oral tiers only) for "must be". The symbol ʌ has its usual phonetic interpretation in this figure. (a) Canonical form ([ˈmʌst#bi]). (b) fluent speech form ([ˈmʌsbi]).

Example (5b) is described as an assimilation rather than a deletion. Nonetheless, the same kind of analysis can be proposed. The bilabial closure gesture may increase its overlap with the preceding alveolar gesture, rendering it effectively inaudible. The overlap of voicing onto the beginning of the bilabial closure yields the [bp] transcription. The possibility of analyzing assimilations in this way is also proposed in Kohler (1976) for German, and in Barry (1984). Brown (1977) also notes that it is misleading to view such changes as replacing one segment with another, but she does not propose a formal alternative.

The combination of assimilation and deletion observed in (5c) can be analyzed in the same way. The bilabial closure gesture (associated with the /p/) increases its overlap with the alveolar closure gesture (associated with the /nd/). The fact that the velic lowering gesture for the /n/ now overlaps the bilabial closure accounts for the appearance of [m]. Thus, these examples of consonant assimilation and consonant deletion are all hypothesized to occur as a result of increasing gestural overlap between gestures on separate oral tiers.

In fact, the most common types of place of articulation assimilations in casual speech do involve gestures on separate oral tiers. At least for RP, Brown (1977) claims that the most common cases involve alveolar stops assimilating to labials or velars (see also Gimson 1962). Thus, the common assimilation types represent two of the three possible combinations of gestures from two separate tiers. One might ask why labial-velar or velar-labial assimilations do not occur (at least not frequently), given the possibility of their overlapping. The answer to this question would involve studying the acoustic and perceptual consequences of overlapping vocal tract constriction movements (we intend to do this using the ability of our model to generate speech). A possible explanation lies in the fact (Kuehn and Moll 1976) that tongue tip movements show higher velocities than do either tongue dorsum or lip movements (which are about equivalent to each other). A slower movement might prove more difficult to hide.

Within tiers. Gestures on the same articulatory tier cannot overlap without perturbing each other, since the same vocal tract variables are employed but with different targets. Thus, even partial overlap of gestures on the same tier leads to *blending* of the observed output characteristics of the two gestures. This same point has been made by Catford (1977), who distinguishes what he calls "contiguous" sequences (which typically involve the same tract variable in the present system), from "heterorganic" sequences (which typically involve different tract variables). The blending of gestures shows itself in spatial changes in one or both of the overlapping gestures.

Variation in the overlap of gestures on the same tier can account for other types of assimilations (somewhat less common, according to Brown 1977) that cannot be accounted for in terms of the "hiding" notion proposed to account for the changes in (5). Some examples are shown in (6) (a,c from Catford 1977; b from Brown 1977):

(6) (a) /tɛn 'θɪŋgz/ → [tEn̪'θɪŋgz] ("ten things")
(b) /kʌm fram/ → [kʌɱfram] ("come from")
(c) /ðɪs'ʃap/ → [ðɪʃ'ʃap] ("this shop")

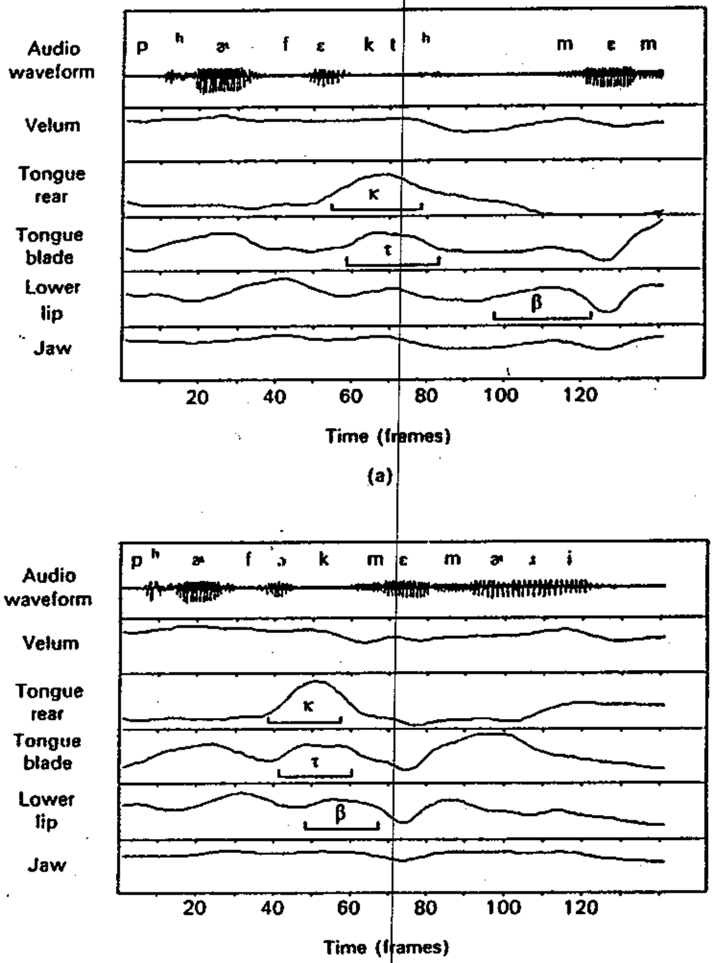
For example, in (6a), the overlapping production of the alveolar closure (associated with the /n/) and the dental fricative leads to a more fronted articulation of the alveolar closure (and perhaps more retracted articulation of the dental fricative). Similar interactions are found between a bilabial stop and labiodental fricative (in 6b), and between alveolar and palatoalveolar fricatives (in 6c). As Catford (1977) notes, the articulations involved (particularly in cases like 6c) show blending of the two goals into a smooth transition, rather than a substitution of one segment for another, as the transcription would indicate.

