

# Articulatory gestures as phonological units\*

702

Catherine P. Browman  
Haskins Laboratories

Louis Goldstein  
Yale University

---

## 1 A gestural phonology

Over the past few years, we have been investigating a particular hypothesis about the nature of the basic 'atoms' out of which phonological structures are formed. The atoms are assumed to be primitive actions of the vocal tract articulators that we call 'gestures'. Informally, a gesture is identified with the formation (and release) of a characteristic constriction within one of the relatively independent articulatory subsystems of the vocal tract (i.e. oral, laryngeal, velic). Within the oral subsystem, constrictions can be formed by the action of one of three relatively independent sets of articulators: the lips, the tongue/blade and the tongue body. As actions, gestures have some intrinsic time associated with them – they are characterisations of movements through space and over time (see Fowler *et al.* 1980).

Within the view we are developing, phonological structures are stable 'constellations' (or 'molecules', to avoid mixing metaphors) assembled out of these gestural atoms. In this paper, we examine some of the evidence for, and some of the consequences of, the assumption that gestures are the basic atoms of phonological structures. First, we attempt to establish that gestures are pre-linguistic discrete units of action that are inherent in the maturation of a developing child and that therefore can be harnessed as elements of a phonological system in the course of development (§ 1.1). We then give a more detailed, formal characterisation of gestures as phonological units within the context of a computational model (§ 1.2), and show that a number of phonological regularities can be captured by representing constellations of gestures (each having inherent duration) using gestural scores (§ 1.3). Finally, we show how the proposed gestural structures relate to proposals of feature geometry (§§ 2-3).

### 1.1 Gestures as pre-linguistic primitives

Gestures are units of action that can be identified by observing the coordinated movements of vocal tract articulators. That is, repeated observations of the production of a given utterance will reveal a characteristic pattern of constrictions being formed and released. The fact that these patterns of (discrete) gestures are similar in structure to the non-linear phonological representations being currently postulated (e.g. Clements 1985; Sagey 1986; Hayes 1986), together with some of the evidence presented in Browman & Goldstein (1986, in press), leads us to make the strong hypothesis that gestures themselves constitute basic phonological units. This hypothesis has the attractive feature that the basic units of phonology can be identified directly with cohesive patterns of movement within the vocal tract. Thus, the phonological system is built out of inherently discrete units of action. This state of affairs would be particularly useful for a child learning to speak. If we assume that discrete gestures (like those that will eventually function as phonological units) emerge in the child's behavioural repertoire *in advance* of any specifically linguistic development, then it is possible to view phonological development as harnessing these action units to be the basic units of phonological structures.

The idea that pre-linguistic gestures are employed in the service of producing early words has been proposed and supported by a number of writers, for example, Fry (1966), Locke (1983), Studdert-Kennedy (1987) and Vihman (in press), where what we identify as 'gestures' are referred to as 'articulatory routines' or the like. The view we are proposing extends this approach by hypothesising that these pre-linguistic gestures actually become the units of contrast. Additional phonological developments involve differentiating and tuning the gestures, and developing patterns of intergestural coordination that correspond to larger phonological structures.

The evidence that gestures are pre-linguistic units of action can be seen in the babbling behaviour of young infants. The descriptions of infant babbling (ages 6–12 months) suggest a predominance of what are transcribed as simple CV syllables (Locke 1986; Oller & Eilers 1982). The 'consonantal' part of these productions can be analysed as simple, gross, constriction manoeuvres of the independent vocal tract subsystems and (within the oral subsystem) the separate oral articulator sets. For example, based on frequency counts obtained from a number of studies, Locke (1983) finds that the consonants in (1) constitute the 'core' babbling inventory: these 12 'consonants' account for about 95% of the babbles of English babies. Similar frequencies obtain in other language environments:

(1) h b d g p t k m n j w s

These transcriptions are not meant to be either systematic phonological representations (the child doesn't have a phonology yet) or narrow

phonetic transcriptions (the child cannot be producing the detailed units of its 'target' language, because, as noted below, there do not seem to be systematic differences in the babbles produced by infants in different language environments). Others have noted the problems inherent in using a transcription that assumes a system of units and relations to describe a behaviour that lacks such a system (e.g. Kent & Murray 1982; Oller 1986; Koopmans-van Beinum & van der Stelt 1986; Studdert-Kennedy 1987). As Studdert-Kennedy (1987) argues, it seems likely that these transcriptions reflect the production by the infant of simple vocal constriction gestures, of the kind that evolve into mature phonological structures (which is why adults can transcribe them using their phonological categories). Thus, /h/ can be interpreted as a laryngeal widening gesture and /b d g/ as 'gross' constriction gestures of the three independent oral articulator sets (lips, tongue tip and tongue body). /p t k/ combine the oral constriction gestures with the laryngeal manoeuvre, and /m n/ combine oral constrictions with velic lowering. These combinations do not necessarily indicate an ability on the part of the infant to coordinate the gestures. Rather, any accidental temporal coincidence of two such gestures would be perceived by the listener as the segments in question.

The analysis outlined above suggests that babbling involves the emergence, in the infant, of simple constriction gestures of independent parts of the vocal tract. As argued by Locke (1986), the pattern of emergence of these actions can be viewed as a function of anatomical and neurophysiological developments, rather than the beginning of language acquisition *per se*. This can be seen, first of all, in the fact that the babbling inventory and its developmental sequence have not been shown to vary as a function of the particular language environment in which the child finds itself (although individual infants may vary considerably from one another in the relative frequencies of particular gestures – Studdert-Kennedy 1987; Vihman in press). In fact, in the large number of studies reviewed by Locke (1983), there appear to be no detectable differences (either instrumentally or perceptually) in the 'consonantal' babbling of infants reared in different language environments. (More recent studies have found some language environment effect on the overall long term spectrum of vocalic utterances (de Boysson-Bardies *et al.* 1986), and on prosody (de Boysson-Bardies *et al.* 1984). Other subtle effects may be uncovered with improvement of analytic techniques.)

Secondly, Locke (1983) notes that the developmental changes in frequency of particular babbled consonants can likely be explained by anatomical developments. Most of the consonants produced by very young infants (less than six months) involve tongue body constrictions, usually transcribed as velars. Some time shortly after the beginning of repetitive canonical babbling (usually in the seventh month), tongue tip and lip constrictions begin to outnumber tongue body constrictions, with tongue tip constrictions eventually dominating. Even deaf infants show a progression that is qualitatively similar, at least in early stages, although their babbling can be distinguished from that of hearing infants on a

number of acoustic measures (Oller & Eilers 1988). Locke suggests an explanation in terms of vocal tract maturation. At birth, the infant's larynx is high, and the tongue virtually fills the oral cavity (Lieberman 1984). This would account for the early dominance of tongue body constrictions. After the larynx drops, tongue tip and lip constrictions – without simultaneous tongue constrictions – are more readily formed. In particular, the closing action of the mandible will then contribute to constrictions at the front of the mouth.

Finally, Locke (1986) notes that the timing of the development of repetitive 'syllabic' babbling coincides with the emergence of repetitive motor behaviours generally. He cites Thelen's (1981) observation of 47 different rhythmic activities that have their peak frequency at 6–7 months. Locke concludes (1986: 145) that 'it thus appears that the syllabic patterning of babble – like the phonetic patterning of its segments – is determined mostly by nonlinguistic developments of vocal tract anatomy and neurophysiology'.

The pre-linguistic vocal gestures become linguistically significant when the child begins to produce its first few words. The child seems to notice the similarity of its babbled patterns to the speech s/he hears (Locke 1986), and begins to produce 'words' (with apparent referential meaning) using the available set of vocal gestures. It is possible to establish that there is a definite relationship between the (non-linguistic) gestures of babbling and the gestures employed in early words by examining individual differences among children. Vihman *et al.* (1985) and Vihman (in press) find that the particular consonants that were produced with high frequency in the babbling of a given child also appear with high frequency in that child's early word productions. Thus, the child is recruiting its well-practised action units for a new task. In fact in some early cases (e.g. 'baby' words like *mama*, etc.), 'recruiting' is too active a notion. Rather, parents are helping the child establish a referential function with sequences that already exist as part of the babbling repertoire (Locke 1986).

Once the child begins producing words (complex units that have to be distinguished one from another), using the available gestures as building blocks, phonology has begun to form. If we compare the child's early productions (using the small set of pre-linguistic gestures) to the gestural structure of the adult forms, it is clear that there are (at least) two important developments that are required to get from one to the other: (1) differentiation and tuning of individual gestures and (2) coordination of the individual gestures belonging to a given word. Let us examine these in turn.

*Differentiation and tuning.* While the repertoire of gestures inherent in the consonants of (1) above employs all of the relatively independent articulator sets, the babbled gestures involve just a single (presumably gross) movement. For example, some kind of closure is involved for oral constriction gestures. In general, however, languages employ gestures

produced with a given articulator set that contrast in the degree of constriction. That is, not only are closure gestures produced but also fricative and wider (approximant) gestures. In addition, the exact location of the constriction formed by a given articulator set may contrast, e.g. in the gestures for /θ/, /s/ and /ʃ/. Thus, a single pre-linguistic constriction gesture must eventually differentiate into a variety of potentially contrastive gestures, tuned with different values of constriction location and degree. Although the location and degree of the constriction formed by a given articulator set are, in principle, physical continua, the differentiated gestures can be categorically distinct. The partitioning of these continua into discrete categories is likely aided by quantal (i.e. non-linear) articulatory-auditory relations of the kind proposed by Stevens (1972, 1989). In addition, Lindblom (1986) has shown how the pressures to keep contrasting words perceptually distinct can lead to discrete clustering along some articulatory/acoustic continua. Even so, the process of differentiation may be lengthy. Nitttrouer *et al.* (1989) present data on fricative production in American English children. They find that differentiation between /s/ and /ʃ/ is increasing in children from ages three to seven, and hasn't yet reached the level shown by adults. In addition, tuning may occur even where differentiation is unnecessary. That is, even if a language has only a single tongue tip closure gesture, its constriction location may be tuned to a particular language-specific value. For example, English stops have an alveolar constriction location, while French stops are more typically dental.

*Coordination.* The various gestures that constitute the atoms of a given word must be organised appropriately. There is some evidence that a child can know what all the relevant gestures are for some particular word, and can produce them all, without either knowing or being able to produce the appropriate organisation. Studdert-Kennedy (1987) presents an example of this kind from Ferguson & Farwell (1975). They list ten attempts by a 15-month old girl to say the word *pen* in a half-hour session, as shown in (2):

- (2) [mã<sup>a</sup>, ʏλ, de<sup>dn</sup>, hɪn, m<sup>b</sup>õ, p<sup>h</sup>ɪn, t<sup>h</sup>ɪt<sup>h</sup>ɪt<sup>h</sup>ɪ, ba<sup>h</sup>, d<sup>h</sup>au<sup>n</sup>, buã]

While these attempts appear radically different, they can be analysed, for the most part, as the set of gestures that constitute *pen* misarranged in various ways: glottal opening, bilabial closure, tongue body lowering, alveolar closure and velum lowering. Eventually, the 'right' organisation is hit upon by the child. The search is presumably aided by the fact that the coordinated structure embodied in the target language is one of a relatively small number of dynamically stable patterns (see also Boucher 1988). The formation of such stable patterns may ultimately be illuminated by research into stable modes in coordinated human action in general (e.g. Haken *et al.* 1985; Schmidt *et al.* 1987) being conducted within the broad context of the non-linear dynamics relevant to problems of pattern formation in physics and biology (e.g. Thompson & Stewart

1976; Glass & Mackey 1988). In addition, aspects of coordination may emerge as the result of keeping a growing number of words perceptually distinct using limited articulatory resources (Lindblom *et al.* 1983).

If phonological structures are assumed to be organised patterns of gestural units, a distinct methodological bonus obtains: the vocal behaviour of infants, even pre-linguistic behaviour, can be described using the same primitives (discrete units of vocal action) that are used to describe the fully elaborated phonological system of adults. This allows the growth of phonological form to be precisely monitored, by observing the development of the primitive gestural structures of infants into the elaborated structures of adults (Best & Wilkenfield 1988 provide an example of this). In addition, some of the thorny problems associated with the transcription of babbling can be obviated. Discrete units of action are present in the infant, and can be so represented, even if adult-like phonological structures have not yet developed. A similar advantage applies to describing various kinds of 'disordered' speech (e.g. Kent 1983; Marshall *et al.* 1988), which may lack the organisation shown in 'normal' adult phonology, making conventional phonological/phonetic transcriptions inappropriate, but which may, nevertheless, be composed of gestural primitives. Of course, all this assumes that it is possible to give an account of adult phonology using gestures as the basic units – it is to that account that we now turn.

## 1.2 The nature of phonological gestures

In conjunction with our colleagues Elliot Saltzman and Philip Rubin at Haskins Laboratories, we are developing a computational model that produces speech beginning with a representation of phonological structures in terms of gestures (Browman *et al.* 1984; Browman *et al.* 1986; Saltzman *et al.* 1987; Browman & Goldstein 1987), where a GESTURE is an abstract characterisation of coordinated task-directed movements of articulators within the vocal tract. Each gesture is precisely defined in terms of the parameters of a set of equations for a 'task-dynamic' model (Saltzman & Kelso 1987; Saltzman 1986). When the control regime for a given gesture is active, the equations regulate the coordination of the model's articulators in such a way that the gestural 'task' (the formation of a specified constriction) is reached as the articulator motions unfold over time. Acoustic output is obtained from these articulator motions by means of an articulatory synthesiser (Rubin *et al.* 1981). The gestures for a given utterance are themselves organised into a larger coordinated structure, or constellation, that is represented in a GESTURAL SCORE (discussed in §1.3). The score specifies the sets of values of the dynamic parameters for each gesture, and the temporal intervals during which each gesture is active. While we use analyses of articulatory movement data to determine the parameter values for the gestures and gestural scores, there is nevertheless a striking convergence between the structures we derive through these analyses and phonological structures currently being

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

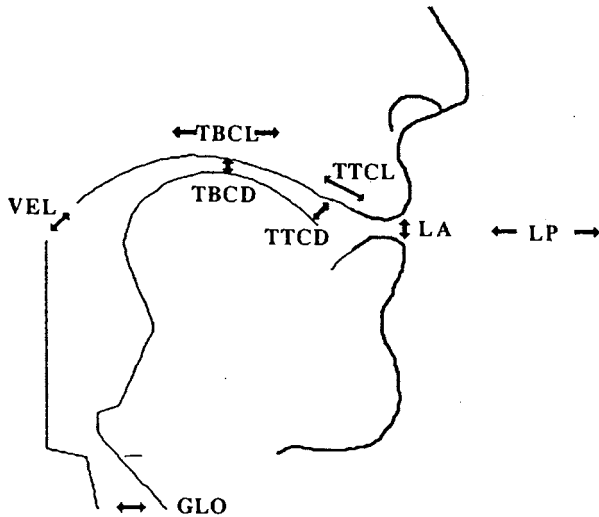


Figure 1

Tract variables and contributing articulators of computational model.

proposed in other frameworks (e.g. Plotkin 1976; Ewen 1982; Lass 1984; Clements 1985; Sagey 1986; Anderson & Ewen 1987; McCarthy 1989).

Within task dynamics, the goal for a given gesture is specified in terms of independent task dimensions, called VOCAL TRACT VARIABLES. Each tract variable is associated with the specific sets of articulators whose movements determine the value of that variable. For example, one such tract variable is Lip Aperture (LA), corresponding to the vertical distance between the two lips. Three articulators can contribute to changing LA: the jaw, vertical displacement of the lower lip with respect to the jaw, and vertical displacement of the upper lip. The current set of tract variables in the computational model, and their associated articulators, can be seen in Fig. 1. Within the task dynamic model, the control regime for a given

gesture coordinates the ongoing movements of these articulators in a flexible, but task-specific manner, according to the demands of other concurrently active gestures. The motion associated with each of a gesture's tract variables is specified in terms of an equation for a second-order dynamical system.<sup>1</sup> The equilibrium position parameter of the equation  $x_0$  specifies the tract variable target that will be achieved, and the stiffness parameter  $k$  specifies (roughly) the time required to get to target. These parameters are tuned differently for different gestures. In addition, their values can be modified by stress.

