Gestural Structures and Phonological Patterns*

Catherine P. Browman and Louis Goldstein†

INTRODUCTION

In this paper, we present a particular view of phonology, one in which lexical items are construed as characterizing the activity of the vocal tract (and its articulators) as a sound-producing system. Central to this characterization is the concept of dynamically defined articular gestures. Such gestures are, we argue, useful primitives for characterizing phonological patterns as well as for analyzing the activity of the vocal tract articulators. Gestural structures can function as lexical representations because distinctiveness is captured by these structures and phonological processes can be seen as operating on them. In addition, the gestural approach has two benefits: it serves to analyze phonological patterns according to their source (articular, acoustic, perceptual, etc.), and it captures the underlying unity among apparently disparate patterns, not only within phonology but also among phonology, perception, and production. These benefits lead to a simpler, more general account of a number of phenomena.

A gestural lexicon is part of the computational model of speech production that we are currently developing with Elliot Saltzman, Philip Rubin and others at Haskins Laboratories (Browman, Goldstein, Kelso, Rubin, & Saltzman, 1984; Browman, Goldstein, Saltzman, & Smith, 1986; Saltzman, Rubin, Goldstein, & Browman, 1987). Our notion of gesture is based on the concept of coordinative structures (Fowler, Rubin, Remez, & Turvey, 1980) as developed in the task dynamic model (Saltzman, 1986; Saltzman & Kelso, 1987), and is consistent with the view of Liberman and Mattingly (1985) that “gestures...have characteristic invariant properties...as the more remote structures that control the [peripheral] movements” (p. 23). Briefly, the basic assumptions are (1) that a primary task in speaking is to control the coordinated movement of groups of articulators (rather than the individual movements of individual articulators), and (2) that these coordinated movements can be characterized using dynamical equations. Each gesture has its own characteristic coordinative pattern.

The hypothesis that the lexicon is composed of dynamically specified gestures has several implications for the motor theory and modularity claims. First, since the same lexicon is assumed to be accessed when an individual is speaking or listening, the hypothesis implies that a listener ultimately recovers the set of gestures that are part of

Haskins Laboratories

Status Report on Speech Research

1989
a given lexical entry. However, the motor theory is a specific proposal about the *mechanism* of recovery (among several possible proposals, see for example, Fowler and Rosenblum, in press) and the arguments for a gestural lexicon should not be construed as supporting this or any other specific proposal. Second, we should note that the use of dynamical equations is not restricted to the description of motor behavior in speech, but has been used to describe the coordination and control of skilled motor actions in general (Cooke, 1980; Kelso, Holt, Rubin, & Kugler, 1981; Kelso & Tuller, 1984a,b; Kugler, Kelso, & Turvey, 1980). Indeed, in its preliminary version, the task dynamic model we are using for speech was exactly the model used for controlling arm movements, with the articulators of the vocal tract simply substituted for those of the arm. Thus, in this respect the model is not consistent with Liberman and Mattingly's (1985) concept of language or speech as a separate module, with principles unrelated to other domains. However, in another respect, the central role of the task in task dynamics captures the same insight as the “domain-specificity” aspect of the modularity hypothesis—the way in which the vocal tract articulators are yoked is crucially affected by the task to be achieved (Abbs, Gracco, & Cole, 1984; Kelso, Tuller, V.-Bateson, & Fowler, 1984).

In this paper, in which we focus on larger language patterns (lexical and phonological structure), we will be demonstrating that many of the patterns observed both in inventories and in alternations can be understood in terms of their articulatory and dynamic structure. The general principles governing motor behavior also interact with other principles, such as articulatory-acoustic relations, in a task-specific way to give rise to a variety of different patterns.

### 2. LEXICAL ORGANIZATION

In this section, we discuss how gestural structures can fulfill one of the most important functions of lexical representation—distinctiveness. We will examine how contrastive words (or morphemes) differ from one another in terms of their component gestures and their organization. In addition, the same gestural structures have an inherent physical meaning: they directly (without mediation of any “implementation” rules) characterize the articulatory movements of those words.

#### 2.1 Gestures and Distinctiveness

A dynamically defined articulatory gesture characterizes the formation (and release) of a constriction within the vocal tract through the movement of (1) a particular set of articulators, (2) towards a particular constriction location (3) with a specific degree of constriction, and (4) in a characteristic, dynamically-described manner. Gestures are not the movements themselves, but rather abstract characterizations of the movements (see Browman & Goldstein, 1987, and Saltzman et al., 1987, for more detailed descriptions of gestures). Contrastive gestures can be distinguished on the basis of the values of these four attributes.

Because the gestures characterize movements within the vocal tract, they are effectively organized by the anatomy of the vocal tract. For example, at the coarsest level, gestures may differ in the major articulatory subsystems (velic, laryngeal, and oral) employed (attribute (1) above). These choices correspond to contrasts in nasality, voicing, and place. The oral subsystem can be further divided into three distinct articulator sets (or synergies), one for the lips, one for the tongue tip, and one for the tongue body. This (hierarchical) articulatory organization is also incorporated in a number of recent approaches to phonological features (e.g., Clements, 1985; Ladefoged & Halle, 1988; Sagey, 1986). Contrasts in the familiar oral place of articulation involve one of these articulator sets moving to a particular constriction location (attribute (2)).
Gestures may also contrast in the degree of such constriction (attribute (3)), corresponding, for example, to stop-fricative-approximant contrasts. Within our computational model, these three attributes determine the set of "tract variables" (Figure 1). Each oral gesture is represented as a pair of tract variable dynamical equations (one for constriction location, the other for constriction degree), each for the appropriate articulatory synergy. Velic and glottal gestures involve single tract variable equations (Brown & Goldstein, 1987; Saltzman et al., 1987). These dynamical equations include (in addition to specification of target value) dynamic descriptors such as stiffness (related to the rate of movement) and damping. Contrasts in these parameters (attribute (4)) may be relevant, for example, to the distinction between vowels and glides.