Articulatory vs. functional structure. The examples of overlap and blending considered so far have all involved consonantal gestures. The consequences of gestures being on the same or different articulatory tiers are not restricted to consonantal gestures, however. Similar consequences have also been observed between consonantal and vocalic gestures. Thus, while consonant and vowel gestures overlap in any CV utterance, only in the case of velar consonants does this overlap occur on a single articulatory tier (tongue body), and in this case we would expect the gestures to show blending. Indeed they do, and this blending is usually described as fronting of the velar consonant in the context of following front vowels. In contrast, alveolar stops can be produced concurrently with a vowel without changing the place of articulation of the stop. Öhman's (1967) X-rays of the vowel tract during the closures for /idi/, /udu/, and /ada/ show global shapes determined almost completely by the vowel, but with a relatively invariant tongue tip constriction superimposed on the vowel shapes. For /igi/ and /ugu/, however, the actual position of the constriction is shifted.

Thus, articulatory organization (i.e. whether gestures occur on the same or different articulatory tiers) appears to be relevant to questions of blending and overlap, regardless of the functional status of the gestures. That is, it is relevant for both consonantal and vocalic gestures. However, the effect of articulatory structure can also interact with the functional status of the gestures involved. For example, Keating (1985) points out the effect that language-particular distinctiveness requirements can have on coarticulation. She discusses Öhman's (1966) findings that Russian consonants, which involve contrastive secondary tongue body articulations (palatalization vs. velarization), do block V-to-V tongue body coarticulation, whereas English tongue body consonants, which involve contrastive primary articulations, do not block V-to-V tongue body coarticulation. Following Öhman, she models this effect by placing the secondary tongue body articulation for the Russian consonants on the vowel tier. Thus, the functional status of the "consonantal" tongue body gestures in the two languages also affects the amount of blending observed.

19.3.1.1 EVIDENCE FOR HIDDEN GESTURES

If our analysis of the changes involved in examples like (5) is correct, then it should be possible to find articulatory evidence of the "hidden" alveolar gesture. We examined the AT&T X-ray database (described in section 2) for examples of consonantal assimilations and deletions of this kind, by listening to sentences with candidate consonant sequences. Although there were very few examples of assimilations or deletions in the corpus (talking with lead pellets in your mouth and an X-ray gun pointed at your head hardly counts as a "casual" situation), the examples we were able to find do show the hidden gesture. For example, figure 19.13a shows the vertical displacements of lead pellets placed on the velum, tongue dorsum rear, tongue blade, lower lip and lower teeth, along with the acoustic waveform, for the utterance "perfect memory," spoken as a sequence of two words separated by a pause. The phonetic transcription aligned with the acoustic



(b)

Figure 19.13 X-ray pellet trajectories for "perfect memory." (a) Spoken in a word list ([pɚfɛkt#'mɛm...]). (b) spoken in a phrase ([pɚfɛk'mɛm...]).

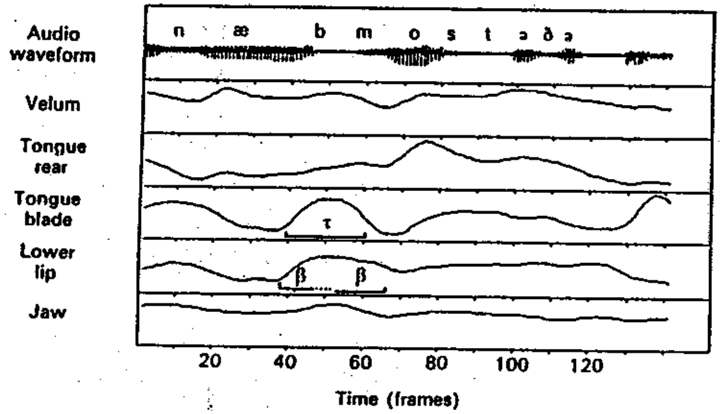


Figure 19.14 X-ray pellet trajectories for "nabbed most" ([ˈnæbmost], spoken in a phrase.

waveform indicates that the /t/ at the end of "perfect" is completely audible and its release is visible in the waveform. The time-course of the velar closure gesture associated with the /k/ in "perfect" is assumed to be reflected in the vertical displacement of the tongue dorsum (tongue rear) pellet. The relevant portion of this pellet trajectory is underlined in the figure and is labeled with the appropriate gestural symbol. Similarly, the portion of the tongue blade displacement associated with the alveolar closure gesture for the /t/ in "perfect," and the portion of the lower lip displacement associated with the bilabial closure gesture for the initial /m/ in "memory" have been marked and labeled in the figure. Note that the velar and alveolar gestures partially overlap, indicating that velar closure is not released until the alveolar closure is formed. Thus, the *onset* of the alveolar gesture is acoustically "hidden" (it takes place during the velar closure), but its release is audible. Note the large amount of time between the release of the alveolar gesture and the onset of the bilabial gesture.

Figure 19.13b shows the same two word sequence spoken as part of a sentence. Here, the final /t/ in "perfect" is deleted in the traditional sense – careful listening reveals no evidence of the /t/, and no /t/ release can be seen in the waveform. However, the alveolar gesture can still be seen quite clearly in the figure. It is even of roughly of the same magnitude as in figure 19.13a. What differs here is that the bilabial gesture for the initial /m/ now overlaps the release of the alveolar gesture. Thus, both the closure and release of the alveolar gesture are now overlapped and there is, therefore, no acoustic evidence of its presence. The existence of this kind of phenomenon is consistent with Fujimura's (1981) iceberg account of the X-ray corpus from which these tokens are taken. He proposed that certain articulatory movements (like those forming the onsets and offsets of the gestures in figure 19.13) remain relatively invariant across phrase positions in which a word may occur, but that they may "float" relative to other icebergs. The hiding we observe here is a consequence of that floating.

Another example of alveolar stop "deletion" where the alveolar closure gesture remains can be seen in figure 19.14. The same pellet trajectories are shown as in the previous figure. Here, the speaker (the same one shown in the previous figure) is producing the phrase "nabbed most" in a sentence. As indicated by the phonetic transcription, the /d/ at the end of "nabbed" has been deleted. The bilabial gestures associated with the /b/ of "nabbed" and the /m/ of "most" here overlap (forming a continuous closure), and the alveolar closure gesture, while quite robust kinematically, is once again irrelevant acoustically.