Gestures are currently specified in terms of one or two tract variables. Velic gestures involve a single tract variable of aperture size, as do glottal gestures. Oral gestures involve pairs of tract variables that specify the constriction degree (LA, TTCD and TBCD) and constriction location (LP, TTCL and TBCL). For simplicity, we will refer to the sets of articulators involved in oral gestures using the name of the end-effector, that is, the name of the single articulator at the end of the chain of articulators forming the particular constriction: the LIPS for LA and LP, the tongue tip (TT) for gestures involving TTCD and TTCL, and the tongue body (TB) for gestures involving TBCD and TBCL. As noted above, each tract variable is modelled using a separate dynamical equation; however, at present the paired tract variables use identical stiffness and are activated and de-activated simultaneously. The damping parameter  $b$  for oral gestures is always set for critical damping – the gestures approach their targets, but do not overshoot them, or 'ring'. Thus, a given oral gesture is specified by the values of three parameters: target values for each of a pair of tract variables, and a stiffness value (used for both equations).

This set of tract variables is not yet complete, of course. Other oral tract variables that need to be implemented include an independent tongue root variable (Ladefoged & Halle 1988), and (as discussed in §2.1) variables for controlling the shape of TT and TB constrictions as seen in the third dimension. Additional laryngeal variables are required to allow for pitch control and for vertical movement of the larynx, required, for example, for ejectives and implosives.

The representations of gestures employing distinct sets of tract variables are categorically distinct within the system outlined here. That is, they are defined using different variables that correspond to different sets of articulators. They provide, therefore, an inherent basis for contrast among gestures (Browman & Goldstein in press). However, for contrasting gestures that employ the same tract variables, the difference between the gestures is in the tuned values of the continuous dynamic parameters (for oral gestures: constriction degree, location and stiffness). That is, unlike the articulator sets being used, the dynamic parameters do not inherently define categorically distinct classes. Nonetheless, we assume that there are stable ranges of parameter values that tend to contrast with one another repeatedly in languages (Ladefoged & Maddieson 1986; Vatikiotis-Bateson 1988). The discrete values might be derived using a combination



of articulatory and auditory constraints applied across the entire lexicon of a language, as proposed by Lindblom *et al.* (1983). In addition, part of the basis for the different ranges might reside in the non-linear relation between the parameter values and their acoustic consequences, as in Stevens' quantal theory (Stevens 1972, 1989).

In order to represent the contrastive ranges of gestural parameter values in a discrete fashion, we employ a set of GESTURAL DESCRIPTORS. These descriptors serve as pointers to the particular articulator set involved in a given gesture, and to the numerical values of the dynamical parameters characterising the gestures. In addition, they can act as classificatory and distinctive features for the purposes of lexical and phonological structure. Every gesture can be specified by a distinct descriptor structure. This functional organisation can be formally represented as in (3), which relates the parameters of the dynamical equations to the symbolic descriptors. Contrasting gestures will differ in at least one of these descriptors:

- (3) Gesture = articulator set (constriction degree, constriction location, constriction shape, stiffness)

CONSTRUCTION DEGREE is always present, and refers to the  $x_0$  value for the constriction degree tract variables (LA, TTCD, TBCD, VEL, or GLO).

CONSTRUCTION LOCATION is relevant only for oral gestures, and refers to the  $x_0$  value for the constriction location tract variables (LP, TTCL, or TBCL).

CONSTRUCTION SHAPE is relevant only for oral gestures, and refers to the  $x_0$  value of constriction shape tract variables. It is not currently implemented.

STIFFNESS refers to the  $k$  value of the tract variables.

Fig. 2 displays the inventory of articulator sets and associated parameters that we posit are required for a general gestural phonology. The parameters correspond to the particular tract variables of the model shown below them. Those parameters with asterisks are not currently implemented.

For the present, we list without comment the possible descriptor values for the constriction degree (CD) and constriction location (CL) dimensions in (4). In §2.1, we will discuss the gestural dimensions and these descriptors in detail, including a comparison with current proposals of featural geometry.

- (4) CD descriptors: closed critical narrow mid wide  
 CL descriptors: protruded labial dental  
 alveolar postalveolar palatal  
 velar uvular pharyngeal

In the phonology of dynamically defined articulatory gestures that we are developing, gestures are posited to be the atoms of phonological structure. It is important to note that such gestures are relatively abstract. That is, the physically continuous movement trajectories are analysed as

Gestures:		
<u>Articulator Set</u>		<u>Dimensions</u>
LIPS	(con degree, con location, LA LP	, stiffness)
TT	(con degree, con location, con shape*, TTCD TTCL	stiffness)
TB	(con degree, con location, con shape*, TBCD TBCL	stiffness)
TR*	(con degree*, con location*,	, stiffness*)
VEL	(con degree, VEL	, stiffness)
GLO	(con degree, con location*, GLO	, stiffness)

Figure 2

Inventory of articulator sets and associated parameters.

resulting from a set of discrete, concurrently active gestural control regimes. They are discrete in two senses: (1) the dynamic parameters of a gesture's control regime remain constant throughout the discrete interval of time during which the gesture is active, and (2) gestures in a language may differ from one another in discrete ways, as represented by different descriptor values. Thus, as argued in Browman & Goldstein (1986) and Browman & Goldstein (in press), the gestures for a given utterance, together with their temporal patterning, perform a dual function. They characterise the actual observed articulator movements (thus obviating the need for any additional implementation rules), and they also function as units of contrast (and more generally capture aspects of phonological patterning). As discussed in those papers, the gesture as a phonological unit differs both from the feature and from the segment (or

root node in current feature geometries). It is a larger unit than the feature, being effectively a unitary constriction action, parameterised jointly by a linked structure of features (descriptor values). Yet it is a smaller unit than the segment: several gestures linked together are necessary to form a unit at the segmental, or higher, levels.

### 1.3 Gestural scores: articulatory tiers and internal duration

In the preceding section, gestures were defined with reference to a dynamical system that shapes patterns of articulatory movements. Each gesture possesses, therefore, not only an inherent spatial aspect (i.e. a tract variable goal) but also an intrinsic temporal aspect (i.e. a gestural stiffness). Much of the power of the gestural approach follows from these basic facts about gestures (combined with their abstractness), since they allow gestures to OVERLAP in time as well as in articulator and/or tract variable space (see also Fujimura 1981a, b; Bell-Berti & Harris 1981; Fowler 1980, 1983). In this section we show how overlap among gestures is represented, and demonstrate that simple changes in the patterns of overlap between neighbouring gestural units can automatically produce a variety of superficially different types of phonetic and phonological variation.

Within the computational model described above, the pattern of organisation, or CONSTELLATION, of gestures corresponding to a given utterance is embodied in a set of phasing principles (see Kelso & Tuller 1987; Nittrouer *et al.* 1988) that specify the spatiotemporal coordination of the gestures (Browman & Goldstein 1986). The pattern of intergestural coordination that results from applying the phasing principles, along with the interval of active control for individual gestures, is displayed in a two-dimensional GESTURAL SCORE, with articulatory tiers on one dimension and temporal information on the other (Browman & Goldstein 1987; Browman *et al.* 1986). A gestural score for the word *palm* (pronounced [pam]) is displayed in Fig. 3a. As can be seen in the figure, the tiers in a gestural score, on the vertical axis, represent the sets of articulators (or the relevant subset thereof) employed by the gestures, while the horizontal dimension codes time.

The boxes in Fig. 3a correspond to individual gestures, labelled by their descriptor values for constriction degree and constriction location (where relevant). For example, the initial oral gesture is a bilabial closure, represented as a constriction of the LIPS, with a [closed] constriction degree, and a [labial] constriction location. The horizontal extent of each box represents the interval of time during which that particular gesture is active. During these activation intervals, which are determined in the computational model from the phasing principles and the inherent stiffnesses of each gesture, the particular set of dynamic parameter values that defines each gesture is actively contributing to shaping the movements of the model articulators. In Fig. 3b, curves are added that show the time-varying tract variable trajectories generated by the task dynamic model

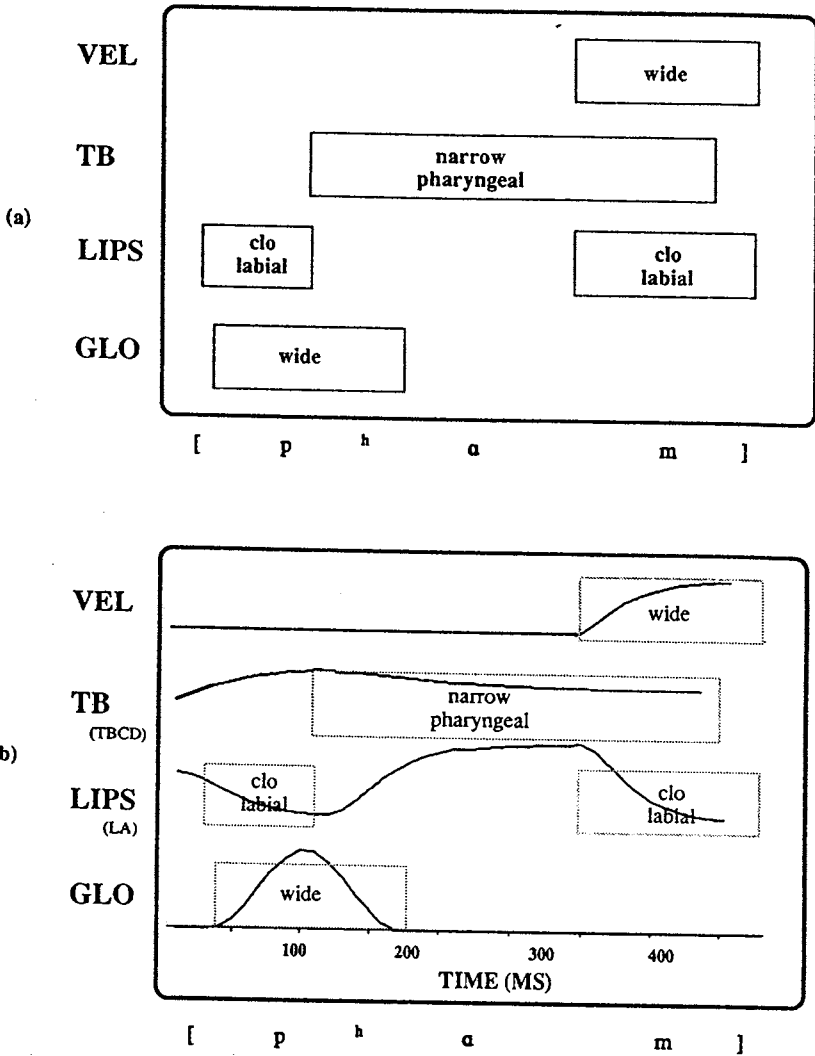


Figure 3

Gestural score for *palm* [pam] using box notation. (a) Activation intervals only; (b) model generated tract variable motions added.

according to the parameters indicated by the boxes. For example, during the activation interval of the initial labial closure, the curve representing LA (the vertical distance between the lips) decreases. As can be seen in these curves, the activation intervals directly capture something about the durations of the movements of the gestures.

Fig. 4 presents an alternative symbolic redisplay<sup>2</sup> of the gestural score. Instead of the 'box' notation of Fig. 3, a 'point' notation is employed that

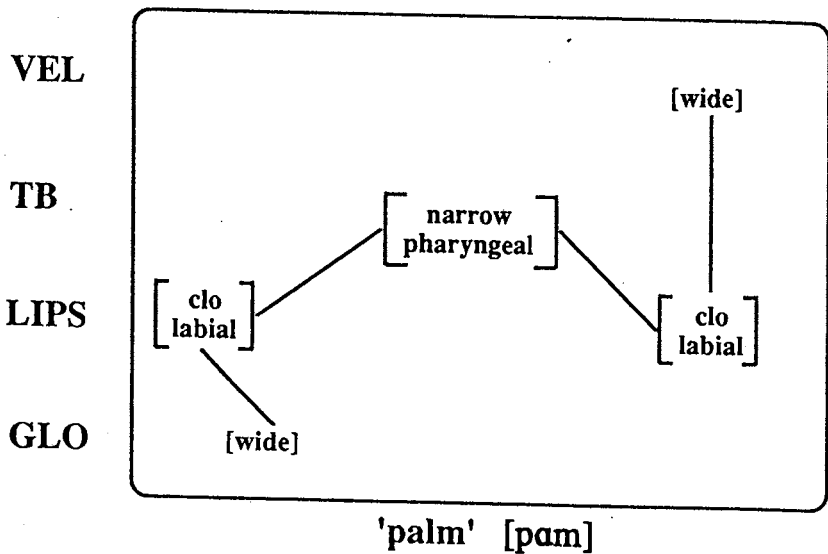


Figure 4

Gestural score for *palm* [pam] using point notation, with association lines added.

references only the gestural descriptors and relative timing of their 'targets'. The extent of activation of the gestures in the box notation is not indicated. In addition, association lines between the gestures have been added. These lines indicate which gestures are phased with respect to each other. Thus, the display is a shorthand representation of the phasing rules discussed in Browman & Goldstein (1987). The pair of gestures connected by a given association line are coordinated so that a specified phase of one gesture is synchronised with some phase of the other. For example, the peak opening of the GLO [wide] gesture (180 degrees) is synchronised with the release phase (290 degrees) of the LIPS [clo labial] gesture. Also important for the phase rules is the projection of oral constriction gestures onto separate Vowel and Consonant tiers, which are not shown here (Browman & Goldstein 1987, 1988; see also Keating 1985).

The use of point notation highlights the association lines, and, as we shall see in §2.2.2, is useful for the purpose of comparing gestural organisations with feature geometry representations in which individual units lack any extent in time. For the remainder of this section, however, we will be concerned with showing the extent of temporal overlap of gestures, and therefore will employ the box notation form of the gestural score.

The information represented in the gestural score serves to identify a particular lexical entry. The basic elements in such a lexical unit are the gestures, which, as we have already seen, can contrast with one another by means of differing descriptor values. In addition, gestural scores for

different lexical items can contrast in terms of the presence *vs.* absence of particular gestures (Browman & Goldstein 1986, in press; Goldstein & Browman 1986). Notice that one implication of taking gestures as basic units is that the resulting lexical representation is inherently under-specified, that is, it contains no specifications for irrelevant features. When a given articulator is not involved in any specified gesture, it is attracted to a 'neutral' position specific to that articulator (Saltzman *et al.* 1988; Saltzman & Munhall in press).

The fact that each gesture has an extent in time, and therefore can overlap with other gestures, has a variety of phonological and phonetic consequences. Overlap between invariantly specified gestures can automatically generate contextual variation of superficially different sorts: (1) acoustic non-invariance, such as the different formant transitions that result when an invariant consonant gesture overlaps different vowel gestures (Lieberman & Mattingly 1985); (2) allophonic variation, such as the nasalised vowel that is produced by overlap between a syllable-final velic opening gesture and the vowel gesture (Krakow 1989); and (3) various kinds of 'coarticulation', such as the context-dependent vocal tract shapes for reduced schwa vowels that result from overlap by the neighbouring full vowels (Fowler 1981; Browman & Goldstein 1989). Here, however, we will focus on the implications of directly representing overlap among phonological units for the phonological/phonetic alternations in fluent speech.

Browman & Goldstein (1987) have proposed that the wide variety of differences that have been observed between the canonical pronunciation of a word and its pronunciation in fluent contexts (e.g. Brown 1977; Shockey 1974) all result from two simple kinds of changes to the gestural score: (1) reduction in the magnitude of individual gestures (in both time and space) and (2) increase in overlap among gestures. That paper showed how these two very general processes might account for variations that have traditionally been described as segment deletions, insertions, assimilations and weakenings. The reason that increased overlap, in particular, can account for such different types of alternations is that the articulatory and acoustic consequences of increased overlap will vary depending on the nature of the overlapping gestures. We will illustrate this using some of the examples of deletion and assimilation presented in Browman & Goldstein (1987), and then compare the gestural account with the treatment of such examples in non-linear phonological theories that do not directly represent overlap among phonological units.

Let us examine what happens when a gestural score is varied by increasing the overlap between two oral constriction gestures, for example from no overlap to complete synchrony. This 'sliding' will produce different consequences in the articulatory and acoustic output of the model, depending on whether the gestures are on the same or different articulatory tiers, i.e. whether they employ the same or different tract variables. If the gestures are on different articulatory tiers, as in the case of a LIPS closure and a tongue tip (TT) closure, then the resulting tract variable motion for each gesture will be unaffected by the other concurrent

gesture. Their tract variable goals will be met, regardless of the amount of overlap. However, with sufficient overlap, one gesture may completely obscure the other acoustically, rendering it inaudible. We refer to this as gestural 'hiding'. In contrast, when two gestures are on the same articulatory tier, for example, the tongue tip constriction gestures associated with /t/ and /θ/, they cannot overlap without perturbing each other's tract variable motions. The two gestures are in competition – they are attempting to do different tasks with the identical articulatory structures. In this case, the dynamical parameters for the two overlapping gestural control regimes are 'blended' (Saltzman *et al.* 1988; Saltzman & Munhall in press).