Thus, contrasts among gestures involve differences in the attributes (1-4) discussed above. We may ask, however, how the discreteness associated with lexical contrasts emerges from these differences in attributes. First, it should be noted that gestures involving movements of different sets of articulators (attribute 1) differ in an inherently discrete way and are thus automatically candidates for distinctiveness (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The importance of this anatomical source of distinctiveness can be seen in the fact that, for the 317 languages reported in Maddieson (1984), 96.8% have nasals, and about 80% use voicing differences among consonants. Moreover, 98.4% have a contrast among gestures of the lips, tongue tip, and tongue body.

With respect to the other attributes, however, gestures may not differ discretely from an articulatory and/or dynamic perspective (although our knowledge about possible principles governing the control and coordination of speech gestures is, at present, quite limited). In these cases, other forces will tend to create discreteness (and hence the automatic potential for distinctive use). For example, as Stevens (1972) has argued, the potential continuum of constriction degree can be partitioned into relatively discrete regions that produce complete closure or turbulence, based on aerodynamic considerations. The stability of these regions is reflected in the fact that all the languages discussed in Maddieson (1984) have stops; 93.4% have fricatives. For wider constriction degrees, as well as for constriction location (for a given set of articulators), articulatory-acoustic relations (Stevens, 1972) and the predisposition to maximally differentiate the gestures and sounds of a language (e.g., Lindblom, MacNeilage, & Studdert-Kennedy, 1983) will help to localize the distinctive values of these variables in fairly well-defined, stable regions for any one language. These regions may, in fact, differ from language to language (e.g., Disner, 1983; Ladefoged, 1984; Lindblom, 1986), indicating that there is some nondeterminacy in the distribution of these regions. Thus the distinctiveness of individual gestures is a confluence of general articulatory, acoustic, and perceptual pressures towards discreteness, with a certain amount of arbitrariness resolved in language-particular ways.

2.2 Gestural Organization and Distinctiveness

Gestures capture distinctiveness not only individually but also in their organization with respect to each other. Here the fact that gestures are both spatial and temporal becomes crucial. That is, because gestures are characterizations of spatiotemporal articulatory events, it is possible for them to overlap temporally in various ways. This fact of overlap, combined with the underlying anatomical structure, gives rise to different organizations that can be used contrastively, and also leads to a variety of phonological processes. The distinctive use of overlap will be discussed in this section, its role in phonological processes in section 3.
<table>
<thead>
<tr>
<th>tract variable</th>
<th>articulators involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP  lip protrusion</td>
<td>upper &amp; lower lips, jaw</td>
</tr>
<tr>
<td>LA  lip aperture</td>
<td>upper &amp; lower lips, jaw</td>
</tr>
<tr>
<td>TTCL  tongue tip constrict location</td>
<td>tongue tip, body, jaw</td>
</tr>
<tr>
<td>TTCD  tongue tip constrict degree</td>
<td>tongue tip, body, jaw</td>
</tr>
<tr>
<td>TBCL  tongue body constrict location</td>
<td>tongue body, jaw</td>
</tr>
<tr>
<td>TBCD  tongue body constrict degree</td>
<td>tongue body, jaw</td>
</tr>
<tr>
<td>VEL  velic aperture</td>
<td>velum</td>
</tr>
<tr>
<td>GLO  glottal aperture</td>
<td>glottis</td>
</tr>
</tbody>
</table>

Figure 1. Tract variables and associated articulators used in the computational model of phonology and speech production described in the text.
2.2.1 Gestural Scores

A lexical item typically consists of a characteristic organization of several gestures. Within the computational model being developed, the concept of phase relations (Kelso & Tuller, 1985) is used to coordinate the set of gestures for a given lexical item. This results in a representation that we call a gestural score (Browman et al., 1986). Figure 2a shows a schematic gestural score for the item “spam.” The five separate rows can be thought of as “tiers”: one each for the velic and glottal subsystems, and three for the oral subsystem, representing the three articulatory synergies. Each box represents a single gesture, its horizontal extent indicating the interval of time during which it is active. (For more details on gestural scores and their use, see Browman & Goldstein, 1987.)

Figure 2 compares two gestural scores, one for “spam” (a) and one for “Sam” (b). There are two additional aspects of distinctiveness that appear when these scores are compared. The only difference between the scores is in the presence of the first bilabial closure gesture in “spam.” The first implication of this is that distinctiveness can be conveyed by the presence or absence of gestures (Browman & Goldstein, 1986a; Goldstein & Browman, 1986), which is an inherently discrete property. The second implication is that distinctiveness is a function of an entire constellation of gestures, and that more complex constellations need not correspond to concatenated segment-sized constellations. That is, the difference between “spam” and “Sam” is represented by the difference between two gestures (alveolar fricative and bilabial closure) vs. one gesture (alveolar fricative), in each case co-occurring with a single glottal gesture. This contrasts with a segment based description of the distinction as two segments vs. one segment, where each segment includes both a glottal and an oral constriction specification (see Browman & Goldstein, 1986a, for further discussion). In this sense, the gestural structures are topologically similar to autosegmental structures postulated on the basis of evidence from phonological alternations (e.g., Clements & Keyser, 1983; Hayes, 1986). Note that, in contrast to segment-based theories, which effectively require simultaneous coordination among all the features composing a segment, there are no a priori constraints on intergestural organization within the gestural framework (other than anatomical structure). The relative “tightness” of cohesion among particular constellations of gestures is a matter for continuing research.