The X-ray data also provide an example of assimilation in which an alveolar closure gesture is hidden by bilabial gestures. Figure 19.15 shows x-ray pellet trajectories for a (second) speaker producing the phrase "seven plus seven times..." Pellet locations are the same as for the first speaker, except that there is no velum pellet. As indicated in the transcription, the version in figure 19.15b contains an assimilation of the final /n/ to [m]. Note however, that the alveolar closure gesture (as outlined on the tongue blade trajectory) is still present in the assimilated version. Comparing the two figures, it is not completely clear how to account for their different acoustic properties. They do not show the radical timing differences shown in figure 19.13. The difference may simply reside in the alveolar gesture being somewhat reduced in magnitude in the assimilated version. As suggested earlier, reduction in gestural magnitude is the other major kind of change that we expect to observe in casual speech, and we will return to this aspect of change in section 19.3.2. The example does clearly show, however, the presence of a hidden gesture (though perhaps somewhat reduced) in an assimilation.

While the amount of data we have analyzed is still quite small, these tokens do demonstrate that gestural overlap can lead to apparent assimilations and deletions. Further supporting data were found by Barry (1985) who analyzed assimilations in a manner similar to that proposed here. He presents electropalatographic evidence of "residual" alveolar articulations in cases of alveolar-velar and alveolar-bilabial assimilations. Some electropalatographic evidence of such "residual" alveolars is also presented by Hardcastle and Roach (1979) and for German by Kohler (1976). Barry also shows cases of assimilation that do not include these residual articulations. Such cases may involve reduction in the magnitude of the hidden gesture (as suggested for figure 19.15b), and it is possible that such reduced gestures would not show up in a technique that depends on actual contact. Alternatively, some cases of assimilation may involve complete gesture deletion. Even deletion, however, can be seen as an extreme reduction, and thus as an endpoint in a continuum of gestural reduction, leaving the underlying representation unchanged.

Indirect evidence in support of the hidden gesture analysis of alveolar stop deletion can be found in variable rule studies of deletion of final /t,d/ in clusters (Labov, Cohen, Robins and Lewis 1968; Guy, 1975; Guy, 1980; Neu, 1980). These studies show that the deletion is more likely to take place if the following word begins with a consonant than if it begins with a vowel. In studies with enough data to permit analysis according to the initial segment of the following word (Guy, 1980; Neu, 1980), the greatest deletion probabilities occur when the following word begins with (true) consonants, followed in order by liquids, glides and vowels.

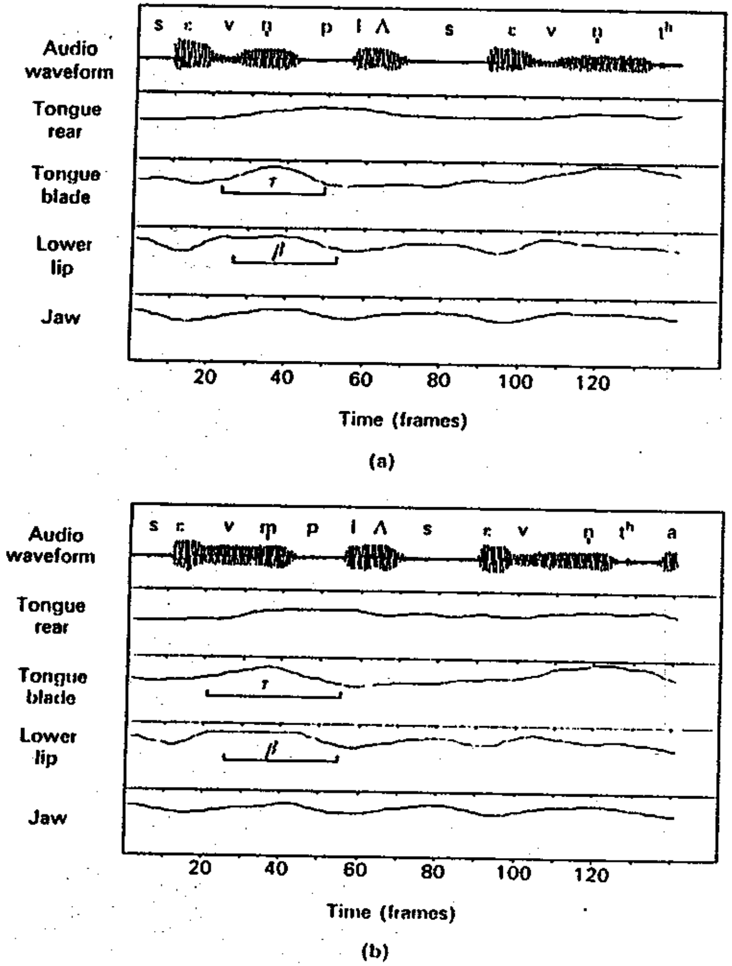


Figure 19.15 X-ray pellet trajectories for "seven plus seven times." (a) Not assimilated ([sɛvn̩#plʌs...]), (b) assimilated ([sɛvm̩plʌs...]).

This consonant-to-vowel ordering is exactly what we would expect when we consider the consequences of gestural overlap. In general, given comparable overlap patterns, the more extreme the constriction associated with a gesture, the better able that gesture is to acoustically mask (or aerodynamically interfere with) another gesture with which it is co-occurring (cf. Mattingly 1981). Thus, overlap by a following consonant (stop or fricative) gesture would be most able to contribute to hiding of an alveolar closure gesture, while a following vowel would presumably not contribute at all to hiding the gesture (indeed, as noted above, overlap of consonants with vowels is the usual case). In the case of a following vowel, the alveolar gesture itself must either be severely reduced or overlapped completely by a preceding consonant gesture. Thus, the ordering of probabilities on deletion of final /t,d/ in clusters could follow directly from the view of deletion that we are proposing here, without these differential probabilities needing to be "learned" as part of a rule. This is consistent with Guy's (1980) demonstration that this consonant-liquid-glide-vowel ordering is robust across dialects and across individual talkers, while the ordering of other factors that contribute to deletion probability (e.g. morphological status of the final /t,d/) may differ greatly from dialect to dialect.

19.3.1.2 THE EMERGENCE OF VARIATION IN GESTURAL ORGANIZATION

While the evidence summarized supports an overlapping gestural analysis of consonant assimilations and deletions, we would also like to understand how such variation in gestural scores arises. If we think of gestural scores as providing an organization of gestures that allows them to overlap, yet to all be perceived, why, and under what circumstances, should this organization "fail," in the sense that some gestures become imperceptible?