Browman & Goldstein (1987) presented examples of articulations in fluent speech that showed the hiding and blending behaviour predicted by the model. Examples of hiding are transcribed in (5):

- (5) a. /pə<sup>r</sup>fəkt 'mɛmə<sup>r</sup>ri/ → [pə<sup>r</sup>fək'mɛmə<sup>r</sup>ri]  
 b. /sɛvən 'plʌs/ → [sɛvɪn'plʌs]

In (5a), careful listening to a speaker's production of *perfect memory* produced in a fluent sentence context failed to reveal any audible /t/ at the end of *perfect*, although the /t/ was audible when the two words were produced as separate phrases in a word list. This deletion of the final /t/ is an example of a general (variable) process in English that deletes final /t/ or /d/ in clusters, particularly before initial obstruents (Guy 1980). However, the articulatory data for the speaker examined in Browman & Goldstein (1987) showed that nothing was actually deleted from a gestural viewpoint. The alveolar closure gesture at the end of 'perfect' was produced in the fluent context, with much the same magnitude as when the two words were produced in isolation, but it was completely overlapped by the constrictions of the preceding velar closure and the following labial closure. Thus, the alveolar closure gesture was acoustically hidden. This increase in overlap is represented in Fig. 5, which shows the (partial) gestural scores posited for the two versions of this utterance, based on the observed articulatory movements. The gestures for the first syllable of *memory* (shown as shaded boxes) are well separated from the gestures for the last syllable of *perfect* (shown as unshaded boxes) in the word list version in Fig. 5a, but they slide earlier in time in the fluent version in Fig. 5b, producing substantial overlap among three closure gestures. Note that these three overlapping gestures are all on separate tiers.

In (5b), *seven plus* shows an apparent assimilation, rather than deletion, and was produced when the phrase was produced at a fast rate. Assimilation of final alveolar stops and nasals to a following labial (or velar) stop is a common connected speech process in English (Gimson 1962; Brown 1977). Here again, however, the articulatory data in Browman & Goldstein (1987) showed that the actual change was 'hiding' due to increased overlap: the alveolar closure gesture at the end of *seven* was still produced by the speaker in the 'assimilated' version, but it was hidden by

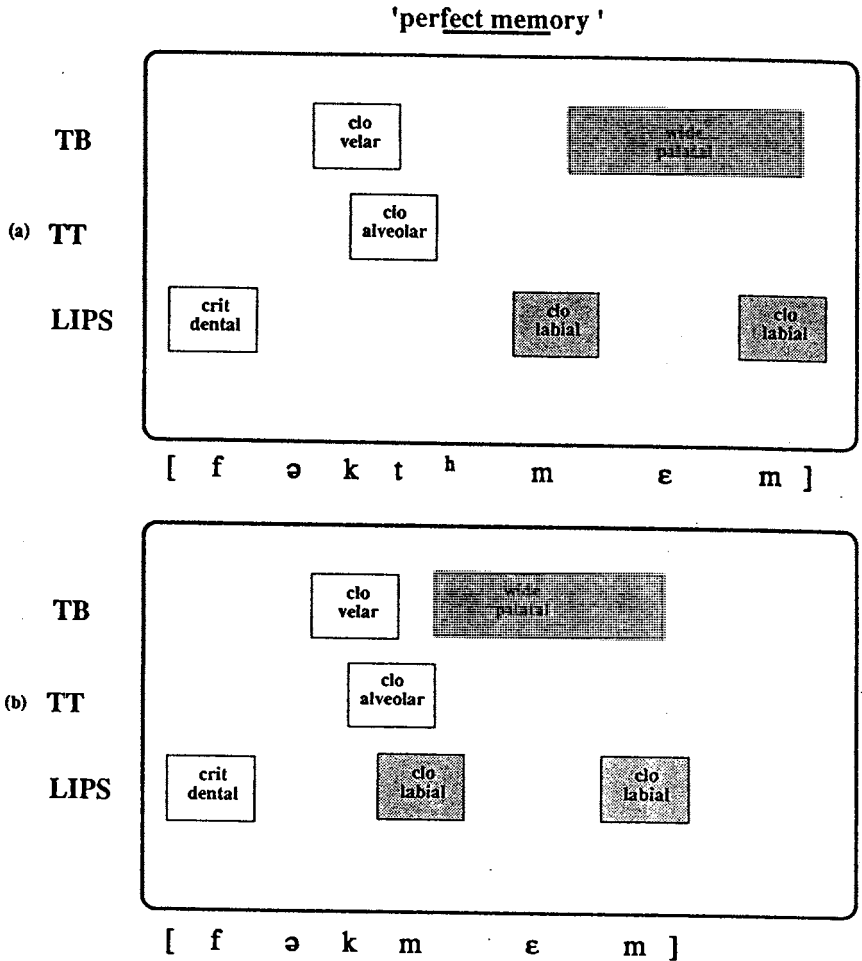


Figure 5

Partial gestural score for two versions of *perfect memory*, posited from observed articulator movements. Last syllable of *perfect* shown in unshaded boxes; first syllable of *memory* shown in shaded boxes. Only oral tiers are shown. (a) Words spoken in word list; (b) words spoken as part of fluent phrase.

the preceding labial fricative and following labial stop. The changes in the posited gestural score can be seen in Fig. 6. Because the velum lowering gesture (VEL [wide]) at the end of *seven* in Fig. 6b overlaps the labial closure, the hiding is perceived as assimilation rather than deletion. Evidence for such hidden gestures (in some cases having reduced magnitude) has also been provided by a number of electropalatographic studies, where they have been taken as evidence of 'partial assimilation' (Barry 1985; Hardcastle & Roach 1979; Kohler 1976 (for German)). Thus, from a gestural point of view, deletion (5a) and assimilation (5b) of



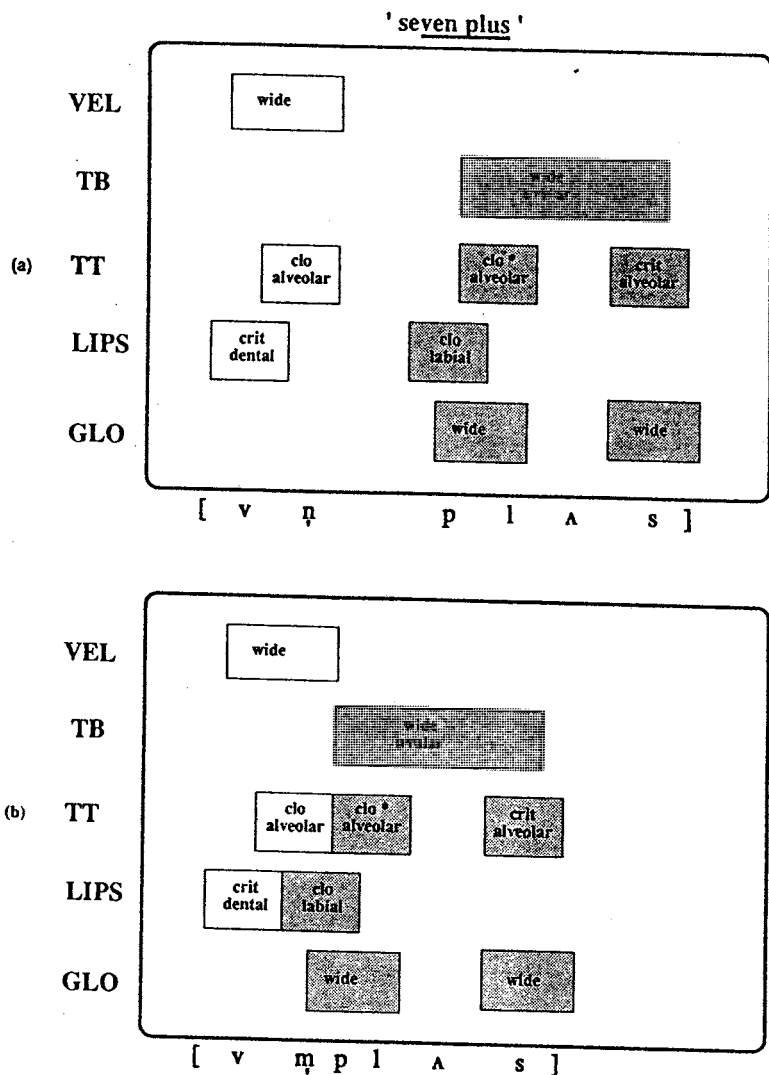


Figure 6

Partial gestural score for two versions of *seven plus*, posited from observed articulator movements. Last syllable of *seven* shown in unshaded boxes; *plus* shown in shaded boxes. The starred [alveolar clo] gesture indicates that laterality is not represented in these scores. (a) Spoken at slow rate; (b) spoken at fast rate.

final alveolars may involve exactly the same process – increase of overlap resulting in a hidden gesture.

When overlap is increased between two gestures on the same articulatory tier, rather than on different articulatory tiers as in the above examples, the increased overlap results in blending between the dynamical

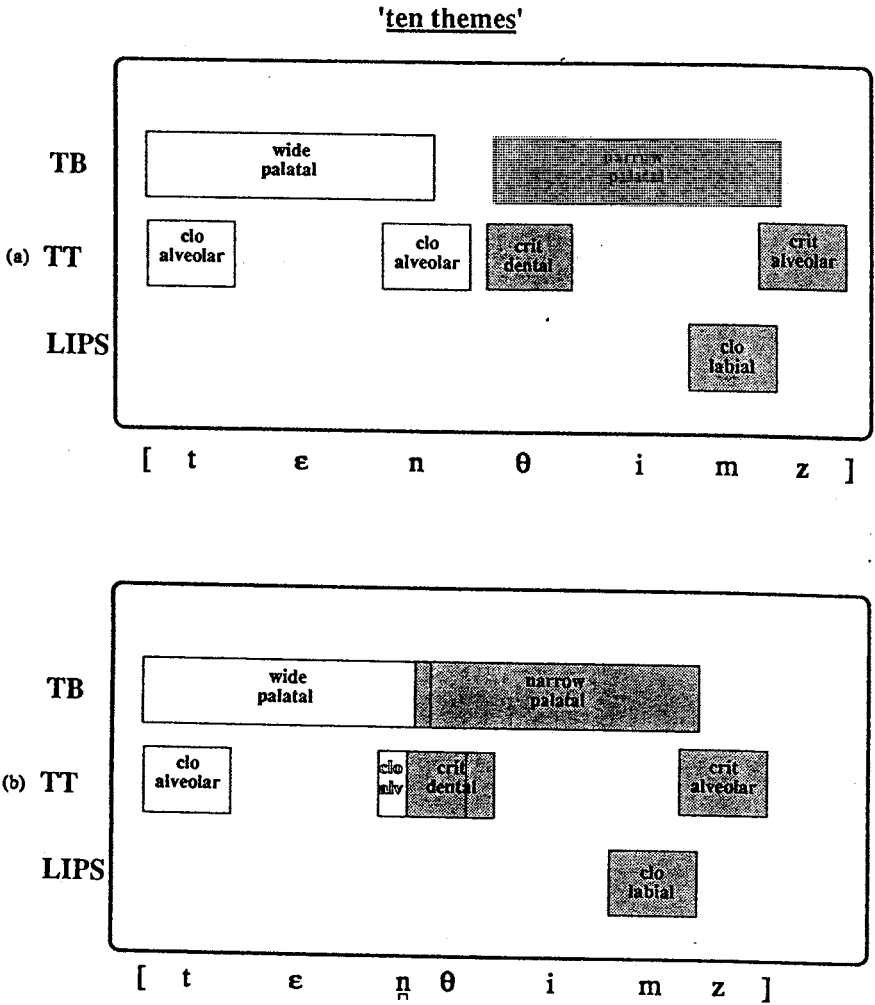


Figure 7

Hypothesised gestural score for two versions of *ten themes*. *ten* shown in unshaded boxes; *themes* shown in shaded boxes. Only oral tiers are shown. (a) Spoken with pause between words; (b) spoken as part of fluent phrase.

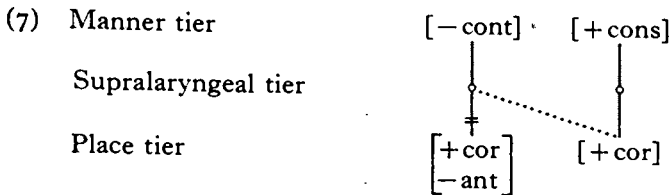
parameters of the two gestures rather than hiding. The trajectories of the tract variables shared by the two gestures are affected by the differing amounts of overlap. Evidence for such blending can be seen in examples like (6):

(6) /tɛn θimz/ → [tɛn̩θimz]

The apparent assimilation in this case, as well as in many other cases that involve conflicting requirements for the same tract variables, has been

characterised by Catford (1977) as involving an accommodation between the two units. This kind of accommodation is exactly what is predicted by parameter blending, assuming that the same underlying mechanism of increased gestural overlap occurs here as in the examples in (5). The (partial) gestural scores hypothesised for (6), showing the increase in overlap, are displayed in Fig. 7. The hypothesis of gestural overlap, and consequent blending, makes a specific prediction: the observed motion of the TT tract variables resulting from overlap and blending should differ from the motion exhibited by either of the individual gestures alone. In particular, the location of the constriction should not be identical to that of either an alveolar or a dental, but rather should fall somewhere in between. If this prediction is confirmed, then three (superficially) different fluent speech processes – deletion of final alveolar stops in clusters, assimilation of final alveolar stops and nasals to following labials and velars, and assimilation of final alveolar stops to other tongue tip consonants – can all be accounted for as the consequence of increasing overlap between gestures in fluent speech.

How does the gestural analysis of these fluent speech alternations compare with analyses proposed by other theories of non-linear phonology? Assimilations such as those in (6) have been analysed by Clements (1985) as resulting from a rule that operates on sequences of alveolar stops or nasals followed by [+coronal] consonants. The rule, whose effect is shown in (7), delinks the place node of the first segment and associates the place node of the second segment to the first (by spreading):

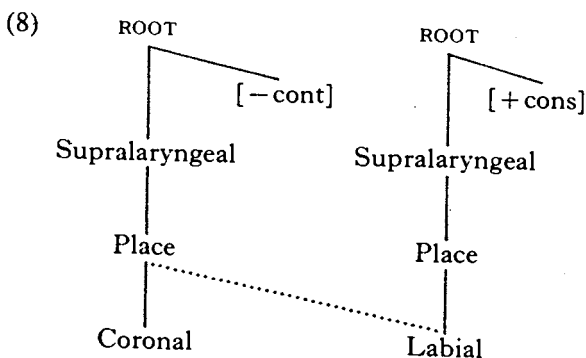


Since the delinked features are assumed to be deleted, by convention, and not realised phonetically, the analysis in (7) predicts that the place of articulation of the assimilated sequence should be indistinguishable from that of the second consonant when produced alone. This claim differs from that made by the blending analysis, which predicts that the assimilated sequence should show the influence of both consonants. These conflicting predictions can be directly tested.

The overlap analysis accounts for a wider range of phenomena than just the assimilation in (6). The deletion of final alveolars (in clusters) and the assimilation of final alveolars to following stops in (5) were also shown to be cases of hiding due to increased gestural overlap. Clements' analysis does not handle these additional cases, and cannot be extended to do so without major reinterpretation of autosegmental formalisms. To see that this is the case, suppose Clements' analysis is extended by eliminating the [+coronal] requirement on the second segment (for fluent speech). This

would produce an assimilation for a case like (5b), but it would not be consistent with the data showing that the alveolar closure gesture is, in fact, still produced. To interpret a delinked gesture as one that is articulatorily produced, but auditorily hidden, would require a major change in the assumptions of the framework.

Within the framework of Sagey (1986), the cases of hiding in (5) could be handled, but the analysis would fail to capture the parallelism between these cases and the blending case in (6). In Sagey's framework, there are separate class nodes for each of the independent oral articulators, corresponding to the articulatory tiers. Thus, an assimilation like that in (5b) could be handled as in (8): the labial node of the second segment is spread to the preceding segment's place node, effectively creating a 'complex' segment involving two articulations:



However, the example in (6) could not be handled in this way. In Sagey's framework, complex segments can only be created in case there are different articulator nodes, whereas in (6), the same articulator is involved. Thus, an analysis in Sagey's framework will not treat the examples in (5) and (6) as resulting from a single underlying process. If the specific prediction made by the overlap and blending analysis for cases like (6) proves correct, this would be evidence that a unitary process (overlap) is indeed involved – a unity that is directly captured in the overlapping gesture approach, but not in Sagey's framework.