2.2.2 Gestural Overlap

The gestural scores in Figure 2 show substantial temporal overlap among the various gestures. In this section, we examine how differences in degree of overlap can be used distinctively in the case of two gestures, each with approximately the same extent in time as could occur with some “singleton” consonants, for example syllable-initial [n] (velic and oral gestures) and [t] (glottal and oral gestures). The different possibilities for overlap of such gestures are exemplified in Figure 3, where (as in the gestural scores of Figure 2), the horizontal extent of each rectangle represents the temporal interval during which a particular gesture is active. There is a potential continuum ranging from complete synchrony (displayed in row (a)) through partial overlap (row (b)) to minimal overlap (row (c)). Depending on the particular articulatory subsystems involved (shown in the different columns in the figure), as well as the amount of overlap, these gestural combinations have been categorized (by phonologists and phoneticians) as being very different phenomena, as indicated by the labels for each example in the figure. Various, and in some cases ad hoc, phonological features have been employed to capture contrasts in overlap patterns (see Browman & Goldstein, 1986a). We would propose, however, that direct analysis of these organizations in terms of degree of overlap leads to a simpler and more explanatory description of the distribution of these structures in phonological inventories and of their role in phonological processes.

*Gestural Structures and Phonological Patterns*
Figure 2. (a) Gestural score for the word *spam*. (b) Gestural score for the word *Sam*. 
Figure 3. Examples of gestural overlap. Each box represents the temporal interval during which a particular gesture is active. (a) Complete synchrony. (b) Partial overlap. (c) Minimal overlap. The leftmost column contrasts degrees of overlap for velic plus oral gestures, the middle column for glottal plus oral gestures, and the rightmost column for two oral gestures.
The simplest case is when two gestures from different articulatory subsystems overlap completely, both having approximately the same extent in time (Figure 3a, left and middle). This case, which (in some sense) maps in a 1:1 fashion onto a segmentation of the acoustic signal, is often analyzed as a segmental unit, for example, (initial) nasal stops or voiceless unaspirated stops (velic plus oral or glottal plus oral, respectively).

What happens to this simple case, however, when one of the gestures is slid slightly so that the overlap is partial rather than complete (Figure 3b, left and middle)? An aspirated, rather than unaspirated stop, for example, maps onto a sequence of two distinct acoustic events—voiceless stop and aspiration—leading to difficulties, in a purely temporal acoustic measure, in assigning the aspiration to the preceding stop or the following vowel. From an articulatory point of view, however, there are still two gestures simply organized slightly differently, such that there is not complete synchrony between the two (see Lisker & Abramson, 1964; Löfqvist, 1980). About 50 languages in Maddieson’s (1984) survey contrast these two patterns of organization (voiceless unaspirated vs. voiceless aspirated). The similar overlap differences between the velic and oral gestures are contrastive in 15 of the 18 languages with prenasalized stops, i.e., prenasalized stops contrast with nasals (see also Browman & Goldstein, 1986a; Herbert, 1986).

The status of overlap between two gestures within the same articulatory subsystem (oral-oral: Figure 3, right column) differs in several respects from that of overlap between gestures from different articulatory subsystems (oral-glottal or oral-velic). Overlapping oral (consonant) gestures are relatively rare, and seem to be restricted to two types. First, synchronous double oral articulations (Figure 3a: right) involving stops are possible, although they occur in only 6.3% of the languages surveyed by Maddieson (1984); moreover, they are apparently restricted to labial and velar closure gestures. Relatively synchronous alveolar and velar closure gestures also occur, albeit rarely, but appear to be associated with a different airstream mechanism—velaric suction—that produces clicks (Traill, 1985). (In fact, Marchal [1987] suggests the possibility that the suction associated with clicks could develop automatically from the increased overlap of sequential velar and alveolar closures, as he observes currently in French.) Labial-alveolar double articulations have been reported for such languages as Bura and Margi, but the data of Maddieson (1983) show that the gestures involved display, at most, only partial overlap (but cf. Sagey, 1986, for further discussion of the phonological status of these structures). Second, stops may show at least partial overlap with a gesture of constriction degree wider than stop or fricative (Figure 3b: right). For example, labialization of stops is observed in about 13% of the languages in Maddieson (1984) and roughly the same proportion of languages show palatalization. Other than these cases, the possibilities for overlap of oral closure gestures with other constriction degrees (e.g., fricatives) is controversial (see Sagey, 1986). Thus, while it is clear that languages can have two synchronous oral closure gestures, and can also have two such gestures with minimal overlap (that is, consonant clusters as in Figure 3c: right), it is not clear whether any single language contrasts degrees of overlap for oral stops and fricatives.