Recall that for V C(C)(C) V utterances, we proposed that the oral consonant gestures are phased with respect to the immediately preceding one (statement 3), as long as the sequence conforms to the language's syllable structure constraints. The original syllable affiliation of the consonants does not matter – what is important is the fact of well-formedness of the sequence. For those cases we examined, (e.g., [s # pl] vs. [# spl]), the sequences conform to possible syllable onsets, even when there is an intervening word boundary (i.e. [s # pl]). The cases that lead to complete gestural overlap, however, involve sequences that do not form possible syllable onsets (or codas), i.e. alveolar closure-bilabial closure and alveolar closure-velar closure sequences. We propose that in these cases (for example the one shown in figure 19.12b), the phasing principles in (3), and possibly (2) and (4) as well, are not followed. While it is not yet clear exactly what kind of phasing will be involved in these cases, we expect that the structure will *not* involve phasing the consonants to one another in sequence. The lack of such sequential phasing would then allow the kind of overlap we saw in figures 19.13b, 19.14 and 19.15b to emerge.

The view we are proposing then suggests that the gestural organization in a language is exquisitely tuned to allow the successive oral gestures in syllable onsets and codas to overlap partially, without obscuring the information in these gestures.

This view has been propounded by Mattingly (1981), who argues that "the syllable has more than a phonological or prosodic role; it is the means by which phonetic influences [cf. gestures] are scheduled so as to maximize parallel transmission" (p. 418). As Mattingly suggests, this organization guarantees, for example, that the constriction and release of an [l] will not occur completely during the bilabial closure in a [pla] utterance. However, for sequences that are not possible syllable onsets (or codas), we hypothesize that the production system does not have the same kind of tight organization available. Thus, for such sequences, variation in degree of overlap is possible, even to the point of obscuring the parallel transmission of information (i.e. one gesture may hide another).

Some indirect evidence supports the view that status as possible cluster correlates with tightness of gestural organization. Alveolar stops frequently assimilate in alveolar-labial and alveolar-velar sequences, but assimilation (in either direction) is rare in labial-alveolar and velar-alveolar sequences. This could simply be due to the relative scarcity of word-final labial and velars compared to final alveolars (see Gimson 1962). A more interesting interpretation, from the current perspective, attributes these asymmetries to syllable structure differences: the labial-alveolar and velar-alveolar sequences are all possible syllable codas, and thus would be expected to show a tight phasing organization that prevents complete overlap. However, the details of the actual articulatory movements in such cases need to be examined before our proposal can be explicitly worked out and evaluated. In addition, the possibility that postconsonantal final alveolars are not always part of the syllable core (e.g. Fujimura and Lovins 1978) needs to be addressed.

19.3.2 *Other casual speech processes*

The aspect of an articulatory phonology that makes it potentially powerful for describing continuous speech is that diverse types of phonetic alternation – segment insertion, deletion, assimilation and weakening – can all be described in terms of changes in gesture magnitude or intergestural overlap. That is, these alternations, which might require a number of unrelated segmental rules, can be given a unified account in the gestural framework. In the previous subsection, we showed how variation in intergestural overlap can give rise to apparent consonant deletions and assimilations in casual speech. In this subsection, we suggest how variation in gestural overlap, and also gestural magnitude, might yield some other types of alternations.

In addition to consonant deletions, *schwa deletions* are common in casual speech (Brown 1977; Dalby 1984). Just as with the apparent consonantal deletions described above, such apparent schwa deletions might result from variation in gestural overlap. For example, the apparent deletion of the second vowel in "difficult" (Shockey 1973) might instead be an increase in overlap between the labiodental fricative gesture and the following velar closure gesture, so that the fricative is not released before the closure is formed (see Catford 1977, on open vs. close transitions).

Apparent *segment insertions* might also arise from variation in gestural overlap, as the following examples suggest:

(7) (a) /ˈsʌmθɪŋ/ → [ˈsʌmpθɪŋ] ("something")
(b) /ˈsæmsən] → [ˈsæmpsən] ("Samson")

A number of authors (e.g. Ohala 1974; Anderson 1976) have analyzed the epenthetic stop in such nasal-fricative sequences as arising from variation in the relative timing of the velic closure gesture and the preceding oral closure gesture. In particular, if denasalization precedes the release of the closure gesture, then a short interval of oral closure will be produced. These changes in timing could be directly accommodated within our gestural structures.

Brown (1977) also identifies a class of *weakenings*, or lenitions, in casual speech. Typical examples involve stop consonants weakening to corresponding fricatives (or approximants), as shown in (8):

(8) (a) /bɪˈkʌz/ → [pxʌz] ("because")
(b) /ˈmʌst bi/ → [ˈmʌsβi] ("must be")

These changes might be described as decreases in the magnitude of individual gestures. The reduction in amplitude of movement associated with the gesture then leads to the possibility of an incomplete acoustic closure. (Additionally, reductions in magnitude may combine with increased overlap, leading to greater likelihood of a gesture being "hidden" by another gesture: see figure 19.15.) Such reductions of movement amplitude often, but not always, occur in fast speech (Lindblom 1983; Munhall, Ostry, and Parush, 1985; Kuehn and Moll 1976; Gay 1981). It may also be the case that gestural magnitudes are reduced simply because the speaker is paying less attention to the utterance (Dressler and Wodak 1982; Barry 1984).

Reduction of gestural magnitude may be involved in changes other than weakenings. For example, the classic cases of intervocalic *voicing assimilation* could be described as a reduction in the magnitude of the glottal opening-and-closing gesture responsible for the voicelessness. If the magnitude of the opening is reduced sufficiently, devoicing might not take place at all. Data from Japanese (Hirose, Niimi, Honda and Sawashima 1985) indicate that the separation between the vocal folds at the point where voicing ceases at the beginning of an intervocalic voiceless stop is much larger than at the point where voicing begins again at the end of the stop. This suggests that if the magnitude of the abduction gestures were slightly reduced, the critical value of vocal fold separation for devoicing might never be reached. This is likely what is going on the data of Lisker, Abramson, Cooper, and Schvey (1969). Using transillumination to measure glottal opening in running speech in English, they found that the vast majority (89%) of English /ptk/ (excluding cases following initial /s/ and environments that allow flapping) were, in fact, produced with glottal opening, but of these, 11% showed no break in glottal pulsing. While the magnitude of glottal opening was not measured, we would hypothesize that these cases involved a decrease in magnitude of the opening gesture.