Finally we note that, in general, the gestural approach gives a more constrained and explanatory account of the casual speech changes. All changes are hypothesised to result from two simple mechanisms, which are intrinsically related to the talker's goals of speed and fluency – reduce the size of individual gestures, and increase their overlap. The detailed changes that emerge from these processes are epiphenomenal consequences of the 'blind' application of these principles, and they are explicitly generated as such by our model. Moreover, the gestural approach makes predictions about the kinds of fluent speech variation expected in other languages. Given differences between languages in the canonical gestural scores for lexical items (due to language-specific phasing principles), the same casual speech processes are predicted to have

different consequences in different languages. For example, in a language such as Georgian in which stops in the canonical form are always released, word-finally and in clusters (Anderson 1974), the canonical gestural score should show less overlap between neighbouring stops than is the case in English. The gestural model predicts that overlap between gestures will increase in casual speech. But in a language such as Georgian, an increase of the same magnitude as in English would not be sufficient to cause hiding. Thus, no casual speech assimilations and deletions would be predicted for such a language, at least when the increase of gestural overlap is the same magnitude as for English.

The usefulness of internal duration and overlap among phonological elements has begun to be recognised by phonologists (Sagey 1988; Hammond 1988). For example, Sagey (1986: 20–21) has argued that phonological association lines, which serve to link the features on separate tiers, 'represent the relation of overlap in time... Thus... the elements that they link... must have internal duration'. However, in non-gestural phonologies, the consequences of such overlap cannot be evaluated without additional principles specifying how overlapping units interact. In contrast, in a gestural phonology, the nature of the interaction is implicit in the definition of the gestures themselves, as dynamical control regimes for a (model) physical system. Thus, one of the virtues of the gestural approach is that the *consequences* of overlap are tightly constrained and made explicit in terms of a physical model. Explicit predictions (e.g. about the relation between the constrictions formed by the tongue tip in *themes* and in *ten themes*, as well as about language differences) can be made that test hypotheses about phonological structures.

In summary, a gestural phonology characterises the movements of the vocal tract articulators during speech in terms of a minimal set of discrete gestural units and patterns of spatiotemporal organisation among those units, using dynamic and articulatory models that make explicit the articulatory and acoustic consequences of a particular gestural organisation (i.e. gestural score). We have argued that a number of phonological properties of utterances are inherent in these explicit gestural constellations, and thus do not require postulation of additional phonological structure. Distinctiveness (Browman & Goldstein 1986, in press) and syllable structure (Browman & Goldstein 1988) can both be seen in gestural scores. In addition, a number of postlexical phonological alternations (Browman & Goldstein 1987) can be better described in terms of gestures and changes in their organisation (overlap) than in terms of other kinds of representations.

## 2 Relation between gestural structures and feature geometry

In the remainder of this paper, we look more closely at the relation between gestural structures and recent proposals of feature geometry. The comparison shows that there is much overall similarity and compatibility

between feature geometry and the geometry of phonological gestures (§2.1). Nevertheless, there are some differences. Most importantly, we show that the gesture is a cohesive unit, that is, a coordinated action of a set of articulators, moving to achieve a constriction that is specified with respect to its *degree* as well as its location. We argue that the gesture, and gestural scores, could usefully be incorporated into feature geometry (§2.2). The gestural treatment of constriction degree as part of a gestural unit leads, however, to an apparent disparity with how manner features are currently handled in feature geometry. We conclude, therefore (§3), by proposing a hierarchical TUBE GEOMETRY that resolves this apparent disparity, and that also, we argue, clarifies the nature of manner features and how they should be treated within feature geometry, or any phonological approach.

## 2.1 Articulatory geometry and gestural descriptors

In this section, the details of the gestural descriptor structures – the distinctive categories of the tract variable parameters – are laid out. Many aspects of these structures are rooted in long tradition. Jespersen (1914), for example, suggested an ‘alphabetic’ system of specifying articulatory place along the upper tract, the articulatory organ involved, and the degree of constriction. Pike (1943), in a more elaborated alphabetic system, included variables for impressionistic characterisation of the articulatory movements, including ‘crests’, ‘troughs’ and ‘glides’ (movements between crests and troughs). More recently, various authors (e.g. Venneman & Ladefoged 1973; Campbell 1974; Halle 1982; Sagey 1986) have argued that phonological patterns are often formed on the basis of the moving articulator used (the articulator set, in the current system). The gestural structures described in this paper differ from these accounts primarily in the explicitness of the functional organisation of the articulators into articulatory gestures, and in the use of a dynamic model to characterise the coordinated movements among the articulators.

The articulatory explicitness of the gestural approach leads to a clear-cut distinction between features of input and features of output. That is, a feature such as ‘sonority’ has very little to do with articulation, and a great deal to do with acoustics (Ladefoged 1988a, b). This difference can be captured by contrasting the *input* to the speech production mechanism – the individual gestures in the lexical entry – and the *output* – the articulatory, aerodynamic and acoustic consequences of combining several gestures in different parts of the vocal tract. The gestural descriptors, then, characterise the input mechanism – they are the ‘features’ of a purely articulatory phonology. Most traditional feature systems, however, represent a conflation of articulatory and acoustic properties. In order to avoid confusion with such combined feature systems, and in order to emphasise that gestural descriptors are purely articulatory, we retain the

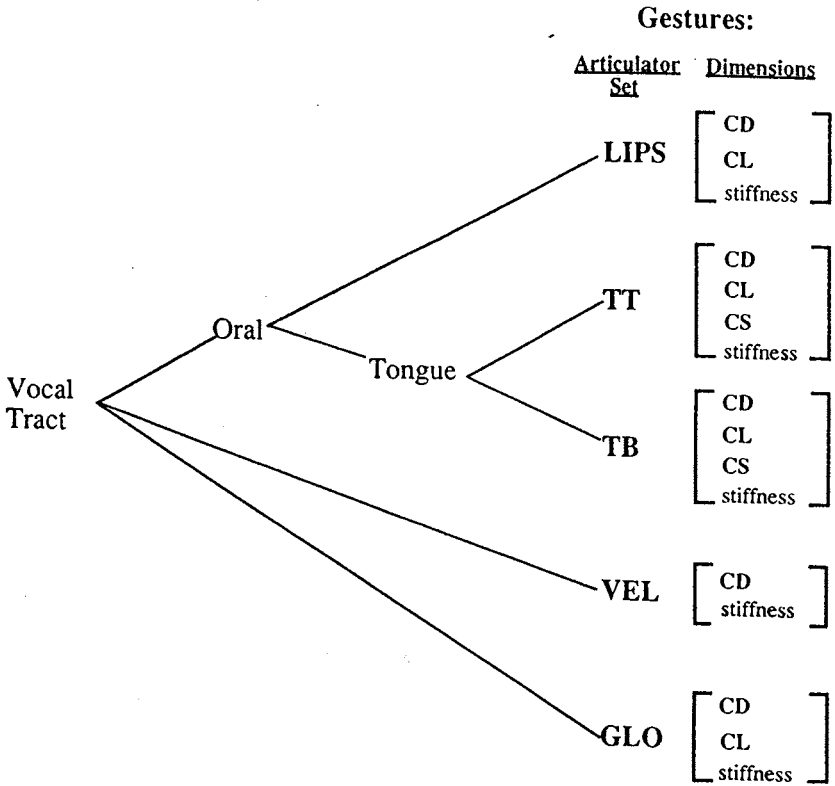


Figure 8  
Articulatory geometry tree.

non-standard terminology developed in the computational model for the category names of the gestural descriptor values.

In §2.1.1, we discuss the descriptors corresponding to the articulator sets and show how they are embedded in a hierarchical articulatory geometry, comparable to recent proposals of feature geometry. In §§2.1.2-5, we examine the other descriptors in turn, demonstrating that they can be used to define natural classes, and showing how the differences between these descriptors and other feature systems stem from their strictly articulatory and/or dynamic status.

*2.1.1 Anatomical hierarchy and articulator sets.* The gestural descriptors listed in Fig. 2 are not hierarchically organised. That is, each descriptor occupies a separate dimension describing the movement of a gesture. However, there is an implicit hierarchy for the sets of articulators involved. This can be seen in Fig. 8, which redisplay (most of) the inventory of articulator sets and associated parameter dimensions from

Fig. 2, and adds a grouping of gestures into a hierarchy based on articulatory independence. Because the gestures characterise movements within the vocal tract, they are effectively organised by the anatomy of the vocal tract. TT and TB gestures share the individual articulators of tongue body and jaw, so they define a class of Tongue gestures. Similarly, both LIPS and Tongue gestures use the jaw, and are combined in the class of Oral gestures. Finally, Oral, velic (VEL), and glottal (GLO) constitute relatively independent subsystems that combine to form a description of the overall vocal tract. This anatomical hierarchy constitutes, in effect, an articulatory geometry.

The importance of an articulatory geometry has repeatedly been noted, for example in the use of articulatory subsystems by Abercrombie (1967) and Anderson (1974). More recently, it has formed the basis of a number of proposals of feature geometry (Clements 1985; Sagey 1986; Ladefoged & Halle 1988; Ladefoged 1988a, b; McCarthy 1988). The hierarchy of the moving articulators is central to all these proposals of featural organisation, although the evidence discussed may be phonological patterns rather than physiological structure. Leaving aside manner features (to be discussed in §§ 2.2.1 and 3), the various hierarchies differ primarily in the inclusion of a Supralaryngeal node in the feature geometries of Clements (1985) and Sagey (1986), and a Tongue node in the articulatory geometry diagrammed in Fig. 8.

There is no Supralaryngeal node included in the articulatory geometry of Fig. 8, because this geometry effectively organises the vocal tract *input*. As we will argue in § 3, the Supralaryngeal node is important in characterising the *output* of the vocal tract. However, using the criterion of articulatory independence, we see no anatomical reason to combine any of the three major subsystems into a higher node in the input hierarchy. Rather than being part of a universal geometry, further combinations of the laryngeal, velic and Oral subsystems into class nodes such as Supralaryngeal, or Central (Oral) *vs.* Peripheral (velic and laryngeal), should be invoked where necessitated by language-particular organisation. For example, the central-peripheral distinction was argued to exist for Toba Batak (Hayes 1986).

The Tongue node in Fig. 8 is proposed on anatomical grounds, again using the criterion of articulatory independence (and relatedness). That is, the TT articulator set shares two articulators with the TB articulator set – the tongue body and the jaw. Put another way, the tongue is an integral structure that is clearly separate from the lips. There is some evidence for this node in phonological patterns as well. A number of articulations made with the tongue cannot be clearly categorised as being made with either TT or TB. For example, laterals in Kuman (Papua New Guinea) alternate between velars and coronals (Lynch 1983, cited in McCarthy 1988). They are always Tongue articulations, but are not categorisable as exclusively TT or TB. Rather, they require some reference to both articulator sets. This is also the case for English laterals, which can



alternate between syllable-initial coronals and syllable-final velars (Ladefoged 1982).

Palatal and palato-alveolar consonants are another type of articulation that falls between TT and TB articulations (Ladefoged & Maddieson 1986; Keating 1988; Recasens in preparation). Keating (1988) suggests that the intermediate status of palatals – partly TT and partly TB – can be handled by treating palatal consonants as complex segments, represented under both the tongue body and tongue tip/blade nodes. However, without the higher level Tongue node, such a representation equates complex segments such as labio-velars, consisting of articulations of two separate articulators (lips and tongue), with palatals, arguably a single articulation of a single predorsal region of the tongue (Recasens in preparation). With the inclusion of the Tongue node, labio-velars and palatals would be similar in being double Oral articulations, but different in being LIPS plus Tongue articulations *vs.* two Tongue articulations. Thus, the closeness of the articulations is reflected in the level of the nodes.

This evidence is suggestive, but not conclusive, on the role of the Tongue node in phonological patterns. We predict that more evidence of phonological patterns based on the anatomical interdependence of the parts of the tongue should exist. One type of evidence should result from the fact that one portion of the tongue cannot move completely independently of the other portions. This lack of independence can be seen, for example, in the suggestion that the tongue body may have a characteristically more backed shape for consonants using the tip/blade in dentals as opposed to alveolars (Ladefoged & Maddieson 1986; Stevens *et al.* 1986). We would expect to find blocking rules based on such considerations.

2.1.2 *Constriction degree (CD)*. CD is the analogue within the gestural approach of the manner feature(s). However, it is crucial to note that, unlike the manner classes in feature geometry, CD is first and foremost an attribute of the gesture – the constriction made by the moving set of articulators – and therefore, at the gestural level, is solely an articulatory characterisation. (This point will be elaborated in §2.2.1.) In our model, CD is a continuum divided into the following discrete ranges: [closed], [critical], [narrow], [mid] and [wide] (Ladefoged 1988b refers to such a partitioning of a continuum as an ‘ordered set’ of values). The two most closed categories correspond approximately to acoustic stops and fricatives; the names used for these categories indicate that these values are articulatory rather than acoustic. Thus, the second degree of constriction, labelled [critical], indicates that critical degree of constriction for a gesture at which some particular aerodynamic consequences could obtain if there were appropriate air flow and muscular tension. That is, the critical constriction value permits frication (turbulence) or voicing, depending on the set of articulators involved (oral or laryngeal). Similarly, [closed]

refers to a tight articulatory closure for that particular gesture; the overall state of the vocal tract might cause this closure to be, acoustically, either a stop or a sonorant. This, in turn, will be determined by the combined effects of the concurrently active gestures. §3 will discuss in greater detail how we account for natural classes that depend on the consequences of combining gestures.

The categorical distinctions among [closed], [critical] and the wider values (as a group) are clearly based on quantal articulatory-acoustic relations (Stevens 1972). The basis for the distinction among wider values is not as easy to find in articulatory-acoustic relations, although [narrow] might be identified with Catford's (1977) [approximant] category. Catford is also careful to define 'articulatory stricture' in solely articulatory terms. He defines [approximant] as a constriction just wide enough to yield turbulent flow under high airflow conditions such as in open glottis for voicelessness, but laminar flow under low airflow conditions such as in voicing. The other descriptors are required to distinguish among vowels. For example, contrasts among front vowels differing in height are represented as [palatal] constriction locations (cf. Wood 1982) with [narrow], [mid] or [wide] CDs, where these categories might be established on the basis of sufficient perceptual and articulatory contrast in the vowel system (Lindblom 1986; Lindblom *et al.* 1983). If additional differentiation is required, values are combined to indicate intermediate values (e.g. [narrow mid]). In addition, [wide] *vs.* [narrow] can be used to distinguish the size of glottal aperture – the CD for GLO – associated with aspirated and unaspirated stops, respectively.

2.1.3 *Constriction location (CL)*. Unlike CD, which differs from its featural analogue of manner by being articulatory rather than acoustic, both CL and its featural analogue of place are articulatory in definition. However, CL differs from place features in not being hierarchically related to the articulator set. That is, the set of articulators moving to make a constriction, and the location of that constriction, are two independent (albeit highly related) dimensions of a gesture. Thus, we use the label 'Oral' rather than 'place' in the articulatory hierarchy, not only to emphasise the anatomical nature of the hierarchy, but also to avoid the conflation of moving articulator and location along the tract that 'place' conveys.