Thus, when we examine contrast in terms of gestures and their organization, we find restrictions or “gaps” in phonological inventories. These restrictions involve the distinctive use of overlap among oral gestures (compared to the relatively freer use of overlap contrasts across articulatory subsystems) and can probably be accounted for by a combination of anatomical and acoustic factors. Labials and velars, for example, use different sets of articulators (lips and tongue), so that it is possible to produce them simultaneously (see Halle, 1982); they also have similar, mutually reinforcing effects on the acoustic signal, both having comparatively more energy in the lower rather than

Browman & Goldstein
upper end of the spectrum (Jakobson, Fant, & Halle, 1969). Ohala and Lorentz (1978) argue that no other combination of constrictions formed with independent articulators shows this type of acoustic compatibility, and that this could account for the predominance of labiovelars among "double" articulations. Constraints on overlapping of gestures with different constriction degrees can also be understood in terms of the aerodynamic or acoustic interaction of the two gestures. It is simply not possible to have the oral tract simultaneously completely closed and also open enough to permit, for example, frication, even if the two constrictions involve different articulator sets and different locations. The acoustic consequences of the oral tract as a whole will be dominated by the narrowest constriction. Note that this points up an inherent difference between the use of two gestures within the same or within different articulatory subsystems. It is perfectly possible, for example, to have the oral tract closed while the nasal tract remains open: the acoustic consequences of the open nasal tract will not be obscured. As Mattingly (1981) points out, this type of conflict may constrain the organization of oral gestures in order to maintain the perceptual recoverability.

More generally, we may ask, as we did in the case of gestural attributes, how the potential continuum of intergestural overlap (particularly when involving different articulatory subsystems) is partitioned into the three discrete contrasting overlap patterns exemplified in Figure 3. While the answer here is quite speculative, again a combination of dynamic articulatory and acoustic factors may be involved. Recent research on bimanual rhythmic movements has demonstrated discrete, stable coordinative modes (in phase vs. out of phase) whose properties can be understood in terms of differential coupling of non-linear oscillators (e.g., Kay, Kelso, Saltzman, & Schöner, 1987; Turvey, Rosenblum, Kugler, & Schmidt, 1986). To the extent that the coordination of different speech gestures can be analyzed in a similar way, it may be possible to discover analogous coupling modes that correspond to contrastive patterns of gestural overlap. In addition, while the overlap between two gestures forms an articulatory continuum, certain qualitative acoustic events may emerge at some critical degree of overlap. For example, as the glottal gesture slides further to the right between rows (a) and (b) of Figure 3 (middle column), at some point perceptible aspiration will be generated. Such emerging acoustic properties could also contribute to the partitioning of the overlap continuum.

In summary, lexical distinctiveness can be represented in gestural structures in three different ways: differences in gestural attributes, presence vs. absence of particular gestures, and differences in gestural overlap. In many cases, the gestural framework provides a natural basis for discrete categorization, particularly when supplemented with additional aerodynamic, acoustic, and perceptual principles. In addition, certain general tendencies found in the phonological inventories of human languages can be rationalized when the elements of these inventories are described as gestural structures.

3. PHONOLOGICAL PROCESSES

The role of gestures and gestural overlap in phonology is not limited to their ability to capture the distinctions among lexical items. Gestural overlap, in particular, can help explain much variability observed during speaking. For example, as argued in the early formulation of the motor theory (Liberman et al., 1967), much coarticulation can be explained in terms of the overlap of vowel and consonant gestures ("parallel transmission"), and this idea has been incorporated into our gesture-based computational model (Browman & Goldstein, 1987; Saltzman et al., 1987). In addition, as will be exemplified below, simple changes in gestural overlap can account for more extreme forms of variation such as apparent segment assimilations, deletions, and
insertions (Browman & Goldstein, 1987). The explanatory power of gestural overlap springs from the fact that gestures are abstract characterizations of the actual movements of the articulators, movements that occur in space and over time. When two gestures using the same articulator set co-occur, the movements of the articulators will be affected by both gestures. Even when two co-occurring gestures use different sets of articulators, the nature of their overlap can lead, as we will see, to interesting discontinuous effects on the acoustic signal.

The consequences of overlapping gestures are particularly clear in casual speech (Browman & Goldstein, 1987), where casual speech is defined as that fluent subset of fast speech in which reductions typically occur (Gay, 1981; Lindblom, 1983). Indeed, Browman and Goldstein (1987) have proposed that all variability in casual speech may be due to gestural reduction and/or changes in gestural overlap. The same patterns that occur in casual speech are also observed in "natural" phonological processes (e.g., Sloat, Taylor, & Hoard, 1978) and in many types of historical change (Browman & Goldstein, 1986b; Pagliuca, 1982). In section 3.1, we discuss these common patterns, exemplifying how gestures and gestural overlap can contribute to our understanding of the patterns in each of these areas. In sections 3.2 and 3.3, we discuss synchronic and diachronic patterns (i.e., phonological alternations and historical change) that do not correspond to casual speech processes, showing how in many cases the gestural framework clarifies the nature of the patterns.

3.1 Synchronic and Diachronic Patterns Attributable to Casual Speech Processes

If we assume that a speaker's knowledge of a lexical item includes a specification of its gestures and their organization, it is possible to provide an explanatory account for many (synchronic and diachronic) patterns of phonological and phonetic variation. First, many kinds of variation that have been described by allophonic or low-level phonetic rules can be modeled as the automatic consequence of talking, in which overlapping gestures are produced concurrently. Second, additional kinds of variation can be modeled as consequences of two general principles governing the realization of gestures in casual speech: reduction and increase in overlap. For both of these types of variation, no explicit changes in the talker's representation of the items need to be assumed. In addition, for many kinds of variation that cannot be modeled in this way, we can nevertheless establish a relationship between specific casual speech processes and the development (over historical time) of parallel changes in the talker's representation of the gestural structures themselves (how this might occur is discussed below). Such changes to the gestural structures may either be limited to only some of the environments in which a morpheme occurs, which would be a synchronic alternation that can be described by a gestural phonological rule, or the change may affect an item in all its environments, which would be an example of lexical change in a gestural structure. In this section, then, we will discuss patterns originating in the speech production mechanism. We will look first at reduction of individual gestures, and then at consequences of overlapping gestures, noting different consequences for overlap when gestures employ the same or different sets of articulators. Each type of pattern will be exemplified in casual speech, phonological alternation, and historical (lexical) change.