19.4 Summary

We have discussed a computationally explicit representation of articulatory organization, one that provides an analytical, abstract description of articulation using dynamically defined articulatory gestures, arranged in gestural scores.

We first showed how *canonical phonological forms* can be described in the gestural framework, presenting preliminary results that syllable structure is best represented using separate vowel and consonant tiers, such that consonantal gestures overlap the vowel gesture with which they are associated. We also suggested that the vocalic gestures are most closely associated with the leftmost consonantal gesture in a consonant sequence, and that well-formedness of a consonant cluster is revealed by lack of variability in the overlap among the gestures constituting the cluster.

We then showed how it might be possible to describe all reported phonological changes occurring in *casual speech* as consequences of variation in the overlap and magnitude of gestures. That is, in the gestural approach these two processes are seen as underlying the various superficially different observed changes. We presented some examples of "hidden" gestures, in which the articulations associated with the gesture were still observable in fluent speech, although there were no perceptible acoustic consequences of the gesture. We further discussed the importance of articulatory structure in fluent speech: overlap of gestures has very different consequences, depending on whether the gestures are on the same or different articulatory tiers.

The gestural approach to casual speech is *highly constrained* in that casual speech processes may not introduce units (gestures), or alter units except by reducing their magnitude. This means that all the gestures in the surface realization of an item are present in their lexical representation; casual speech processes serve only to modify these gestures, in terms of diminution or deletion of the gestures themselves, or in terms of changes in overlap. Phonological rules of the usual sort, on the other hand, can introduce arbitrary segments, and can change segments in arbitrary ways.

The gestural approach is further constrained by its reliance on articulatory structure, by its use of task dynamics to characterize the movements, and by our insistence on computational explicitness. All of these constraints lead to directions for future research. Will our suggested structures, both for the canonical and casual speech forms, be confirmed by further articulatory studies? Can articulatory structure provide simpler solutions to phonological problems such as the specification of language-particular syllable templates? Do dynamic parameters, that in our system are attributes of gestures, participate in phonological patterns that are inherently different from those patterns involving articulatory structure? Such questions are but part of the challenge for an articulatory phonology.

Appendix A
Mass-spring model

A simple dynamical system consists of a mass attached to the end of a spring. If the mass is pulled, stretching the spring beyond its rest length (equilibrium position), and then released, the system will begin to oscillate. The resultant movement patterns of the mass will be a damped sinusoid described by the solution to the following equation:

$$m\ddot{x} + b\dot{x} + k(x - x_0) = 0$$

where m = mass of the object
b = damping of the system
k = stiffness of the spring
x_0 = rest length of the spring (equilibrium position)
x = instantaneous displacement of the object
$\dot{x}$ = instantaneous velocity of the object
$\ddot{x}$ = instantaneous acceleration of the object

Note that the time-varying motion of the sinusoid is described by an equation whose parameters do not change over time. The equation constitutes a global constraint on the form of the movement; different trajectory shapes can be obtained by substituting different values for the parameters m, k, and x_0. When such an equation is used to model the movements of coordinated sets of articulators, the "object" - motion variable – in the equation is considered to be the task, for example, the task of lip aperture. Thus, the sinusoidal trajectory would describe how lip aperture changes over time. In task dynamics, the task is treated as massless, since it is the motion of an abstract entity, the tract variable, that is being modelled, rather than the movement of physically massive articulators. For further details on task dynamics, see Saltzman (1986).

Appendix B
Phonetic identity and stress

Phonetic identity is conveyed primarily by the values of the equilibrium positions for the tract variables. The values of the equilibrium positions (targets) for the tract variables LA, TBCD, TTCD, GLO, and VEL refer to constriction degree, while the targets for LP, TBCL, and TTCL refer to the location of the constriction with respect to the upper or back wall of the vocal tract (LP refers to lip protrusion). While we have available to us the complete numerical continuum, our initial modeling relies on categorical approximations. Thus, for the oral constriction degree tract variables, there are a maximum of 7 discrete values: closure, "critical" (that constriction appropriate for generating frication), narrow, narrow-mid, mid, mid-wide, and wide. The tongue constriction location variables also have a maximum of 7 discrete values: dental, alveolar, alveo-palatal, palatal, velar, uvular, and pharyngeal. In the current preliminary (and incomplete) formulation, the first three locations utilize TTCL and the last four TBCL. The glottal and velic tract variables are currently not explicitly modeled by task dynamics; instead the acoustic consequences of the articulatory movements are approximated by simple on-off functions. The glottal tract variable has three modes: one approximating an opening-and-closing gesture (for voicelessness), one approximating a (tight) closing-and-opening gesture (for glottal stops), and one for speech mode (voicing); the velic tract variable is binary valued, i.e. either open or closed.

The durational and segmental variations associated with differences in *stress level* can be simply accounted for in terms of dynamically defined articulatory movements (Browman and Goldstein 1985; Kelso, Vatikiotis-Bateson, Saltzman, and Kay 1985; Ostry and Munhall 1985). Decreasing the stiffness k of the "spring" for the stressed vowel results in a slower trajectory, which corresponds to the longer duration associated with stress. Increasing the difference between the rest length (equilibrium position) of the spring (x_0) and the initial position (x) increases the amplitude of the oscillation, which corresponds to the difference in displacement between a reduced and full vowel. For a consonant, say a labial closure gesture, either decreasing the stiffness or increasing the target (equilibrium position) will increase the duration of the acoustic closure, since in both cases there will be a longer period of time during which the lip aperture will be small enough to achieve acoustic closure.

In our current implementation, different stress levels indicate different multiplicative factors used to modify the inherent dynamic parameters. While this approach appears to work well for the gestural scores we have tested so far, at least for stiffness, it needs to be tested in considerably more detail, including being compared to other possible implementations, such as additivity.