The constriction location refers to the location on the upper or back wall of the vocal tract where a gestural constriction occurs, and thus is separate from, but constrained by, the articulator set moving to make the constriction. Fig. 9 expands the articulator sets, the CL descriptor values and the possible relations between them: the locations where a given set of articulators can form a constriction. Notice that each articulator set maps onto a subset of the possible constriction locations (rather than all possible CLs), where the subset is determined by anatomical possibility. Notice also that there is a non-unique mapping between articulator sets and CL values. For example, both TT and TB can form a constriction at

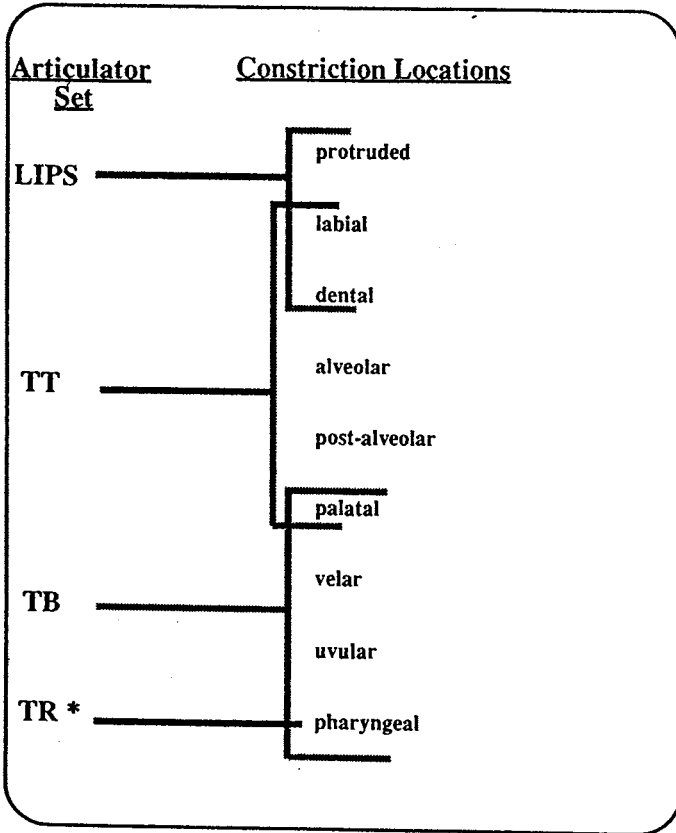


Figure 9

Possible mappings between articulator sets and constriction locations.

the hard palate; indeed, it may be possible for TB to form a constriction even further forward in the mouth. Thus constriction location cannot be subsumed under the moving articulator hierarchy, contrary to the proposals by Ladefoged (1988a) and Sagey (1986), among others.

This can be seen most clearly in the case of the [labial] and [dental] locations, where either the LIPS or TT articulator sets can form constrictions, as shown in (9) (the unusual linguo-labials are discussed in Maddieson 1987). (A similar matrix is presented in Ladefoged & Maddieson 1986, but not as part of a formal representation.) That is, the [labial] constriction location cannot be associated exclusively with the LIPS articulator set. Similarly, the [dental] constriction location cannot be associated exclusively with the TT articulator set. Thus, we view CL as an independent cross-classifying descriptor dimension whose values

cannot be hierarchically subsumed uniquely under particular articulator sets:

		Articulator set	
		LIPS	TT
CL	labial	bilabial	linguo-labial
	dental	labio-dental	dental

We use multivalued (rather than binary-valued) CL descriptors, following Ladefoged (1988b), since the CL descriptor values correspond to categorical ranges of the continuous dynamic parameters. In the front of the mouth, the basis for the discrete ranges of CL presumably involves the actual differentiated anatomical landmarks, so that parameter values are tuned with respect to them. There may, in addition, be relatively invariant auditory properties associated with the different locations (Stevens 1989). For the categories that are further back (palatal and beyond), Wood (1982) hypothesises that the distinct CLs emerge from the alignment of Stevens' quantal considerations with the positioning possibilities allowed by the tongue musculature. To the extent that this set of descriptor values is too limited, it can, again, be extended by combining descriptors, e.g. [palatal velar], or by using 'retracted' and 'advanced' (*Principles of the IPA* 1949).

2.1.4 *Constriction shape (CS)*. In some cases, gestures involving the same articulator sets and the same values of CL and CD may differ in the shape of the constriction, as looked at in the frontal, rather than sagittal, plane. For example, constrictions involving TT gestures may differ as to whether they are formed with the actual tip or blade of the tongue, the shape of the constriction being 'wider' in the third dimension if produced with the blade. The importance of this difference has been built into recent feature systems as apical *vs.* laminal (Ladefoged 1988a) or [distributed] (Sagey 1986). Some method for controlling such differences needs to be built into our computational model. An additional TT tract variable (TTR) that specifies the orientation (angle) of the tongue tip in the sagittal plane with respect to the CL and CD axes is currently being incorporated into the task dynamic model. It is possible that different settings of this variable will be able to produce the required apical/laminal differences, as well as allowing the sublaminar contact involved in extreme retroflex stops (Catford 1977; Ladefoged & Maddieson 1986).

For TB gestures, an additional tract variable is also required to control cross-sectional shape. One of the relevant shape differences involves the production of laterals, in which at least one of the sides of the tongue does not make firm contact with the molars. Ladefoged (1980) suggests that the articulatory manoeuvre involved is a narrowing of the tongue volume, so that it is pulled away from the sides of the mouth. Such narrowing is an attractive option for an additional TB shaping tract variable. In this

proposal, an alveolar lateral would essentially involve two gestures: a TT closure, and a TB gesture with a [narrowed] value for CS and perhaps defaults for CL and CD. (Some further specification of TBCD would clearly be required for lateral fricatives, as opposed to lateral approximants.) In this sense, laterals would be complex constellations of gestures, as suggested by Ladefoged & Maddieson (1986), and similar to the proposal of Keating (1988) for treating palatals as complex segments. Another role for a TB shaping tract variable might involve bunching for rhotics (Ladefoged 1980). Finally, it is unclear whether an additional shape parameter is required for LIPS gestures. It might be required to describe differences between the two kinds of rounding observed in Swedish (e.g. Lindau 1978; Linker 1982). On the other hand, given proper constraints on lip shape (Abry & Bøe 1986), it may be possible to produce all the required lip shapes with just LA and LP.

2.1.5 *Dynamic descriptors*. The STIFFNESS ( $k$ ) of a gesture is a dynamical parameter, inferred from articulatory motions, that has been shown to vary as a function of gestural CD, stress and speaking rate (Browman & Goldstein 1985; Kelso *et al.* 1985; Ostry & Munhall 1985). In addition, however, we hypothesise that stiffness may be tuned independently, so that it can serve as the primary distinction between two gestures. /j/ and /w/ are two cases in which gestural stiffness as an additional independent parameter may be specified. /j/ is a TB [narrow palatal] gesture and /w/ is a complex formed by a TB [narrow velar] gesture and a LIPS [narrow protruded] gesture. Our current hypothesis is that these gestures have the same CD and CL as for the corresponding vowels (/i/ and /u/), but that they differ in having an [increased] value of stiffness. This is similar to the articulatory description of glides in Catford (1977). Finally, it is possible that the stiffness value that governs the rate of movement into a constriction is related to the actual biomechanical stiffness of the tissues involved. If so, it would be relevant to those gestures that, in fact, require a specific muscular stiffness: trills and taps likely involve characteristic values of oral gesture stiffness, and pitch control and certain phonation types require specified vocal fold stiffnesses (Halle & Stevens 1971; Ladefoged 1988a).

## 2.2 The representation of phonological units

In §2.1 we laid out the details of gestural descriptors and how they are organised by an articulatory geometry. Setting aside for the moment the critical aspects of gestures discussed in §1.3 (internal duration and overlap), the differences between feature geometry and the organisation of gestural descriptors have so far been comparatively minor – primarily the proposed Tongue node and the non-hierarchical relation between constriction location and the moving articulators. In this section, we consider two ways in which a gestural analysis suggests a different organisation of phonological structure from that proposed by the feature geometries

referred to in §2.1.1. First, we present evidence that gestures function as cohesive units of phonological structure (§2.2.1). Second, we discuss the advantages of the gestural score as a phonological notation (§2.2.2).

*2.2.1 The gesture as a unit in phonological patterns.* In Fig. 8, gestures are in effect the terminal nodes of a feature tree. Notice that the constriction degree, as well as constriction location, constriction shape and stiffness, is part of the descriptor bundle. That is, constriction degree is specified directly at the articulator node – it is one dimension of gestural movement. Looked at in terms of the gestural score (e.g. Fig. 3), successive units on a given oral tier contain specifications for both constriction location and degree of that articulator set. This positioning of CD differs from that proposed in current feature geometries. While CL and CS (or their analogues) are typically considered to be dependents of the articulator nodes, CD (or its analogues) is not. Rather, some of the closest analogues to CD – [stricture], [continuant] and [sonorant] – are usually associated with higher levels, either the Supralaryngeal node (Clements 1985) or the Root node (Sagey 1986; Ladefoged & Halle 1988; Ladefoged 1988a, b; McCarthy 1988). Is there any evidence, then, that the unit of the gesture plays a role in phonological organisation? Within the approach of feature geometry, such evidence would consist, for example, of rules in which an articulator set and the degree and location of the constriction it forms either spread or delete together, as a unitary whole.

It is in fact generally assumed, implicitly although not explicitly, that velic and glottal features have this type of unitary gestural organisation. That is, features such as [+nasal] and [–voice] combine the constriction degree (wide) along with the articulator set (velum or glottis). Assuming a default specification of these features ([–nasal] and [+voice]), then denasalisation consists of the deletion of a nasal gesture, and intervocalic voicing of the deletion of a laryngeal gesture. Additional implicit use of a gestural unit can be found in Sagey's (1986) proposal that [round] be subordinate to the Labial node. This is exactly the organisation that a gestural analysis suggests, since [round] is effectively a specification of the degree and nature of the constriction of the lips.

It is in the case of primary oral gestures that proposals positioning CD at the gestural level and those positioning it at higher levels contrast most sharply. The inherent connection among all the component aspects of making a constriction, or in other words, the unitary nature of the gesture, can be seen most clearly when oral gestures are deleted, a phenomenon sometimes described as delinking of the Place (or Supralaryngeal) node – debuccalisation. In a gestural analysis, when the movement of an articulator to make a constriction is deleted, everything about that constriction is deleted, including the constriction degree.

For example, Thráinsson (1978, cited in Clements & Keyser 1983) demonstrates that, in Icelandic, the productive phenomenon whereby the first of a sequence of two identical voiceless aspirated stops is replaced by [h] consists of deleting the first set of supralaryngeal features. This set of

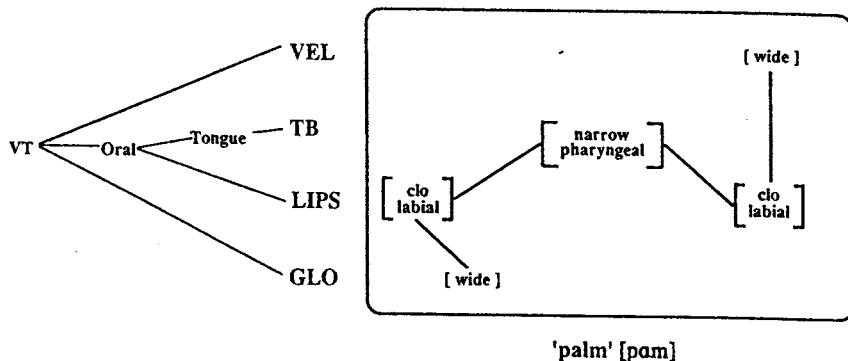


Figure 10

Gestural score for *palm* [pam] in point notation, with articulatory geometry.

supralaryngeal features corresponds to the unit of an oral gesture; it includes the constriction degree, whether described as [clo], [stop] or [-continuant]. Thus, in this example, the entire oral gesture is deleted.

Another example is cited in McCarthy (1988), using data from Straight (1976) and Lombardi (1987). In homorganic stop clusters and affricates (with an intervening word boundary) in Yucatec Maya, the oral portion of the initial stop is deleted – for example, /k#k/ → /h#k/ and (affricate) /ts#t/ → /s#t/. Again in this case the constriction degree is deleted along with the place, supporting a gestural analysis: the entire oral closure gesture is deleted. Note that McCarthy effectively supports this analysis, in spite of his positing [continuant] as dependent on the Root node, when he says ‘a stop becomes a segment with no value for [continuant], which is incompatible with supraglottal articulation’.

Thus, there is some clear evidence in phonological patterning for the association of CD with the articulator node. This aspect of gestural organisation is totally compatible with the basic approach of feature geometry, requiring only that CD be linked to the articulator node rather than to a higher level node. In §3, we will show how this gestural affiliation of CD is also compatible with phonological examples in which CD is seemingly separable from the gesture – where it is ‘left behind’ when a particular articulator-CL combination is deleted, or where it appears not to spread along with the articulatory set and CL. For now, we turn to a second aspect of the gestural approach that could usefully be incorporated into feature geometry, this one involving the use of the gestural score as a two-dimensional projection of the inherently three-dimensional phonological representation.

**2.2.2 Gestural scores, articulatory geometry, and phonological units.** The gestural score, in particular the ‘point’ form in Fig. 4, is topologically similar to a non-linear phonological representation. When combined with

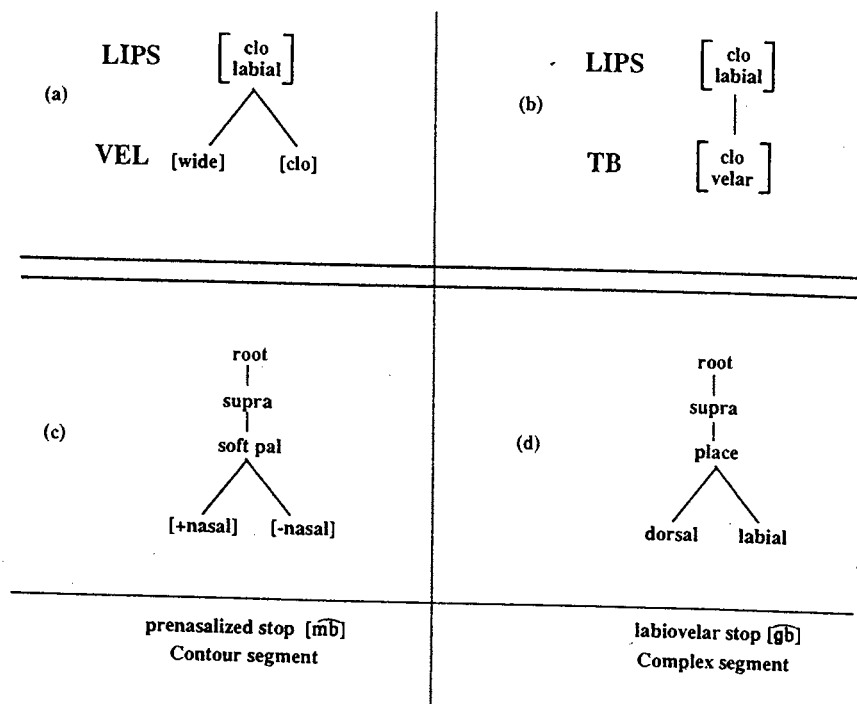


Figure 11

Comparison of gestural score and feature geometry representations for contour and complex segments. (a) Gestural score for contour segment [mb]; (b) gestural score for complex segment [gb]; (c) feature geometry for contour segment [mb]; (d) feature geometry for complex segment [gb].

the hierarchical articulatory geometry, the gestural score captures several relevant dimensions of a non-linear representation simultaneously, in a clear and revealing fashion. Fig. 10 shows a gestural score for *palm*, using point notation and association lines, combined with the articulatory geometry on the left. Note that the geometry tree is represented on its side, rather than the more usual up-down orientation. This seemingly trivial point in fact is a very useful consequence of using gestural scores as phonological notation, as we shall see. It permits spatial organisation such as the articulatory geometry, which is represented on the vertical axis, to be separated from temporal information including sequencing of phonological units, which is represented on the horizontal axis. Such a separation is particularly useful in those instances in which the gestures in a phonological unit are not simultaneous.

For example, prenasalised stops are single phonological units consisting of a sequence of nasal specifications. Fig. 11a depicts a gestural score for prenasalised [mb], which contains a closure gesture on the LIPS tier associated with a sequence of two gestures on the VEL tier, one with



CD = [wide] followed by one with CD = [clo]. Double articulations also constitute a single phonological unit. Fig. 11b displays a gestural score for [gb], which consists of a [clo labial] gesture on the LIPS tier associated with a [clo velar] gesture on the TB tier. Sagey (1986) terms the first type of unit a contour segment, and the second type a complex segment, a terminology that is an apt description of the two figures.

Compare the representations in Figs. 11a and 11b with those in Figs. 11c and 11d, which are Sagey's (1986) feature geometry specifications of the same two phonological units. In the feature geometry representation, the two types of segment appear to be equally sequential, or equally non-sequential. This is a consequence – an unfortunate consequence – of conflating two uses of branching notation, that of indicating (order-free) hierarchical information and that of indicating sequencing information. Thus in Fig. 11d (on the right), the branching lines represent order-free branching in a hierarchical tree. In Fig. 11c, however, the branching lines represent the associations between two ordered elements and a single node on another tier. This conflation results from the particular choice of how to project an inherently three-dimensional phonological structure (feature hierarchy  $\times$  phonological unit constituency  $\times$  time) on to two dimensions.