Brown (1977) discusses a class of weakenings, or lenittions, in casual speech in English, where typical examples involve stop consonants weakening to corresponding fricatives: "because" pronounced as [pxəz], "must be" pronounced as [mʌsβɪ]. These changes are reductions in the magnitude of individual stop gestures such that there is incomplete closure, and are an instance of the general tendency in some types of fast
speech to reduce the movement amplitude (e.g., Lindblom, 1983). Such reductions can occur as regular alternations and therefore be identified as phonological rules in non-casual speech. For example, Spanish voiced stops are pronounced as fricatives when they occur intervocally: *dición* ('diction') pronounced with [\θ] in *la dición*; *guerra* ('war') pronounced with [\γ] in *la guerra* (Sloat et al., 1978). And finally, such reductions may occur not just as alternants but as the sole pronunciation of the word, thereby changing the constriction degree of the gesture in the lexical item. For example, Latin intervocalic /b/ is lenited to a fricative in modern Romance languages, e.g., Latin *habere* ('have'), Italian *avere*, French *avoir* (Lass, 1984).

We may inquire specifically how the articulatory or acoustic output of a casual speech process (in this case reduction) can lead to more permanent or "regular" changes in the gestural structures that underlie the articulatory movement. One possibility is that some speakers become attuned to particular instances of casual speech variation (e.g., reduction in the output magnitude of some gesture), and actually shift (slightly) the value of the constriction degree parameter (CD) for that gesture in that direction. In the stop/fricative cases noted above, reducing the CD of a stop gesture will increase the likelihood that casual speech processes will result in an output that would be categorized as a fricative as opposed to a stop. The greater preponderance of fricative outputs could lead to a further reduction in CD, and, in general, to a systematic drift in this parameter value until a new stable value is reached (the value for a fricative, in this case), and thus, an effective recategorization is achieved. Such stable values would coincide with the discrete parameter ranges discussed in section 2.1. While appeals to drift have been made in other accounts of sound change (e.g., the "allophone" drift of Hockett, 1965), the interaction of this mechanism with the nature of gestural structures increases the range of phenomena to which it may be relevant. In particular, drift may be found not only in the parameters of individual gestures, but also in their overlap, leading to a variety of different articulatory and acoustic consequences. These consequences can be related to phonological processes.

Gestural overlap plays a large role in the formation of phonological patterns. (Indeed, it is conceivable that gestural reduction is partly due to the overlap between vowels and consonants, with their opposite requirements for constriction degree.) The consequence of gestural overlap when two gestures involve the same pair of (oral) tract variables (or, as described in Browman & Goldstein, 1987, they occupy the same articulatory tier) follows automatically from the fact that both use the same articulators, leading to blending of the movements of the two. This kind of assimilation occurs in the following examples of casual speech (some from British English): *eight things* pronounced as [\eɪt \θɪŋz]; *come from* pronounced as [\kɔm\frəm] also; *this year* pronounced as [\θɪ\dʒ\jə] (Brown, 1977; Catford, 1977). Such blendings may also occur in the canonical form of words, particularly because the initial consonant and vowel gestures of a syllable may overlap (Browman & Goldstein, 1987; Fowler, 1983; Öhman, 1966). Such canonical blendings are traditionally described phonologically as instances of allophonic variation, e.g., between front and back /k/ in English—*key* vs. *cay* (Ladefoged, 1982)—and can contribute to sound changes: (pre-)Old English *ceap* (presumed to begin with [\k\]) becoming Modern English *cheap*, but Old English *cuman* (also presumed to begin with [\k\]) becoming Modern English *come* (Arlotto, 1972).

When overlapping gestures use different tract variables (and therefore different sets of articulators), they do not affect each other's movements but rather both contribute to the overall vocal tract shape, acoustic output, and perceptual response. One consequence of increasing the overlap between oral gestures from different articulatory tiers (i.e., using different tract variables) is (perceptual) assimilation. Various authors (Barry, 1985; Browman & Goldstein, 1987; Kohler, 1976) have presented articulatory
evidence that alveolar articulations can occur (possibly reduced in magnitude) in assimilations such as “seven plus” pronounced as [ˈsevmplys] (Browman & Goldstein, 1987). In this example, the following labial gesture overlaps the alveolar gesture to such a degree that the alveolar gesture is effectively “hidden” in the acoustic signal. In addition, the labial gesture also partially overlaps the velic lowering gesture. The result is an apparent assimilation of the place of the nasal to the following consonant (from [n] to [m])—an assimilation that involves no change in the individual gestures, but simply increased overlap among gestures.

Another, more striking, consequence of increasing the overlap between gestures is the percept of deletion rather than assimilation. Apparent deletions, which are common in syllable-final position in fluent speech (Brown, 1977), can result from two gestures, on different articulatory tiers, sliding with respect to each other so that one gesture is effectively hidden. Articulatory evidence for sliding has been observed in the utterances “perfect memory” pronounced as [ˈpəfəkˈmemərəl] and “nabbed most” pronounced as [ˈnaebməwz] (Browman & Goldstein, 1987). In these utterances, all the gestures were present: the final alveolar closures in “perfect” and “nabbed” were observed in the movements of the tip of the tongue. However, because of the increased overlap between “perfect” and “memory” (and “nabbed” and “most”), these final alveolar closures were acoustically hidden by the initial labial closures of the following word, resulting in an apparent deletion—acoustically and perceptually, but not articulatorily.