Note

Our thanks to Ignatius Mattingly, Eric Vatikiotis-Bateson, Kevin Munhall, Carol Fowler, Mary Beckman, John Kingston, and Doug Whalen for thoughtful comments on early drafts of this paper; to Yvonne Manning and Caroline Smith for help with the figures; and to Caroline Smith and Mark Tiede for facilitating the analysis of the X-ray data. This work was supported in part by NSF grant BNS 8520709 and NIH grants HD-01994, NS-13870, NS-13617 to Haskins Laboratories.

References

Anderson, S. R. 1976. Nasal consonants and the internal structure of segments. *Language* 52: 326–344.
Barry, M. C. 1984. Connected speech: processes, motivations, models. *Cambridge Papers in Phonetics and Experimental Linguistics* 3: 1–16.
1985. A palatographic study of connected speech processes. *Cambridge Papers in Phonetics and Experimental Linguistics* 4: 1–16.
Borden, G. J. and T. Gay. 1979. Temporal aspects of articulatory movements for /s/-stop clusters. *Phonetica* 36: 21–31.
Browman, C. P. and L. Goldstein. 1985. Dynamic modeling of phonetic structure. In V. Fromkin (ed.) *Phonetic linguistics*. New York: Academic Press, 35–53.
1986. Towards an articulatory phonology. *Phonology Yearbook* 3: 219–252.
Browman, C. P., L. Goldstein, J. A. S. Kelso, P. Rubin, and E. Saltzman. 1984. Articulatory synthesis from underlying dynamics. *Journal of the Acoustical Society of America* 75: S22–S23. (Abstract).

Browman, C. P., L. Goldstein, E. Saltzman, and C. Smith. 1986. GEST: A computational model for speech production using dynamically defined articulatory gestures. *Journal of the Acoustical Society of America* 80: S97. (Abstract).

Brown, G. 1977. *Listening to Spoken English.* London: Longman.

Catford, J. C. 1977. *Fundamental Problems in Phonetics.* Bloomington, IN: Indiana University Press.

Clements, G. N. 1980. *Vowel Harmony in Nonlinear Generative Phonology: An Autosegmental Model.* Bloomington, IN: Indiana University Linguistics Club.

1985. The geometry of phonological features. *Phonology Yearbook* 2: 225–252.

Clements, G. N. and S. J. Keyser. 1983. *CV Phonology: A Generative Theory of the Syllable.* Cambridge, MA: MIT Press.

Dalby, J. M. 1984. Phonetic structure of fast speech in American English. Ph.D. dissertation, Indiana University.

Dressler, W., and R. Wodak. 1982. Sociophonological methods in the study of sociolinguistic variation in Viennese German. *Language and Society* 11: 339–370.

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

1981. A relationship between coarticulation and compensatory shortening. *Phonetica* 38: 35–50.

1983. Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet. *Journal of Experimental Psychology: General* 112: 386–412.

Fujimura, O. 1981. Temporal organization of articulatory movements as a multidimensional phrasal structure. *Phonetica* 38: 66–83.

Fujimura, O., S. Kiritani, and H. Ishida. 1973. Computer controlled radiography for observation of movements of articulatory and other human organs. *Computers in Biology and Medicine* 3: 371–384.

Fujimura, O. and J. Lovins. 1978. Syllables as concatenative phonetic units. In A. Bell and J. B. Hooper (eds.) *Syllables and Segments.* Amsterdam: North Holland, 107–120.

Gay, T. 1977. Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America* 62: 183–193.

1978. Articulatory units: segments or syllables? In A. Bell and J. B. Hopper (eds.) *Syllables and Segments.* Amsterdam: North Holland, 121–131.

1981. Mechanisms in the control of speech rate. *Phonetica* 38: 148–158.

Gimson, A. C. 1962. *An Introduction to the Pronunciation of English.* London: Edward Arnold.

Goldsmith, J. A. 1976. *Autosegmental Phonology.* Bloomington, IN: Indiana University Linguistics Club.

Goldstein, L. and C. P. Browman. 1986. Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics* 14: 339–342.

Guy, G. R. 1975. Use and application of the Cedergren–Sankoff variable rule program. In R. Fasold and R. Shuy (eds.) *Analyzing Variation in Language.* Washington, DC: Georgetown University Press, 56–69.

1980. Variation in the group and the individual: the case of final stop deletion. In W. Labov (ed.) *Locating Language in Time and Space.* New York: Academic Press, 1–36.

Halle, M. 1982. On distinctive features and their articulatory implementation. *Natural Language and Linguistic Theory* 1: 91–105.

Hardcastle, W. J., and P. J. Roach. 1979. An instrumental investigation of coarticulation in stop consonant sequences. In P. Hollien and H. Hollien (eds.) *Current Issues in The Phonetic Sciences.* Amsterdam: John Benjamins AG, 531–540.

Hayes, B. 1981. *A Metrical Theory of Stress Rules.* Bloomington, IN: Indiana University Linguistics Club.

Hirose, H., S. Niimi, K. Honda, and M. Sawashima. 1985. The relationship between glottal opening and transglottal pressure difference during consonant production. *Annual Bulletin of RILP* 19, 55–64.

Kahn, D. 1976. *Syllable-based Generalizations in English Phonology.* Bloomington, IN: Indiana University Linguistics Club.

Keating, P. A. 1985. CV phonology, experimental phonetics, and coarticulation. *UCLA WPP* 62: 1–13.

Kelso, J. A. S. and B. Tuller. 1985. Intrinsic time in speech production: theory, methodology, and preliminary observations. *Haskins Laboratories. Status Report on Speech Research* SR-81: 23–39.

Kelso, J. A. S., E. Vatikiotis-Bateson, E. L. Saltzman, and B. Kay. 1985. A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling. *JASA* 77: 266–280.

Kent, R. D., and K. Moll. 1975. Articulatory timing in selected consonant sequences. *Brain and Language* 2: 304–323.