In the gestural score, the two-dimensional projection avoids this conflation of the two uses of branching notation. Here nodes always represent gestures and lines are always association (or phasing) lines. Thus any branching, as in Fig. 11a, always indicates temporal information about sequencing. The hierarchical information about the articulatory geometry is present in the organisation of the articulatory tiers, and is thus represented (once) by a sideways tree all the way at the left of the gestural score (as in Fig. 10). In this way, the gestural score retains the virtues of earlier forms of phonological notation in which sequencing information was clearly distinguishable, while also providing the benefits of the spatial geometry that organises the tiers hierarchically. It would, of course, be possible to adopt this kind of representation within feature geometry.

Finally, the reader may have observed that, in discussions to now, constituency in phonological units has been indicated solely by using association lines between gestures. In particular, there has been no separate representation of prosodic phonological structure. This is not an inherent aspect of a phonology of articulatory gestures; rather, it is the result of our current research strategy, which is to see how much structure inheres directly in the relations among gestures, without recourse to higher-level nodes. However, once again, it is possible to integrate the gestural score with other types of phonological representation. To exemplify how the gestural score can be integrated with explicit phonological structure, Fig. 12 displays a mapping between the gestural score for *palm* and a simplified version of the prosodic structure of Selkirk (1988), with the articulatory geometry indicated on the left.

In short, although gestures and gestural scores originated in a description of articulator movement patterns, they nevertheless provide constructs that are useful for phonological representation. Gestures

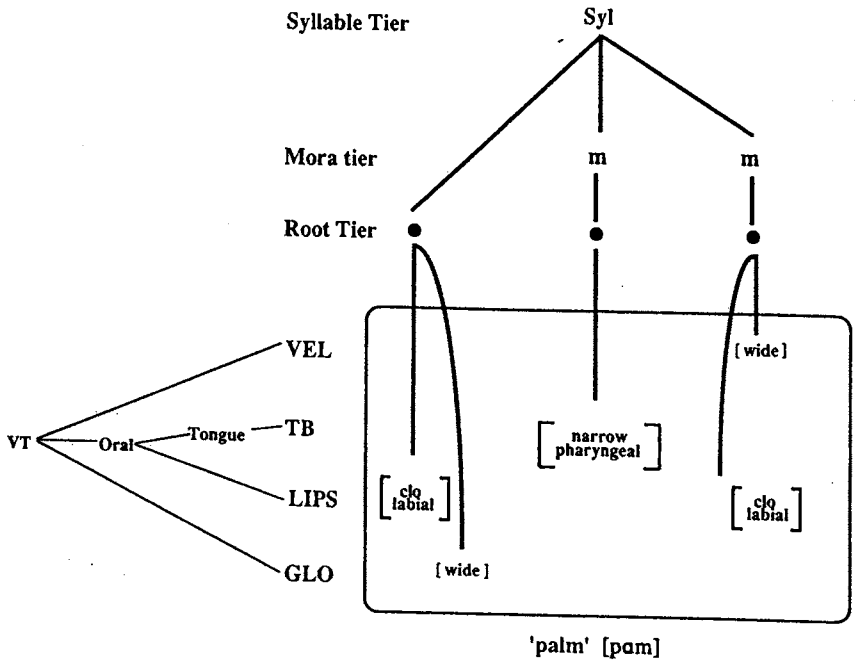


Figure 12

Gestural score for *palm* [pam] in point notation, mapped on to prosodic structure, with articulatory hierarchy on left.

constitute an organisation (combining a moving articulator set with the degree and location of the constriction formed) that can act as a phonological unit. Gestural scores provide a useful notation for non-linear phonological representations, one that permits phonological constituency to be expressed in the same representation with 'phonetic' order information, which is indicated by the relative positions of the gestures in the temporal matrix. The gestures can be grouped into higher-level units, using either association lines or a mapping on to prosodic structure, regardless of the simultaneity, sequentiality or partial overlap of the gestures. In addition, the articulatory geometry organising the tiers can be expressed in the same representation.

### 3 Tube geometry and constriction degree hierarchy

So far, we have been treating individual gestures as the terminal nodes of an anatomical hierarchy. From this perspective, articulatory geometry serves as a way of creating natural classes of gestures on anatomical grounds. In this section, we turn to the vocal tract as a whole, and consider how additional natural classes emerge from the combined effect (articu-

latory, aerodynamic and acoustic consequences) of a set of concurrent gestures. We argue that these natural classes can best be characterised using a hierarchy for constriction degree within the vocal tract based on TUBE GEOMETRY.

Rather than viewing the vocal tract as a set of articulators, organised by the anatomy, tube geometry views it as a set of tubes, connected either in series or in parallel. The sets of articulators move within these tubes, creating constrictions. In other words, articulatory gestures occur within the individual tubes. More than one gesture may be simultaneously active; together, these gestures interact to determine the overall aerodynamic and acoustic output of the entire linked set of tubes. Similarities in the tube consequences of a set of different gestures may lead to similar phonological behaviour: this is the source of acoustic features (Ladefoged 1988a) such as [grave] and [flat] (Jakobson *et al.* 1969) that organise different gestures (in our terms) according to standing wave nodes in the oral tube (Ohala & Lorentz 1977; Ohala 1985).

Within this tube perspective, there is a vocal tract hierarchy that characterises an instantaneous time slice of the output of the vocal tract. Such a hierarchy may appear identical to the feature geometry of Sagey (1986). There are, however, two crucial differences. Unlike Sagey's root node, which 'corresponds neither to anatomy of the vocal tract nor to acoustic properties' (1986: 16), the highest node in the vocal tract hierarchy characterises the physical state of the vocal tract at a single instant in time. In addition, tube geometry characterises manner - CD - at more than just the highest level node. CD is characterised at each level of the hierarchy. Thus, each of the tubes and sets of compound tubes in the hierarchy will have its own effective constriction degree that is completely predictable from the CD of its constituents, and hence ultimately from the CD of the currently active gestures. At the vocal tract level, the effective CD of the supralaryngeal tract, taken together with the initiator power provided by the lungs, determines the nature of the actual airflow through the vocal tract: none (complete occlusion), turbulent flow or laminar flow. This vocal tract hierarchy thus serves as the basis for a hierarchy for CD.

The important point here for phonology is that, in the output system, CD exists simultaneously at all the nodes in the vocal tract hierarchy - it is not isolable to any single node, but rather forms its own CD hierarchy. As we argue in §3.2, all levels of the hierarchy are potentially important in accounting for phonological regularities, and may form the basis of a natural class. First, however, in §3.1, we expand on the nature of tube geometry.

### 3.1 How tube geometry works

Fig. 13 provides a pictorial representation of the tubes that constitute the vocal tract, embedded in the space defined by the anatomical hierarchy of articulatory geometry, in the vertical dimension, and by the constriction

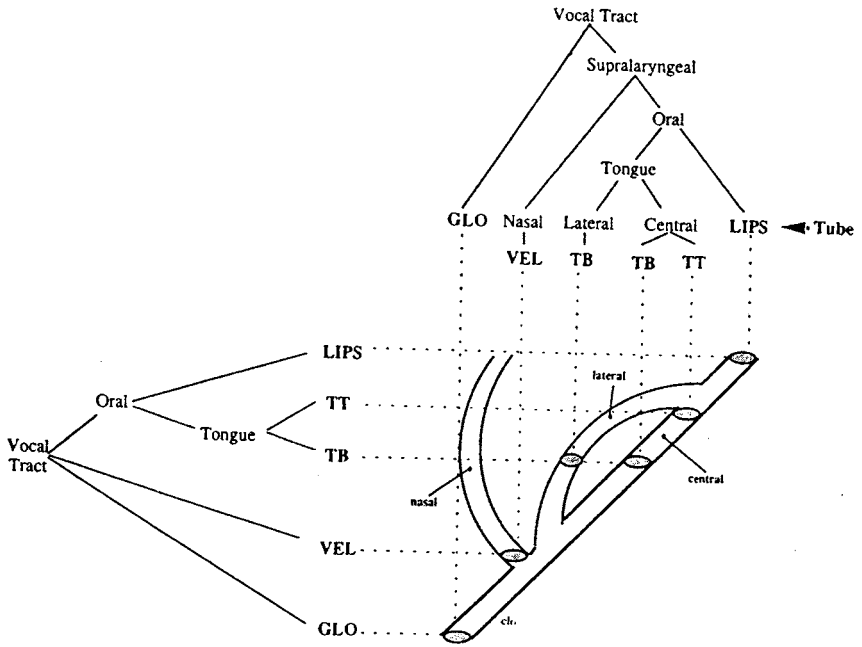


Figure 13  
Vocal tract hierarchy: articulatory and tube geometry.

degree hierarchy of tube geometry, in the horizontal dimension. The constriction action of individual gestures is schematically represented by the small grey disks. Thus, the two dimensions in the figure serve to organise the gestures occurring in the vocal tract tubes, vertically in terms of articulatory geometry, and horizontally in terms of tube geometry. Looking at the tube structure in the centre, we can see that the vocal tract branches into three distinct tubes or airflow channels: a Nasal tube, a Central tongue channel and a Lateral tongue channel. The complex is terminated by GLO gestures at one end. At the other end, the Central and Lateral channels are together terminated by LIPS gestures.

The tube geometry tree, shown at the top of the figure, reflects the combinations of the tubes and their terminators. The tube level of the tree has five nodes, one for each of the three basic tubes and the two terminators. Within each of these basic tubes, the CD will be determined by the CD of the gestures acting within that tube, which are shown as subordinates to the tube nodes. For example, the CD of TT gestures will contribute to determining the CD of the Central tube. TB gestures will contribute to the Central and/or Lateral tube, depending on the constriction shape of the TB gesture. Each of the superordinate nodes corresponds to a tube junction, forming a compound tube from simpler tubes and/or the termination of a tube. Thus, the Central and Lateral

tubes together form a compound tube labelled the Tongue tube. The compound Tongue tube is terminated by the LIPS configuration, which combines with the Tongue tube to form an Oral tube. The Oral and Nasal tubes form another compound tube, the Supralaryngeal, which is terminated by GLO gestures to form the overall Vocal Tract compound tube.

The effective CD at each superordinate node can be predicted from the CD of the tubes being joined and the way they are joined. When tubes are joined in parallel, the effective CD of the compound tube has the CD of the *widest* component tube, that is, the maximum CD. When they are joined in series, the compound tube has the CD of the *narrowest* component tube, that is, the minimum CD. Terminations and multiple constrictions within the same tube work like tubes connected in series. Using these principles, it is possible to 'percolate' CD values from the values for individual gestures up to the various nodes in the hierarchy. Table I shows the possible values for CD of individual gestures, and Table II shows how to determine the CD values at each successive node up through the Supralaryngeal node, referred to hereafter as the Supra node (the Vocal Tract node will be discussed below).

a. LIPS	[CD]	= clo crit open
b. TT	[CD]	= clo crit open
c. TB	[CD, CS = normal]	= clo crit open
d. TB	[CD, CS = narrowed]	= clo crit open
e. VEL	[CD]	= clo open
f. GLO	[CD]	= clo crit open

GLO [crit] is value appropriate for Voicing

[Table I. Possible constriction degree (CD) values at gestural tract variable level: open = (narrow OR mid OR wide)]

a. Nasal	[CD] = VEL [CD]
Lateral	[CD] = TB [CD, CS = narrowed]
Central	[CD] = MIN (TT [CD], TB [CD, CS = normal])
b. Tongue	[CD] = MAX (Central [CD], Lateral [CD])
c. Oral	[CD] = MIN (Tongue [CD], LIPS [CD])
d. Supra	[CD] = MAX (Oral [CD], Nasal [CD])

[Table II. Percolation of CD up through Supralaryngeal node]

The percolation principles follow from aerodynamic considerations. Basically, airflow through a tube system will follow the path of least resistance, as we can illustrate with examples of nasals and laterals. The Oral and Nasal tubes are connected in parallel, forming a compound Supralaryngeal tube. If the Oral tube is [closed] but the Nasal tube is [open], Table II d indicates that the CD of the combined Supra tube will be the wider of the two openings, i.e. [open], as it is in a nasal stop. To take a less obvious example, consider the case where the Nasal tube is [open], but the Oral tube is [crit], i.e. appropriate for turbulence generation under the appropriate airflow conditions. Table II d predicts that in this case as well the Supra CD is [open], implying that there will be no turbulence

generated. This in turn predicts that nasalised fricatives, which consist of exactly the VEL [open] and Oral [crit] gestures, should not exist. That is, given that, at normal airflow rates, the air in this configuration will tend to follow the path of least resistance through the [open] nasal passage, nasalised fricatives would require abnormal degrees of total airflow in order to generate airflow through the [crit] constriction that is sufficient to produce turbulence.

Ohala (1975) has proposed that such nasalised fricatives are, indeed, rare, precisely because they are hard to produce. As Ladefoged & Maddieson (1986) argue, this could account for alternations in which the nasalised counterparts of voiced fricatives are voiced approximants, such as in Guaraní (Gregores & Suárez 1967). However, they also present evidence from Schadeberg (1982) for a nasalised labio-dental fricative in Umbundu. This suggests that the percolation principles may be overridden in certain special cases, perhaps by increased airflow settings.

At the highest level, that of the Vocal Tract, the glottal (GLO) CD and stiffness and the Supra CD combine with initiator (pulmonic) action to determine the actual aerodynamic and acoustic characteristics of airflow through the vocal tract. At this point, then, it becomes more appropriate to label the states of the Vocal Tract in terms of the characteristics of this 'output' airflow, rather than in terms of the CD, which is one of its determining parameters. A gross, but linguistically relevant, characterisation of the output distinguishes the three states defined in Table IIIa: occlusion, noise and resonance. Assuming some 'average' value for initiator power, Table IIIb shows how GLO and Supra CD jointly determine these properties. A complete closure of either system results in occlusion. If GLO is [crit] (appropriate position for voicing, assuming also appropriate stiffness), then it combines with an open Supralaryngeal tract to produce resonance. Any other condition produces noise. In some cases, e.g. GLO [open] and Supra [open], this is weak noise generated at the glottis (aspiration). In other cases, e.g. Supra [crit], the noise is generated in the Oral tube (frication). Further details of the output, such as where the turbulence is generated, and whether voicing accompanies occlusion or noise, are beyond our scope here. Ohala (1983) gives many examples of how the principles involved are relevant to aspects of phonological patterning.

a. <i>VT outputs</i>	
occlusion:	no airflow through VT; silence or low-amplitude voicing
noise:	turbulent airflow
resonance:	laminar airflow with voicing; formant structure
b. VT [CD]	
	= occlusion / Supra [clo] OR GLO [clo]
	= resonance / Supra [open] AND GLO [crit]
	= noise / otherwise

[Table III. *Acoustic consequences at Vocal Tract level*]

Finally, note that the percolation of CD through levels of the tube geometry can be defined at any instant in time, and depends on the actual

size of the constriction within each tube at that point in time, regardless of whether some gesture is actively producing the constriction. Thus, the instantaneous CD is the output consequence of the default values for articulators not under active control, the history of the articulator movements, and the constellation of gestures currently active. Table IV shows the default value we are assuming for each basic tube and terminator. (The default for Lateral is seriously oversimplified, and does not give the right Lateral CD in the case of Central fricatives, although percolation to the Oral level will work correctly.)

Nasal	[CD] = clo
Lateral	[CD] = Central [CD]
Central	[CD] = open
LIPS	[CD] = open
GLO	[CD] = crit

Default values are determined by the effect of the model articulators' neutral configurations within a tube

[Table IV. *Default CD values for basic tubes and terminators*]

### 3.2 The constriction degree hierarchy

The importance of tube geometry for phonology resides in the fact that constriction degree is not isolable to any single level of the vocal tract. Rather, constriction degree exists at a number of levels simultaneously, with the CD at each level in this hierarchy defining a potential natural class. In this section, we present some examples in which CD from different levels is phonologically relevant, and argue that the CD hierarchy is important and clarifying for any featural system, as well as for gestural phonology.

We are proposing (1) that there is a universal geometry for CD in which all nodes in the CD hierarchy are simultaneously present, and (2) that different CD nodes are used to represent different natural classes. This approach differs from that adopted in current feature geometries, in which manner features such as [nasal], [sonorant], [continuant] or [consonantal] are usually represented at a single level in the feature hierarchy. Different geometries, however, choose different levels. For example, [nasal] has been variously considered to be a feature dependent on the highest (Root) node (McCarthy 1988), on the Supralaryngeal node (via a manner node) (Clements 1985), or on a Soft Palate node (Sagey 1986; Ladefoged 1988a, b; Ladefoged & Halle 1988). It is possible that this variability in the treatment of [nasal], and other features, reflects the natural variation in the CD hierarchy, such that different phonological phenomena are captured using CD at different levels in the hierarchy.