In addition to the casual speech examples above, increased overlap of gestures using different tract variables may also be the source of regular phonological alternations and lexical simplifications. In particular, we hypothesize that an oral gesture may be hidden an increasing proportion of the time through drift in the parameter(s) controlling intergestural phasing. Eventually, a regularly hidden gesture may be deleted from the gestural score, either in particular environments (leading to a synchronic rule), or from the lexical entry entirely. Oral gesture deletion may occur regardless of whether, from a strictly segmental point of view, the perceptual consequences of the increased overlap are partial (i.e., corresponding to casual speech assimilations) or total (i.e., corresponding to apparent deletions) in nature. Examples of both types of gesture deletion can be found, where the likely source of the deletion is, in all cases, hiding due to increased gestural overlap.

An example of oral gesture deletion leading to assimilation involves the assimilation of nasals to the place of the following consonant. Such regular synchronic alternations occur commonly in languages of the world: e.g., Yoruba (an African language) [o m fo] ‘he is jumping’, [o n lo] ‘he is going’, [o ŋ ke] ‘he is crying’ (Bamgbose, 1969, in Sagey, 1986). Nasal place assimilation can also be seen in lexical changes in English, for example in the change from Old French conforter to Modern English comfort (Arlotto, 1972). Oral gestural deletion of the total type may be seen in phonological rules that delete word-final consonants. For example in Lardil, an Australian language, all non-tongue tip consonants are deleted word-finally: [jalu] (‘story’) vs. [jalu-k-in] (‘story’ nonfuture suffix) (Kenstowicz & Kisseberth, 1979). Word-final deletions may also occur in lexical simplifications, for example, Ancient Chinese /fap/ (‘law’) and /pat/ (‘eight’) becoming Mandarin /fa/ and /pa/ (Arlotto, 1972).

The hypothesized association between increased overlap and oral gesture deletion can also be seen as underlying the inventory restrictions on the distinctive use of overlapping oral consonantal gestures noted in section 2.2.2. While units consisting of two oral obstruent gestures can arise from overlap, for example in Margi (Hoffman, 1963, in Sagey, 1986), it is clear that there is a strong tendency against such a phenomenon in languages of the world.
Finally, variation in overlap among a whole constellation of gestures can lead to a percept of segment insertion. Several authors (e.g., Anderson, 1976; Ohala, 1974) have analyzed the epenthetic stop in nasal-fricative sequences such as *something* [ˈsʌmpθɪŋ] and *Samson* [ˈsæmpson] as arising from variation in the relative timing of the velar closure gesture and the oral closure gesture. In particular, if denasalization precedes the release of the closure gesture, then a short interval of oral closure will be produced. Such variation can lead to historical change, as in the Old English *pymle* becoming Modern English *thimble*. Note that in these cases, no gesture is ever added. Rather, the categorical historical change involves drift in the gestural organization to a different stable pattern of overlap.

### 3.2 Other Types of Synchronic and Diachronic Patterns

In the previous section, we discussed synchronic and diachronic patterns whose explanation required reference only to the mechanism of speech production, as indicated by their correspondence to casual speech patterns. In this section, we explore some synchronic and diachronic patterns that do not correspond to casual speech variation; that is, they cannot be completely accounted for by the principles of gestural reduction and increase in overlap. A number of these cases have been analyzed as being acoustically or perceptually based. Yet, as we will see, these factors do not appear to operate independently of the articulatory and gestural structures involved, but rather interact with these structures to produce the change (or synchronic alternation). In particular, two additional principles are hypothesized to account for these cases: reassignment of gestural parameters among temporally overlapping gestures (3.2.1) and misparsing of articulatory movements into their underlying discrete gestural regimes (3.2.2).

#### 3.2.1 Reassignment of Gestural Attributes

The ability of listeners to recover the intended gestures from the "single aspect of the acoustic signal" (Liberman et al., 1967, p. 455) that results from gestural overlap is as important in its occasional failure as in its more frequent success. That is, while in the normal course of events, "listeners use coarticulatory information as information for the influencing segment, and... do not integrate it into their perceptual experience of the segment with which it co-occurs in time" (Fowler, 1984, p. 365), the failure of this ability can lead to changes in the lexical gestural structure (as has been proposed by Ohala, 1981). Such a failure may be exemplified by the change from [x] to [l] in English words such as "cough" and "tough," whose vowels were diphthongs with rounded offglides at the time these changes took place. This sound change has typically been attributed solely to the acoustic similarity of labials and velars (e.g., Jonasson, 1971). However, Pagliuca (1982) shows that most such [x]-[l] changes are not purely acoustic, but rather are conditioned by labial and velar articulations occurring in close proximity prior to the change. That is, in many cases the [x]-[l] change consists of a change in the overlap and constriction degree of the gestures, rather than the insertion of a completely new articulation.

For the English examples like "cough" and "tough," Pagliuca (1982) describes the change as due to "the gradual coarticulation of decaying x with the adjacent rounded diphthong" (p. 171). For words undergoing the [x]-[l] change, the rounded diphthong apparently changed to an unrounded monophthong at the same time that the [x] was changing to [l] (Dobson, 1968, in Pagliuca, 1982). In the gestural framework, all these changes might result simply from an increase in overlap (drifting over time) between the second element of the diphthong ([u]) and the following velar fricative. Figure 4(a) shows the hypothesized gestural score before the change, for the VC portion of the word. Both [o] and [u] describe two gestures: a tongue body gesture (uvular narrow or velar...
narrow, respectively), and a narrow bilabial gesture for rounding. Increasing overlap, as shown in Figure 4(b), would automatically shorten the preceding vowel from a diphthong to a monophthong. As shown in the figure, it would also mean that the lip-rounding gesture for [u] co-occurred with the frication from the velar gesture. Only one additional step would then be needed: the attribution of the frication (presumably on the part of the listener) to the labial gesture rather than to the tongue body (velar) gesture. This re-assignment of the constriction degree parameter value appropriate for frication would result in a bilabial fricative (and a velar gesture which subsequently is deleted), leaving the structure in 4(c). The monophthongal vowel also lowers at some point to [ɔ], and the bilabial fricative becomes [f], neither of which is shown in the figure.