Kohler, K. 1976. Die Instabilität wortfinaler Alveolarplosive im Deutschen: eine elektropalatographische Untersuchung [The instability of word-final alveolar plosives in German: An electropalatographic investigation.] *Phonetica* 33: 1–30.

Kuehn, D. P. and K. Moll. 1976. A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics* 4: 303–320.

Labov, W., P. Cohen, C. Robins, and J. Lewis. 1968. A study of the non-standard English of Negro and Puerto Rican speakers in New York City. Cooperative Research Report 3288. Vols. I and II. New York: Columbia University. (Reprinted by U.S. Regional Survey, 204 North 35th St., Philadelphia, PA 19104)

Ladefoged, P., and I. Maddieson, I. 1986. Some of the sounds of the world's languages. *UCLA Working Papers in Phonetics* 64: 1–137.

Liberman, A., F. Cooper, D. Shankweiler, and M. Studdert-Kennedy. 1967. Perception of the speech code. *Psychological Review* 74: 431–436.

Lindblom, B. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35: 1773–1781.

1983. Economy of speech gestures. In P. F. MacNeilage (ed.) *The Production of Speech.* New York: Springer-Verlag, 217–245.

Lisker, L., A. S. Abramson, F. S. Cooper, and M. H. Schvey. 1969. Transillumination of the larynx in running speech. *Journal of the Acoustical Society of America* 45: 1544–1546.

Macchi, M. J. 1985. Segmental and suprasegmental features and lip and jaw articulators. Ph.D. dissertation, New York University.

Mattingly, I. G. 1981. Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, and J. Anderson (eds.) *The Cognitive Representation of Speech.* Amsterdam: North-Holland. 415–420.

McCarthy, J. J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12: 373–418.

Miller, J. E., and O. Fujimura. 1982. Graphic displays of combined presentations of acoustic and articulatory information. *The Bell System Technical Journal* 61: 799–810.

Munhall, K. G., D. J. Ostry, and A. Parush. 1985. Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance* 11: 457-474.

Neu, H. 1980. Ranking of constraints on /t,d/ deletion in American English: A statistical analysis. In W. Labov (ed.) *Locating Language in Time and Space*. New York: Academic Press, 37-54.

Ohala, J. J. 1974. Experimental historical phonology. In J. M. Anderson and C. Jones (eds.) *Historical Linguistics* Vol. 2. Amsterdam: North-Holland. 353–389.

Ohman, S. E. G. 1966. Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 39: 151--168.

1967. Numerical model of coarticulation. *Journal of the Acoustical Society of America* 41: 310–320.

Oshika, B., V. Zue, R. Weeks, H. Neu, and J. Aurbach. 1975. The role of phonological rules in speech understanding research. *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-23*, 104-112.

Ostry, D. J. and K. Munhall. 1985. Control of rate and duration of speech movements. *Journal of the Acoustical Society of America* 77: 640–648.

Rubin, P., T. Baer, and P. Mermelstein 1981. An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America* 70: 321–328.

Saltzman, E. 1986. Task dynamic coordination of the speech articulators: a preliminary model. In H. Heuer and C. Fromm (eds.) *Generation and Modulation of Action Patterns*. Experimental Brain Research Series 15. New York: Springer-Verlag, 129-144.

Saltzman, E. and J. A. S. Kelso. 1987. Skilled actions: a task dynamic approach. *Psychological Review* 94: 84-106.

Selkirk, E. O. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.

Shockey, L. 1984. Phonetic and phonological properties of connected speech. *Ohio State Working Papers in Linguistics*. 17: iv 143.

Thrainsson, H. 1978. On the phonology of Icelandic preaspiration. *Nordic Journal of Linguistics*. 1(1): 3-54.

Vennemann, T., and P. Ladefoged. 1973. Phonetic features and phonological features. *Lingua* 32: 61–74.

Zwicky, A. 1972. Note on a phonological hierarchy in English. In R. Stockwell and R. Macauly (eds.) *Linguistic Change and Generative Theory*. Bloomington: Indiana University Press.

Browman, C. P., L. Goldstein, E. Saltzman, and C. Smith. 1986. GEST: A computational model for speech production using dynamically defined articulatory gestures. *Journal of the Acoustical Society of America* 80: S97. (Abstract).

Brown, G. 1977. *Listening to Spoken English.* London: Longman.

Catford, J. C. 1977. *Fundamental Problems in Phonetics.* Bloomington, IN: Indiana University Press.

Clements, G. N. 1980. *Vowel Harmony in Nonlinear Generative Phonology: An Autosegmental Model.* Bloomington, IN: Indiana University Linguistics Club.

1985. The geometry of phonological features. *Phonology Yearbook* 2: 225–252.

Clements, G. N. and S. J. Keyser. 1983. *CV Phonology: A Generative Theory of the Syllable.* Cambridge, MA: MIT Press.

Dalby, J. M. 1984. Phonetic structure of fast speech in American English. Ph.D. dissertation, Indiana University.

Dressler, W., and R. Wodak. 1982. Sociophonological methods in the study of sociolinguistic variation in Viennese German. *Language and Society* 11: 339–370.

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

1981. A relationship between coarticulation and compensatory shortening. *Phonetica* 38: 35–50.

1983. Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet. *Journal of Experimental Psychology: General* 112: 386–412.

Fujimura, O. 1981. Temporal organization of articulatory movements as a multidimensional phrasal structure. *Phonetica* 38: 66–83.

Fujimura, O., S. Kiritani, and H. Ishida. 1973. Computer controlled radiography for observation of movements of articulatory and other human organs. *Computers in Biology and Medicine* 3: 371–384.

Fujimura, O. and J. Lovins. 1978. Syllables as concatenative phonetic units. In A. Bell and J. B. Hooper (eds.) *Syllables and Segments.* Amsterdam: North Holland, 107–120.

Gay, T. 1977. Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America* 62: 183–193.

1978. Articulatory units: segments or syllables? In A. Bell and J. B. Hopper (eds.) *Syllables and Segments.* Amsterdam: North Holland, 121–131.

1981. Mechanisms in the control of speech rate. *Phonetica* 38: 148–158.

Gimson, A. C. 1962. *An Introduction to the Pronunciation of English.* London: Edward Arnold.