To see how this might work, consider the four hierarchies in Fig. 14, which use the tube geometry at the top of Fig. 13 to characterise the linked CD structures for a vowel, a lateral, a nasal and an oral stop. Since tube geometry plays the same role in the representation as the articulatory

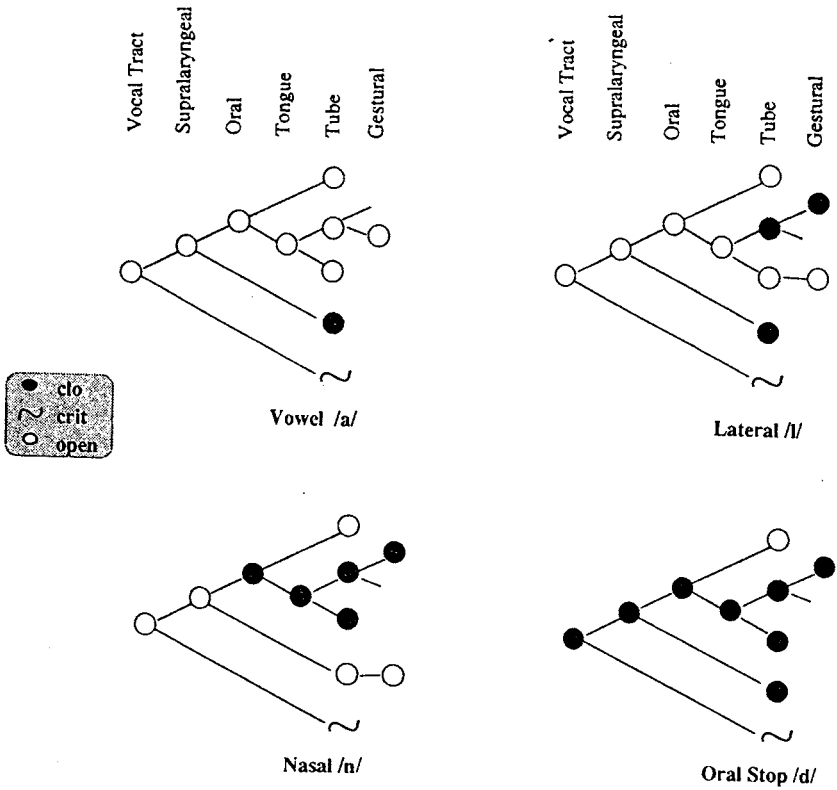


Figure 14  
CD hierarchies for vowel, lateral, nasal and oral stop.

geometry in Fig. 10, the tube hierarchies are rotated 90 degrees just as the anatomical hierarchy was. The circles are a graphic representation of the CD for each node, with the filled circles indicating CD = [clo], the wavy lines indicating CD = [crit], and the open circles indicating CD = [open] (the symbols indicate occlusion, noise and resonance, respectively, at the Vocal Tract level). These CD hierarchies express the various classificatory (featural) similarities and dissimilarities among the four segments, but they do so at different levels.

For example, the nasal has the same characterisation as the vowel and lateral at the Supralaryngeal and Vocal Tract levels, but diverges at lower levels. This aspect of the CD hierarchy thus defines a phonological natural class consisting of nasals, laterals and vowels, where the class is characterised by Supra [open] and VT ['open' = resonance]. This natural class is typically represented by the feature [sonorant], although different feature systems select one or the other of these levels in their definition of sonority. For example, Ladefoged's (1988a, b) acoustic feature distinction of [sonorant] utilises the identity of the value in the CD hierarchy at the



VT level. That is, for Ladefoged [+sonorant] differs from [−sonorant] in terms of output at the VT level. However, for Chomsky & Halle (1968) and Stevens (1972), it is the identity at the Supralaryngeal level that characterises sonority, since they consider /h/ and /ʔ/ to be sonorants even though they differ from the nasal, vowel and lateral at the VT level.

At the Oral level (and below), the nasal and stop display the same linked CD structure, which differs from that of the lateral and that of the vowel. That is, nasals and stops form a natural class in having Oral [clo], whereas laterals and vowels form a class in having Oral [open]. This difference in CD resides at a lower level of the hierarchy than the Vocal Tract or Supralaryngeal levels. And at the lowest levels, the Tube and Gestural, the nasal differs from the stop, lateral and vowel in having both a Nasal [CD] and VEL [CD] that are [open]. Thus, constriction degree at various levels can serve to categorise and distinguish phonological units in different ways.

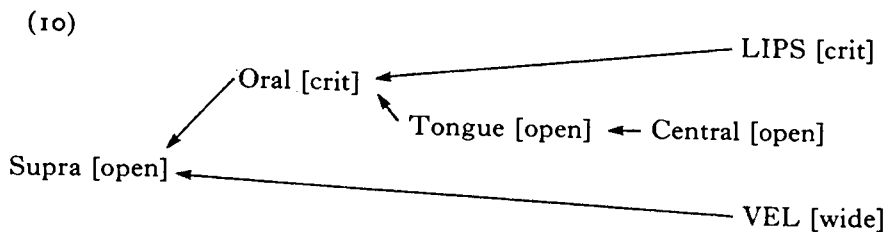
Within traditional feature systems, each of the natural classes described above receives a separate name, which obscures the systematic relation among the various classes. Moreover, even in feature geometry, when manner features are restricted to a single level there is no principled representation of the hierarchical relation among the natural classes. However, feature geometries sometimes incorporate pieces of the CD hierarchy, for example, in the assignment of [nasal] and [continuant] as dependents of two different nodes in the hierarchy (Soft Palate and Root: Sagey 1986), or the assignment of [sonorant] as part of the Root node itself, but [continuant] as a dependent on the Root node (McCarthy 1988). Note, however, that hierarchical relations among manner classes have to be stipulated in feature geometries, whereas such relations are inherent in the CD hierarchy. In addition, the percolation principles of tube geometry provide a mechanism for relating values of the different levels to each other, again something that would need to be stipulated in a hierarchy not based on tube geometry.

All the levels of the CD hierarchy appear to be useful for establishing natural classes, and for relating the CD natural classes to one another. In addition, various kinds of phonological patterns, such as phonological alternations, can be examined in light of this hierarchy. In particular, we can investigate whether regularities are best expressed as processes that treat CD as tied to a particular gesture (as an 'input' parameter) or as a consequence of gestural combination at the various 'output' levels of the CD hierarchy. For example, the cases discussed in §2.2.1 (e.g. /k#k/ → /h#k/) can be best described as processes that delete entire gestures. CD is tied to the gesture and is deleted along with the articulator set and CL. This deletion automatically accounts (by percolation) for the CD changes at other levels of the hierarchy.

In other cases, however, there is much overdetermination – the phonological behaviour can be described equally well from more than one perspective. For example, McCarthy (1988) discusses the common instances in which /s/ → /h/, and glottalised consonants /p' t' k'/ → /ʔ/.

Both of these examples can be straightforwardly analysed as deletion of the oral gesture, as in the cases in §2.2.1. But it is also true that, in both cases, the Vocal Tract output CD is unchanged by the gestural deletion. In the first case, it remains noisy, and in the second it remains an occlusion. Thus, the description of the phonological behaviour could also focus on the (apparent) relative independence of the VT [CD] and the articulator set involved – the articulator set(s) changes, but CD does not. The equivalence of the two perspectives results from that fact that the percolated values of VT [CD] are the same for the two gestures alone and for their combination. Thus, deletion of one or the other will not change the VT [CD]. Examples of this kind, of which there are likely to be many, can be viewed from a dual perspective.

One example of a process for which it seems (at first glance) that a dual perspective cannot be maintained involves assimilation. Steriade (in press) has argued that assimilation is problematic for the gestural approach, since sometimes only the place, and not the manner, features are assimilated. For example, in Kpelle nasals assimilate in place but not manner to a following stop or fricative, so that /N-f/ becomes a sequence (broadly) transcribed as [mv] (Sagey 1986). As Steriade points out, this separation of the place and manner features appears to argue against a gestural analysis, in which the assimilation results from increased overlap between the oral gesture for the fricative (LIPS [crit den]) and the velum lowering gesture (VEL [wide]). Since the nasal does not become a nasal fricative, it would appear either that there is no increased overlap, or that the LIPS [crit] gesture changes its CD when it overlaps the velum lowering gesture. From the perspective of the CD hierarchy, the Supra CD of the nasal is the same before and after the assimilation ([open]), and therefore the Supra level (or VT level) is the significant one for CD, rather than the Gestural level. However, the percolation principles from tube geometry predict that the Supra CD will be [open] even if the nasal is overlapped by a fricative gesture, as derived in (10):

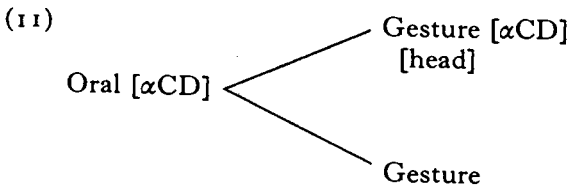


There could also be a TT gesture in the derivation that is either hidden or deleted – the Supra CD would still be [open]. Thus, an increase in gestural overlap could produce both the place assimilation as well as the correct CD at the Supra level, without any change of the Gestural CD. (An articulatory study is needed to determine whether the Gestural CD does indeed change.)

The examples presented so far are processes that are either best

described with CD attached firmly to a given gesture as it undergoes change (e.g. deletion or sliding) or are at least equally well described that way. However, there are phenomena whose description requires a 'loosening' of the relation between CD and the gesture. One such situation, in which the Gestural CD is effectively separated from its articulator set, occurs in historical change (Browman & Goldstein in press), in particular in the types of historical change that Ohala (1981) terms 'listener-based' sound changes. Such cases arise particularly when gestural overlap leads to ambiguity in the acoustic signal that the listener can parse into one of two gestural patterns. In such cases, Browman & Goldstein (in press) argue that there is a (historical) 'reassignment' of the constriction degree parameter between overlapping gestures. This analysis is used to account for the /x/ → /f/ changes in English words like *cough*, based on an argument originally put forth by Pagliuca (1982). Briefly, the analysis proposes that the [crit] descriptor for the TB gesture for /x/ is reassigned to an overlapping LIPS gesture.

A related synchronic situation, involving two overlapping gestures in a complex segment, is discussed in Sagey (1986). Her analysis suggests that manner features must be represented on more than a single node in a feature hierarchy. Specifically, Sagey demonstrates that multiple oral gestures cohering in a complex segment may be restricted to a single distinctive constriction degree. She represents this single distinctive constriction degree at the highest level in the hierarchy, but still must represent which particular articulation bears the contrastive CD. Thus, CD is specified at two different levels. Sagey diagrams the relation between the two levels by a looping arrow drawn between the distinctive level and the 'major' articulator making the distinctive constriction. In the CD hierarchy, the Oral CD would be the lowest possible distinctive level for these double articulations. The gesture that carries the distinctive CD could be marked as [head], and would automatically agree in CD with the mother node, as in (11):



Moreover, in examples such as this, the percolation principles do not contribute to determining the CD of the distinctive node, except in a negative way to be discussed shortly. Sagey argues that the distinctive CD cannot be predicted from physical principles, since in Margi labio-coronals, the coronal articulation is major, and hence in /p̄s/ the less radical constriction is the distinctive constriction, rather than the more radical one. Thus, the relation between the Oral and Gestural levels in (11) must be a statement about a functional phonological unit, a gestural con-

stellation, rather than a statement characterising an instantaneous time slice in terms of tube geometry. That is, only one of the gestures in the constellation may bear a distinctive CD. Nevertheless, the importance of physical overlap relations can also be seen in this example. Maddieson & Ladefoged (1989) have shown that complex segments such as [gb] are not completely synchronous, apparently lending support to the distinction between phonetic ordering, on the one hand, and phonological unordering, on the other hand. However, we argue that the ordering of gestures within a complex segment *is* phonologically important in exactly the case of a phonological unit like Margi /ps̄/, where the distinctive value of CD at higher levels is not predicted from the lower level CDs by the percolation principles. If two gestures overlap completely (i.e. are precisely coextensive), and there are no additional phonetic cues of the sort discussed in Maddieson & Ladefoged, then the percolation principles will determine the higher level constriction degree throughout the entire time-course of the phonological unit, and the distinctive CD will fail to be conveyed. For example, in the case of /ps̄/, if the two gestures were precisely aligned, the (distinctive) frication would never appear at the VT level. Only if the gestures are slightly offset can the distinctive CD be communicated.

In general, the CD hierarchy affords a structure within which a typology of phonological processes can be developed, based on the CD level that seems most relevant. It is an interesting research challenge to develop this typology, and to ask how it is related to other ways of categorising the processes. For example, are there systematic differences in relevant CD level that correlate with whether the process involves spreading (sliding) rules or deletion rules, or with whether the process is prelexical or postlexical?

#### 4 Summary

We have argued that dynamically defined articulatory gestures are the appropriate units to serve as the atoms of phonological representation. Gestures are a natural unit, not only because they involve task-oriented movements of the articulators, but because they arguably emerge as pre-linguistic discrete units of action in infants. The use of gestures, rather than constellations of gestures as in Root nodes, as basic units of description makes it possible to characterise a variety of language patterns in which gestural organisation varies. Such patterns range from the misorderings of disordered speech through phonological rules involving gestural overlap and deletion to historical changes in which the overlap of gestures provides a crucial explanatory element.

Gestures can participate in language patterns involving overlap because they are spatiotemporal in nature and therefore have internal duration. In addition, gestures differ from current theories of feature geometry by including the constriction degree as an inherent part of the gesture. Since

the gestural constrictions occur in the vocal tract, which can be characterised in terms of tube geometry, all the levels of the vocal tract will be constricted, leading to a constriction degree hierarchy. The values of the constriction degree at each higher level node in the hierarchy can be predicted on the basis of the percolation principles and tube geometry. In this way, the use of gestures as atoms can be reconciled with the use of constriction degree at various levels in the vocal tract (or feature geometry) hierarchy.

The phonological notation developed for the gestural approach might usefully be incorporated, in whole or in part, into other phonologies. Five components of the notation were discussed, all derived from the basic premise that gestures are the primitive phonological unit, organised into gestural scores. These components include (1) constriction degree as a subordinate of the articulator node and (2) stiffness (duration) as a subordinate of the articulator node. That is, both CD and duration are inherent to the gesture. The gestures are arranged in gestural scores using (3) articulatory tiers, with (4) the relevant geometry (articulatory, tube or feature) indicated to the left of the score and (5) structural information above the score, if desired. Association lines can also be used to indicate how the gestures are combined into phonological units. Thus, gestures can serve both as characterisations of articulatory movement data and as the atoms of phonological representation.

NOTES

\* This paper has benefited from criticisms from Cathi Best, Alice Faber, Elliot Saltzman, Michael Studdert-Kennedy, Eric Vatikiotis-Bateson, Doug Whalen and two anonymous reviewers. Our thanks to Mark Tiede for help with manuscript preparation, and Zefang Wang for help with the graphics. This work was supported by NSF grant BNS-8520709 and NIH grants HD-01994 and NS-13617 to Haskins Laboratories.

[1] A simple dynamical system consists of a mass attached to the end of a spring – a damped mass-spring model. 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 equation below. 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 tract variable, for example, lip aperture (LA). Thus, the sinusoidal trajectory would describe how lip aperture changes over time:

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

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.

[2] For ease of reference to the gesture, it is possible to use either a bundle of gestural descriptors (or a selected subset) or gestural symbols. We have been trying different approaches to the question of what gestural symbols should be; our present best estimate is that gestures should be treated like archiphonemes. Thus, our current proposal is to use the capitalised form of the voiced IPA

symbol for oral gestures, capitalised and diacritised {H} for glottal gestures, and {±N} for [velic clo] and [velic open] gestures, respectively. In order to clearly distinguish gestural symbols from other phonetic symbols, we enclose them in curly brackets: { }. This approach should permit gestural descriptions to draw upon the full symbol resources of IPA, rather than attempting to develop an additional set of symbols. However, we welcome comments from others on this decision, particularly on the proposal to capitalise gestural symbols, a choice that minimises confusion with phonemic transcriptions – but that leads to a conflict with uvular symbols in the current IPA system.