Ohala (1981) also argues for the importance of (partially) overlapping articulations in historical change, although his emphasis is on the explanatory role of the listener rather than of articulatory overlap, per se. For example, Ohala (1981) attributes many examples of assimilation to the (mis)attribution of some acoustic effect to the segment temporally coincident with the effect rather than to segments in the environment: 8th century Tibetan /nub/ 'west' becoming /nu:/ but /lus/ 'body' becoming /ly:/ (Michailovsky, 1975, in Ohala, 1981). In this example, Ohala proposes that, because of coarticulation, /lus/ is auditorily [lys], and therefore the lexical change to a front vowel in /ly:/ results from a re-assignment of the acoustic effects of the overlapping /s/ to the vowel articulation, as the /s/ is deleted. (An analysis along these same lines for similar changes in the history of Lisu is presented in Javkin, 1979). In gestural terms, the situation in Tibetan can be represented as in Figure 5. When the alveolar fricative gesture in (a) is deleted (in (b)), the constriction location of the tongue body gesture is recategorized as palatal, rather than velar, in order to account for effects that were originally due to the overlapping alveolar gesture. As Ohala notes, the acoustic effect on the vowel in [lys] (in (a)) need not result from an articulatory fronting of the tongue body itself—a partially overlapping tongue tip constriction will produce the same auditory effect as a fronted tongue body constriction. This is consistent with the analysis of overlap presented in section 3.1. That is, since the alveolar and vowel gestures are on separate tiers, they interact through the acoustic effect of the gestures occurring simultaneously, rather than through actual blending of constriction targets. Finally, note that the increased length of the vowel in /ly:/ is also automatically accounted for by this analysis. As shown in the figure, the duration of the vowel gestures (tongue body and lips) are assumed to not change (from (a) to (b)), but the part of their duration that is hidden by the overlapping alveolar gesture at the earlier stage is uncovered when the alveolar is deleted. This type of explanation for "compensatory lengthening" phenomena was proposed by Fowler (1983).

Thus, while acoustic and perceptual factors are relevant in accounting for the changes described in this section, the overlapping of two gestures and their interaction are also crucial to the account. Note that these changes do not involve adding articulations that were not there to begin with; rather they involve changes in the parameters of gestures that are already present.

3.2.2 Gestural Misparsing

Other examples of historical change appear to involve introduction of a gesture that was not present at an earlier stage. One such apparent example is the historical introduction of nasalized vowels in words without a nasal consonant in the environment: Sanskrit /sarpa/ ('snake') becoming Hindi /sāp/, Sanskrit /sāsā/ ('breath') becoming Hindi /sās/ (Ohala & Amador, 1981). As analyzed by Ohala and Amador (1981), such "spontaneous" nasalization is acoustically and perceptually based. The acoustic effects on the vowel of high air flow through the open glottis.
Figure 4. Hypothesized gestural scores for three stages of sound change for English cough. Only the VC portion of the word is shown. (a) is the earliest stage and (c) is the latest.
Figure 5. Hypothesized gestural scores before and after Tibetan sound change: /lus/ → /lyː/. 

Browman & Goldstein
(especially for fricatives) are re-interpreted by the listener as nasalization, leading to the introduction of a velar opening gesture (in gestural terms). However, there is an alternative (or perhaps complementary) articulatory account suggested by Ignatius Mattingly (personal communication). In general, for oral constriction gestures that are not nasalized, the associated velum height has been found to vary directly with the gesture's constriction degree, decreasing in the series: obstruents—high vowels—low vowels (Bell-Berti, 1980). A relatively low velum position is found during non-nasalized low vowels, even in a language (such as Hindi) that contrasts nasalization on low vowels (Henderson, 1984). Thus, in the normal production of utterances like /sa/, the velum will lower rapidly from the consonant to the vowel of this utterance. This rapid change in velum position may be misinterpreted by a listener (or by a child learning the language) as an explicit velar lowering gesture. This account would be most plausible in the case of low vowel nasalization (as in the cited Hindi examples), and to accept it, we would need to know how often such spontaneous nasalization is associated with low vs. high vowels. To summarize this proposal, if we hypothesize that the continuous articulatory movement (velum lowering) associated with a word is parsed by the language learner into the discrete gestures (and organization) of the gestural score, then this case involves a misparsing; velum lowering is assigned to an explicit gesture. Note however, that in this view, the "new" added gesture is based on an articulation that is already present, rather than being "suddenly" introduced by listeners with no articulatory basis.