Goldsmith, J. A. 1976. *Autosegmental Phonology.* Bloomington, IN: Indiana University Linguistics Club.

Goldstein, L. and C. P. Browman. 1986. Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics* 14: 339–342.

Guy, G. R. 1975. Use and application of the Cedergren–Sankoff variable rule program. In R. Fasold and R. Shuy (eds.) *Analyzing Variation in Language.* Washington, DC: Georgetown University Press, 56–69.

1980. Variation in the group and the individual: the case of final stop deletion. In W. Labov (ed.) *Locating Language in Time and Space.* New York: Academic Press, 1–36.

Halle, M. 1982. On distinctive features and their articulatory implementation. *Natural Language and Linguistic Theory* 1: 91–105.

Hardcastle, W. J., and P. J. Roach. 1979. An instrumental investigation of coarticulation in stop consonant sequences. In P. Hollien and H. Hollien (eds.) *Current Issues in The Phonetic Sciences.* Amsterdam: John Benjamins AG, 531–540.

Hayes, B. 1981. *A Metrical Theory of Stress Rules.* Bloomington, IN: Indiana University Linguistics Club.

Hirose, H., S. Niimi, K. Honda, and M. Sawashima. 1985. The relationship between glottal opening and transglottal pressure difference during consonant production. *Annual Bulletin of RILP* 19, 55–64.

Kahn, D. 1976. *Syllable-based Generalizations in English Phonology.* Bloomington, IN: Indiana University Linguistics Club.

Keating, P. A. 1985. CV phonology, experimental phonetics, and coarticulation. *UCLA WPP* 62: 1–13.

Kelso, J. A. S. and B. Tuller. 1985. Intrinsic time in speech production; theory, methodology, and preliminary observations. *Haskins Laboratories. Status Report on Speech Research* SR-81: 23–39.

Kelso, J. A. S., E. Vatikiotis-Bateson, E. L. Saltzman, and B. Kay. 1985. A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling. *JASA* 77: 266–280.

Kent, R. D., and K. Moll. 1975. Articulatory timing in selected consonant sequences. *Brain and Language* 2: 304–323.

Kohler, K. 1976. Die Instabilität wortfinaler Alveolarplosive im Deutschen: eine elektropalatographische Untersuchung [The instability of word-final alveolar plosives in German: An electropalatographic investigation.] *Phonetica* 33: 1–30.

Kuehn, D. P. and K. Moll. 1976. A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics* 4: 303–320.

Labov, W., P. Cohen, C. Robins, and J. Lewis. 1968. A study of the non-standard English of Negro and Puerto Rican speakers in New York City. Cooperative Research Report 3288. Vols. I and II. New York: Columbia University. (Reprinted by U.S. Regional Survey, 204 North 35th St., Philadelphia, PA 19104)

Ladefoged, P., and I. Maddieson, I. 1986. Some of the sounds of the world's languages. *UCLA Working Papers in Phonetics* 64: 1–137.

Liberman, A., F. Cooper, D. Shankweiler, and M. Studdert-Kennedy. 1967. Perception of the speech code. *Psychological Review* 74: 431–436.

Lindblom, B. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35: 1773–1781.

1983. Economy of speech gestures. In P. F. MacNeilage (ed.) *The Production of Speech.* New York: Springer-Verlag, 217–245.

Lisker, L., A. S. Abramson, F. S. Cooper, and M. H. Schvey. 1969. Transillumination of the larynx in running speech. *Journal of the Acoustical Society of America* 45: 1544–1546.

Macchi, M. J. 1985. Segmental and suprasegmental features and lip and jaw articulators. Ph.D. dissertation, New York University.

Mattingly, I. G. 1981. Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, and J. Anderson (eds.) *The Cognitive Representation of Speech.* Amsterdam: North-Holland. 415–420.

McCarthy, J. J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12: 373–418.

Miller, J. E., and O. Fujimura. 1982. Graphic displays of combined presentations of acoustic and articulatory information. *The Bell System Technical Journal* 61: 799–810.

Munhall, K. G., D. J. Ostry, and A. Parush. 1985. Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance* 11: 457-474.

Neu, H. 1980. Ranking of constraints on /t,d/ deletion in American English: A statistical analysis. In W. Labov (ed.) *Locating Language in Time and Space*. New York: Academic Press, 37-54.

Ohala, J. J. 1974. Experimental historical phonology. In J. M. Anderson and C. Jones (eds.) *Historical Linguistics* Vol. 2. Amsterdam: North-Holland. 353-389.

Ohman, S. E. G. 1966. Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 39: 151-168.

1967. Numerical model of coarticulation. *Journal of the Acoustical Society of America* 41: 310-320.

Oshika, B., V. Zue, R. Weeks, H. Neu, and J. Aurbach. 1975. The role of phonological rules in speech understanding research. *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-23*, 104-112.

Ostry, D. J. and K. Munhall. 1985. Control of rate and duration of speech movements. *Journal of the Acoustical Society of America* 77: 640-648.

Rubin, P., T. Baer, and P. Mermelstein 1981. An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America* 70: 321-328.

Saltzman, E. 1986. Task dynamic coordination of the speech articulators: a preliminary model. In H. Heuer and C. Fromm (eds.) *Generation and Modulation of Action Patterns*. Experimental Brain Research Series 15. New York: Springer-Verlag, 129-144.

Saltzman, E. and J. A. S. Kelso. 1987. Skilled actions: a task dynamic approach. *Psychological Review* 94: 84-106.

Selkirk, E. O. 1984. *Phonology and Syntax: The Relation between Sound and Structure.* Cambridge, MA: MIT Press.

Shockey, L. 1984. Phonetic and phonological properties of connected speech. *Ohio State Working Papers in Linguistics.* 17: iv 143.

Thrainsson, H. 1978. On the phonology of Icelandic preaspiration. *Nordic Journal of Linguistics.* 1(1): 3 54.

Vennemann, T., and P. Ladefoged. 1973. Phonetic features and phonological features. *Lingua* 32: 61-74.

Zwicky, A. 1972. Note on a phonological hierarchy in English. In R. Stockwell and R. Macauly (eds.) *Linguistic Change and Generative Theory*. Bloomington: Indiana University Press.