## REFERENCES

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Abry, C. & L.-J. Bøe (1986). 'Laws' for lips. *Speech Communication* 5. 97–104.
- Anderson, J. M. & C. J. Ewen (1987). *Principles of dependency phonology*. Cambridge: Cambridge University Press.
- Anderson, S. R. (1974). *The organization of phonology*. New York: Academic Press.
- Barry, M. C. (1985). A palatographic study of connected speech processes. *Cambridge Papers in Phonetics and Experimental Linguistics* 4. 1–16.
- Bell-Berti, F. & K. S. Harris (1981). A temporal model of speech production. *Phonetica* 38. 9–20.
- Best, C. T. & D. Wilkenfeld (1988). Phonologically-motivated substitutions in a 20–22 month old's imitations of intervocalic alveolar stops. *JASA* 84. S114.
- Boucher, V. (1988). For a phonology without 'rules': peripheral timing-structures and allophonic patterns in French and English. In S. Embleton (ed.) *The 14th LACUS Forum 1987*. Lake Bluff, Ill.: Linguistic Association of Canada and the U.S. 123–140.
- Boysson-Bardies, B. de, L. Sargart & C. Durand (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language* 11. 1–15.
- Boysson-Bardies, B. de, L. Sargart, P. Halle & C. Durand (1986). Acoustic investigations of cross-linguistic variability in babbling. In Lindblom & Zetterström (1986). 113–126.
- Browman, C. P. & L. Goldstein (1985). Dynamic modeling of phonetic structure. In Fromkin (1985). 35–53.
- Browman, C. P. & L. Goldstein (1986). Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.
- Browman, C. P. & L. Goldstein (1987). Tiers in Articulatory Phonology, with some implications for casual speech. *Haskins Laboratories Status Report on Speech Research SR-92*. 1–30. To appear in Kingston & Beckman (in press).
- Browman, C. P. & L. Goldstein (1988). Some notes on syllable structure in articulatory phonology. *Phonetica* 45. 140–155.
- Browman, C. P. & L. Goldstein (1989). 'Targetless' schwa: an articulatory analysis. Paper presented at the 2nd Conference on Laboratory Phonology, Edinburgh.
- Browman, C. P. & L. Goldstein (in press). Gestural structures and phonological patterns. In Mattingly & Studdert-Kennedy (in press).
- Browman, C. P., L. Goldstein, J. A. S. Kelso, P. Rubin & E. Saltzman (1984). Articulatory synthesis from underlying dynamics. *JASA* 75. S22–S23.
- Browman, C. P., L. Goldstein, E. Saltzman & C. Smith (1986). GEST: a computational model for speech production using dynamically defined articulatory gestures. *JASA* 80. S97.
- Brown, G. (1977). *Listening to spoken English*. London: Longman.
- Campbell, L. (1974). Phonological features: problems and proposals. *Lg* 50. 52–65.

- Catford, J. C. (1977). *Fundamental problems in phonetics*. Bloomington: Indiana University Press.
- Chomsky, N. & M. Halle (1968). *The sound pattern of English*. New York: Harper & Row.
- Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook* 2. 225-252.
- Clements, G. N. & S. J. Keyser (1983). *CV phonology: a generative theory of the syllable*. Cambridge, Mass.: MIT Press.
- Ewen, C. J. (1982). The internal structure of complex segments. In H. van der Hulst & N. Smith (eds.) *The structure of phonological representations*. Part 2. Dordrecht: Foris. 27-67.
- Ferguson, C. A. & C. B. Farwell (1975). Words and sounds in early language acquisition. *Lg* 51. 419-439.
- Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *JPh* 8. 113-133.
- Fowler, C. A. (1981). A relationship between coarticulation and compensatory shortening. *Phonetica* 38. 35-50.
- Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: cyclic production of vowels in sequences of monosyllable stress feet. *Journal of Experimental Psychology: General* 112. 386-412.
- Fowler, C. A., P. Rubin, R. E. Remez & M. T. Turvey (1980). Implications for speech production of a general theory of action. In B. Butterworth (ed.) *Language production*. New York: Academic Press. 373-420.
- Fromkin, V. (ed.) (1985). *Phonetic linguistics: essays in honor of Peter Ladefoged*. New York: Academic Press.
- Fry, D. B. (1966). The development of the phonological system in the normal and the deaf child. In F. Smith & G. Miller (eds.) *The genesis of language: a psycholinguistic approach*. Cambridge, Mass.: MIT Press. 187-206.
- Fujimura, O. (1981a). Temporal organization of articulatory movements as a multi-dimensional phrasal structure. *Phonetica* 38. 66-83.
- Fujimura, O. (1981b). Elementary gestures and temporal organization: what does an articulatory constraint mean? In T. Myers, J. Laver & J. Anderson (eds.) *The cognitive representation of speech*. Amsterdam: North-Holland. 101-110.
- Gimson, A. C. (1962). *An introduction to the pronunciation of English*. London: Arnold.
- Glass, L. & M. C. Mackey (1988). *From clocks to chaos*. Princeton, NJ: Princeton University Press.
- Goldstein, L. & C. P. Browman (1986). Representation of voicing contrasts using articulatory gestures. *JPh* 14. 339-342.
- Gregores, E. & J. A. Suárez (1967). *A description of colloquial Guaraní*. The Hague: Mouton.
- Guy, G. R. (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.
- Haken, H., J. A. S. Kelso & H. Bunz (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics* 51. 347-356.
- Halle, M. (1982). On distinctive features and their articulatory implementation. *NLLT* 1. 91-105.
- Halle, M. & K. Stevens (1971). A note on laryngeal features. *MIT Quarterly Progress Report* 11. 198-213.
- Hammond, M. (1988). On deriving the well-formedness condition. *LI* 19. 319-325.
- Hardcastle, W. J. & P. J. Roach (1979). An instrumental investigation of coarticulation in stop consonant sequences. In P. Hollien & H. Hollien (eds.) *Current issues in the phonetic sciences*. Amsterdam: John Benjamins. 531-540.
- Hayes, B. (1986). Assimilation as spreading in Toba Batak. *LI* 17. 467-499.

- Jakobson, R., C. G. M. Fant & M. Halle (1969). *Preliminaries to speech analysis: the distinctive features and their correlates*. Cambridge, Mass.: MIT Press.
- Jespersen, O. (1914). *Lehrbuch der Phonetik*. Leipzig: Teubner.
- Keating, P. A. (1985). CV phonology, experimental phonetics, and coarticulation. *UCLA Working Papers in Phonetics* 62. 1-13.
- Keating, P. A. (1988). Palatals as complex segments: X-ray evidence. *UCLA Working Papers in Phonetics* 69. 77-91.
- Kelso, J. A. S. & B. Tuller (1987). Intrinsic time in speech production: theory, methodology, and preliminary observations. In E. Keller & M. Gopnik (eds.) *Motor and sensory processes of language*. Hillsdale, NJ: Lawrence Erlbaum. 203-222.
- Kelso, J. A. S., E. Vatikiotis-Bateson, E. L. Saltzman & B. Kay (1985). A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling. *JASA* 77. 266-280.
- Kent, R. D. (1983). The segmental organization of speech. In MacNeilage (1983). 57-89.
- Kent, R. D. & A. D. Murray (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *JASA* 72. 353-365.
- Kingston, J. & M. E. Beckman (eds.) (in press). *Papers in laboratory phonology 1: between the grammar and the physics of speech*. Cambridge: Cambridge University Press.
- Kohler, K. (1976). Die Instabilität wortfinaler Alveolarplosive im Deutschen: eine elektropalatographische Untersuchung. *Phonetica* 33. 1-30.
- Koopmans-van Beinum, F. J. & J. M. van der Stelt (1986). Early stages in the development of speech movements. In Lindblom & Zetterström (1986). 37-50.
- Krakow, R. A. (1989). *The articulatory organization of syllables: a kinematic analysis of labial and velar gestures*. PhD dissertation, Yale University.
- Ladefoged, P. (1980). What are linguistic sounds made of? *Lg* 56. 485-502.
- Ladefoged, P. (1982). *A course in phonetics*. 2nd edn. New York: Harcourt Brace Jovanovich.
- Ladefoged, P. (1988a). Hierarchical features of the International Phonetic Alphabet. *UCLA Working Papers in Phonetics* 70. 1-12.
- Ladefoged, P. (1988b). The many interfaces between phonetics and phonology. *UCLA Working Papers in Phonetics* 70. 13-23.
- Ladefoged, P. & M. Halle (1988). Some major features of the International Phonetic Alphabet. *Lg* 64. 577-582.
- Ladefoged, P. & I. Maddieson (1986). Some of the sounds of the world's languages. *UCLA Working Papers in Phonetics* 64. 1-137.
- Lass, R. (1984). *Phonology: an introduction to basic concepts*. Cambridge: Cambridge University Press.
- Lieberman, A. M. & I. G. Mattingly (1985). The motor theory of speech perception revised. *Cognition* 21. 1-36.
- Lieberman, P. (1984). *The biology and evolution of language*. Cambridge, Mass.: Harvard University Press.
- Lindau, M. (1978). Vowel features. *Lg* 54. 541-563.
- Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (eds.) *Experimental phonology*. Orlando, FL: Academic Press. 13-44.
- Lindblom, B., P. F. MacNeilage & M. Studdert-Kennedy (1983). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie & Ö. Dahl (eds.) *Explanations of linguistic universals*. The Hague: Mouton. 182-203.
- Lindblom, B. & R. Zetterström (eds.) (1986). *Precursors of early speech*. Basingstoke: MacMillan.
- Linker, W. (1982). Articulatory and acoustic correlates of labial activity in vowels: a cross-linguistic study. *UCLA Working Papers in Phonetics* 56. 1-134.



- Locke, J. L. (1983). *Phonological acquisition and change*. New York: Academic Press.
- Locke, J. L. (1986). The linguistic significance of babbling. In Lindblom & Zetterström (1986). 143-160.
- Lombardi, L. (1987). On the representation of the affricate. Ms, University of Massachusetts, Amherst.
- Lynch, J. (1983). On the Kuman liquids. *Language and Linguistics in Melanesia* 14. 98-112.
- McCarthy, J. J. (1988). Feature geometry and dependency: a review. *Phonetica* 45. 84-108.
- McCarthy, J. J. (1989). Linear ordering in phonological representation. *LI* 20. 71-99.
- MacNeilage, P. F. (ed.) (1983). *The production of speech*. New York: Springer.
- Maddieson, I. (1987). Revision of the IPA: linguo-labials as a test case. *Journal of the International Phonetic Association* 17. 26-30.
- Maddieson, I. & P. Ladefoged (1989). Multiply articulated segments and the feature hierarchy. *UCLA Working Papers in Phonetics* 72. 116-138.
- Marshall, R. C., J. Gandour & J. Windsor (1988). Selective impairment of phonation: a case study. *Brain and Language* 35. 313-339.
- Mattingly, I. G. & M. Studdert-Kennedy (eds.) (in press). *Modularity and the motor theory of speech perception*. Hillsdale, NJ: Lawrence Erlbaum.
- Nittrouer, S., M. Studdert-Kennedy & R. S. McGowan (1989). The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research* 32. 120-132.
- Nittrouer, S., K. Munhall, J. A. S. Kelso, B. Tuller & K. S. Harris (1988). Patterns of interarticulator phasing and their relation to linguistic structure. *JASA* 84. 1653-1661.
- Ohala, J. J. (1975). Phonetic explanations for nasal sound patterns. In C. A. Ferguson, L. M. Hyman & J. J. Ohala (eds.) *Nasalfest*. Stanford: Stanford University. 289-316.
- Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick & M. F. Miller (eds.) *Papers from the parasession on language and behavior*. Chicago: Chicago Linguistic Society. 178-203.
- Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In MacNeilage (1983). 189-216.
- Ohala, J. J. (1985). Around 'flat'. In Fromkin (1985). 223-241.
- Ohala, J. J. & J. Lorentz (1977). The story of [w]: an exercise in the phonetic explanation for sound patterns. *BLS* 3. 577-599.
- Oller, D. K. (1986). Metaphonology and infant vocalizations. In Lindblom & Zetterström (1986). 21-36.
- Oller, D. K. & R. E. Eilers (1982). Similarity of babbling in Spanish- and English-learning babies. *Journal of Child Language* 9. 565-577.
- Oller, D. K. & R. E. Eilers (1988). The role of audition in infant babbling. *Child Development* 59. 441-449.
- Ostry, D. J. & K. Munhall (1985). Control of rate and duration of speech movements. *JASA* 77. 640-648.
- Pagliuca, W. (1982). *Prolegomena to a theory of articulatory evolution*. PhD dissertation, State University of New York at Buffalo. Ann Arbor: University Microfilms International.
- Pike, K. L. (1943). *Phonetics*. Ann Arbor: University of Michigan Press.
- Plotkin, V. Y. (1976). Systems of ultimate units. *Phonetica* 33. 81-92.
- Principles of the International Phonetic Association* (1949, reprinted 1978). London: University College.
- Recasens, D. (in preparation). The articulatory characteristics of alveopalatal and palatal consonants. New Haven: Haskins Laboratories.

- Rubin, P., T. Baer & P. Mermelstein (1981). An articulatory synthesizer for perceptual research. *JASA* 70. 321-328.
- Sagey, E. C. (1986). *The representation of features and relations in non-linear phonology*. PhD dissertation, MIT.
- Sagey, E. C. (1988). On the ill-formedness of crossing association lines. *LI* 19. 109-118.
- Saltzman, E. (1986). Task dynamic coordination of the speech articulators: a preliminary model. In H. Heuer & C. Fromm (eds.) *Generation and modulation of action patterns*. New York: Springer. 129-144.
- Saltzman, E., L. Goldstein, C. P. Browman & P. E. Rubin (1988). Dynamics of gestural blending during speech production. Paper presented at 1st Annual International Neural Network Society (INNS), Boston.
- Saltzman, E. & J. A. S. Kelso (1987). Skilled actions: a task dynamic approach. *Psychological Review* 94. 84-106.
- Saltzman, E. & K. Munhall (in press). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*.
- Saltzman, E., P. E. Rubin, L. Goldstein & C. P. Browman (1987). Task-dynamic modeling of interarticulator coordination. *JASA* 82. S15.
- Schadeberg, T. C. (1982). Nasalization in Umbundu. *Journal of African Languages and Linguistics* 4. 109-132.
- Schmidt, R. C., C. Carello & M. T. Turvey (1987). Visual coupling of biological oscillators. *PAW Review* 2. 22-24.
- Selkirk, E. (1988). A two-root theory of length. To appear in *NELS* 19.
- Shockey, L. (1974). Phonetic and phonological properties of connected speech. *Ohio State Working Papers in Linguistics* 17. iv-143.
- Steriade, D. (in press). Gestures and autosegments: comments on Browman and Goldstein's 'Gestures in articulatory phonology'. In Kingston & Beckman (in press).
- Stevens, K. N. (1972). The quantal nature of speech: evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (eds.) *Human communications: a unified view*. New York: McGraw-Hill. 51-66.
- Stevens, K. N. (1989). On the quantal nature of speech. *JPh* 17. 3-45.
- Stevens, K. N., S. J. Keyser & H. Kawasaki (1986). Toward a phonetic and phonological theory of redundant features. In J. S. Perkell & D. H. Klatt (eds.) *Invariance and variability in speech processes*. Hillsdale, NJ: Lawrence Erlbaum. 426-449.
- Straight, H. S. (1976). *The acquisition of Maya phonology: variation in Yucatec child language*. New York: Garland.
- Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. MacKay, W. Prinz & E. Scheerer (eds.) *Language perception and production*. London: Academic Press. 67-84.
- Thelen, E. (1981). Rhythmical behavior in infancy: an ethological perspective. *Developmental Psychology* 17. 237-257.
- Thompson, J. M. T. & H. B. Stewart (1976). *Nonlinear dynamics and chaos*. New York: John Wiley & Sons.
- Thráinsson, H. (1978). On the phonology of Icelandic preaspiration. *Nordic Journal of Linguistics* 1. 3-54.
- Vatikiotis-Bateson, E. (1988). Linguistic structure and articulatory dynamics. Indiana University Linguistics Club.
- Vennemann, T. & P. Ladefoged (1973). Phonetic features and phonological features. *Lingua* 32. 61-74.
- Vihman, M. M. (in press). Ontogeny of phonetic gestures: speech production. In Mattingly & Studdert-Kennedy (in press).

- Vihman, M. M., M. A. Macken, R. Miller, H. Simmons & J. Miller (1985). From babbling to speech: a re-assessment of the continuity issue. *Lg* 61. 397-445.
- Wood, S. (1982). X-ray and model studies of vowel articulation. *Working Papers in Linguistics, Lund University* 23.