This kind of misparsing is very similar to that proposed by Ohala (1981) to account for various kinds of historical dissimilations. An acoustic (or articulatory) effect is attributed to coarticulation with segments in the environment rather than to the segment itself. For example, in the case of pre-Shona /kumwa/ becoming Shona /kumya/ (Guthrie 1967-70 and Mkanganwi, 1972, in Ohala, 1981), Ohala proposes that listeners attribute the labiality of the /w/ to the preceding labial consonant, and therefore factor it out, leaving the velar component /y/. Ohala's analysis translated into gestural terms is shown in Figure 6. In (a), the superimposed curve illustrates the waxing and waning of lip constriction (over time) that might be expected when the lips are under the control of the two successive lip gestures shown in the score. In (b) we see that roughly the same lip constriction curve would be expected even if there is only a bilabial closure gesture. As in the nasalization case above, we can view this change in terms of how the observed articulatory movements are parsed into the gestures that could give rise to them. The pattern of lip movement is attributed to a single gesture rather than to a pair of overlapping lip gestures. In this case, the misparsing results in too few gestures (thus deleting one), while in the nasalization case, the misparsing results in an additional gesture.

A related pattern of labial dissimilation occurs synchronically in Cantonese. As described by Yip (1988), labial dissimilation in Cantonese operates on both labial consonants and rounded vowels, that is, on gestures involving the lips. A co-occurrence restriction prevents more than a single gesture using the lips from occurring in the same syllable (except that a back rounded vowel may co-occur with a preceding, but not following, labial consonant). Moreover, this co-occurrence restriction is used productively in a Cantonese "secret language" called La-mi. As discussed in Yip (1988), this secret language uses a form of reduplication in which /C1V(C2)/ becomes /LV(C2) C1i(C2)/, e.g., /yat/-~/lat yit/, /kei/-~/lei ki/. However, /sap/ does not become /lap sip/ but rather /lap sit/; similarly, /tim/ becomes /lim t'im/. That is, the co-occurrence restriction includes the entire two syllables in this secret language, so that a final labial stop is replaced with an alveolar stop in order to preserve the general restriction on labial gestures within a unit. Note that in this secret language, a
Figure 6. Hypothesized gestural scores before and after Shona sound change: /kumwa/ — /kumya/. The curve superimposed on the bilabial gestures shows expected the degree of lip constriction over time resulting from these gestures.
distinctive gesture is neither deleted nor added (as in earlier examples of this section),
but rather is replaced by another oral gesture that uses a different set of articulators
(i.e., is on a different tier). Moreover, the example shows that a dissimilatory pattern,
once set up, can be extended beyond the original misparsed environments that
presumably led to the change.

3.3 Patterns Not Attributable to Gestures or Gestural Overlap

Some phonological alternations are so complex as to not permit an adequate
description using gestural principles (even with the acoustic and perceptual
interactions described in the previous sections). Kenstowicz and Kisseberth (1979)
describe a morphophonemic process in Chimwi:ni (a Bantu language spoken in
Somalia) that exemplifies the degree of complexity that languages can attain. The
process involves an interaction between the perfective suffix for verbs and the final
consonant of the stem. The form of the perfective suffix for verbs either contains the
voiceless lateral alveolar fricative [ɭ], or, if the stem-final consonant is [s, z, j, ŋ], it
contains a [z]. This appears to be a kind of assimilatory process. But the class of [s, z, j, ŋ]
is not a natural class articulatorily: while they are all central, and all use the front part
of the tongue, other central tongue tip consonants ([d], [n]) occur with [ɭ] rather than [z].

The behavior of the stem-final consonant adds to the eye-glazing complexity of the
perfective form. Stem-final consonants that are stops in the infinitive form generally
correspond to the central alveolar fricatives [s] and [z] in the perfective form. That is,
the oral gesture can use different articulators in the two forms (for example, [pl]/[s]), as
well as having a different constriction degree (voicing is unchanged). However, [k]
corresponds to [j] not [s], thereby showing some slight hint of articulatory conditioning,
either of the specific articulator or front/back constriction location.

While the Chimwi:ni example displays a certain amount of articulatory patterning, it
is not possible to provide a general statement of the patterns contributing to the
alternations, even using highly abstract projections of gestures (or their acoustic
consequences). At some point in every language, the patterns begin to take on a life of
their own, loosened from their articulatory and/or acoustic underpinnings, and
perhaps respecting other sets of principles (cf. Anderson, 1981). Phonologists have
attempted to describe these patterns (e.g. "crazy rules," Bach & Harms, 1972) as
emerging from the interaction of a number of independent rules, each of which, by
itself, can be simply understood. (Rule telescopy [Wang, 1969] and rule inversion
[Vennemann, 1972] are examples of such interaction).

In summary, we have attempted to show how lexical representations can be viewed as
gestural structures, and to show that to view them thus contributes to an understanding
of phonological inventories and processes. While other principles and sources of
constraint are no doubt required to completely explicate patterns in phonology, we
have been surprised by the range of phenomena that can be handled with the relatively
simple assumptions that we are making: that phonological structures consist of
temporally overlapping, dynamically defined gestural units, that the output of these
structures may be systematically altered (in highly constrained ways) in casual speech,
and that variations in output of these structures can lead (historically) to changes in the
values of gestural parameters—both through drift and through mechanisms such as re­
assignment and misparsing. Indeed, several of the cases that we present in this paper
were initially chosen by us to illustrate where the gestural approach would fail. Once
the gestural analyses were made explicit, however, they were more insightful than even
we expected. We suggest that it is interesting and important to see just how much such
simple structures and principles can contribute to understanding phonological
patterns.
ACKNOWLEDGMENT

This work was supported by NSF grant BNS 8520709 and NIH grants HD-01994 and NS-13617 to Haskins Laboratories. Our thanks to Margaret Dunn, Alice Faber, Ignatius Mattingly, Elliot Saltzman, and Douglas Whalen for helpful comments.

REFERENCES


**Gestural Structures and Phonological Patterns**


FOOTNOTES


†Also Yale University.