Phonetica
Editor: K. Kohler, Kiel

Separatum
Publisher: S. Karger AG, Basel
Printed in Switzerland

Catherine P. Browman
Louis Goldstein

Haskins Laboratories and
Department of Linguistics,
Yale University,
New Haven, Conn., USA

Target Article

Phonetica 1992;49:155–180

Articulatory Phonology: An Overview

Abstract

An overview of the basic ideas of articulatory phonology is presented, along with selected examples of phonological patterning for which the approach seems to provide a particularly insightful account. In articulatory phonology, the basic units of phonological contrast are *gestures,* which are also abstract characterizations of articulatory events, each with an intrinsic time or duration. Utterances are modeled as organized patterns *(constellations)* of gestures, in which gestural units may overlap in time. The phonological structures defined in this way provide a set of articulatorily based natural classes. Moreover, the patterns of overlapping organization can be used to specify important aspects of the phonological structure of particular languages, and to account, in a coherent and general way, for a variety of different types of phonological variation. Such variation includes allophonic variation and fluent speech alternations, as well as 'coarticulation' and speech errors. Finally, it is suggested that the gestural approach clarifies our understanding of phonological development, by positing that prelinguistic units of action are harnessed into (gestural) phonological structures through differentiation and coordination.

Catherine P. Browman
Haskins Laboratories
270 Crown Street
New Haven, CT 06511 (USA)

1 Introduction

Gestures are characterizations of discrete, physically real events that unfold during the speech production process. Articulatory phonology attempts to describe lexical units in terms of these events and their interrelations, which means that gestures are basic units of contrast among lexical items as well as units of articulatory action. From our perspective, phonology is a set of relations among physically real events, a characterization of the systems and patterns that these events, the gestures, enter into. Thus, gestures are phonological events in the sense of Bird and Klein [1990].

While gestures are primitive phonological units, they do not correspond to either features or segments. Rather, they sometimes give the appearance of corresponding to features, and sometimes to segments. The issues discussed throughout are intended, among other things, to help clarify the differences among gestures, features, and segments. In addition, we will emphasize the following point throughout this paper: gestures and gestural organization can be used to capture both categorical and gradient information. Section 1 will present an overview of articulatory phonology, touching on a number of key aspects. Sections 2 and 3 will expand on examples in which a gestural analysis appears particularly fruitful. We will end in Section 4 with a discussion of how articulatory gestures provide a felicitous framework for dealing with language development.

1.1. Gestures as Dynamic Articulatory Structures

Gestures are events that unfold during speech production and whose consequences can be observed in the movements of the speech articulators. These events consist of the formation and release of constrictions in the vocal tract. To help in explicitly modeling these events, gestures are defined in terms of *task dynamics* [Saltzman, 1986; Saltzman and Kelso, 1987; Saltzman and Munhall, 1989]. Task dynamics has been used to model different kinds of coordinated multi-articulator actions, including those involved in reaching and those involved in speaking. In the case of speech, the tasks involve the formation of various constrictions relevant to the particular language being spoken. Task dynamics describes such tasks using damped second-order dynamical equations to characterize the movements [see Browman and Goldstein, 1990, and Hawkins, 1992, for further discussions of the use of task dynamics to characterize speech].

One important aspect of task dynamics is that it is the motion of *tract variables* and not the motion of individual articulators that is characterized dynamically. A tract variable characterizes a dimension of vocal tract constriction; the articulators that contribute to the formation and release of this constriction are organized into a coordinative structure [Turvey, 1977; Fowler et al., 1980]. For example, the tract variable of lip aperture is affected by the action of three articulators: the upper lip, the lower lip, and the jaw. The current tract variables, and their component articulators, are displayed in figure 1. An individual tract variable control regime is specified in terms of the set of articulators used to achieve a constriction and the values of the parameters in the dynamic equation describing its movement: target (rest position), stiffness, and damping. These parameters provide a kind of internal structure for a control regime that underlies the spatiotemporal event in all its instances. A *gesture* in articulatory phonology is specified using a set of related tract variables. For example, in the oral tract the constriction location and degree are two dimensions of the same constriction, and therefore are considered related tract variables. In figure 1, related

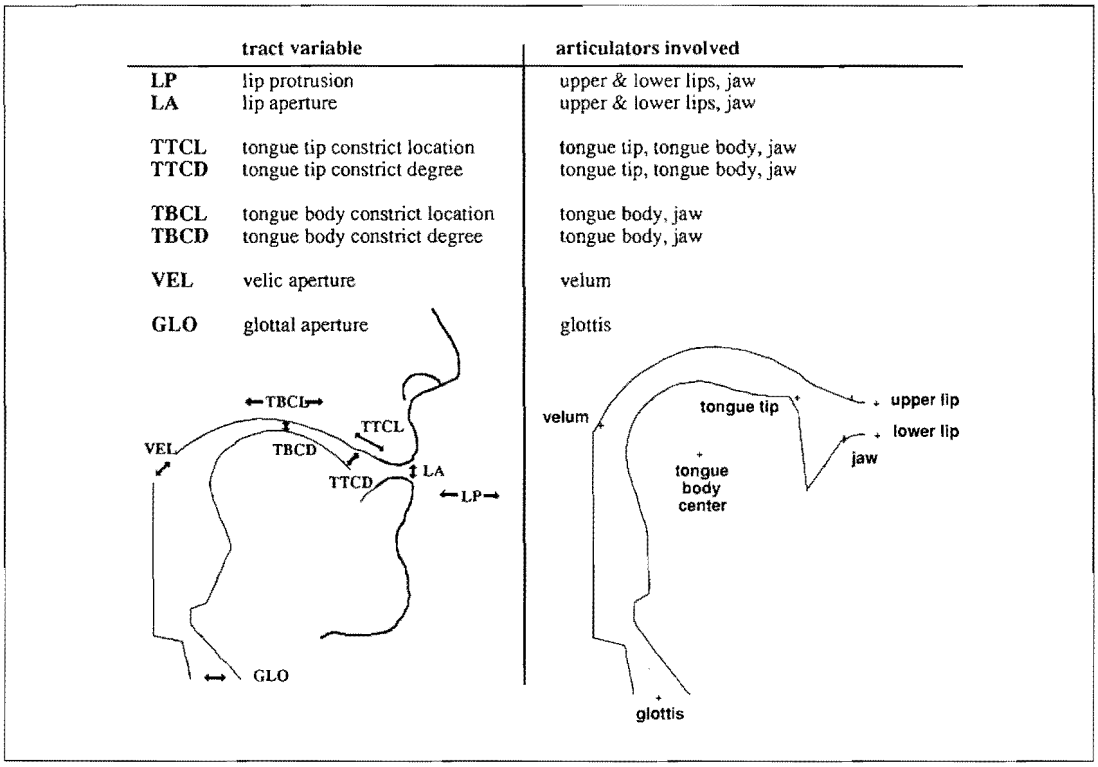

Fig. 1. Tract variables and associated articulators.

tract variables contain the same first letter(s) in their names. Note that this means that each gesture is a local constriction, defined with respect to one of the five tract variable sets shown in the figure (Lips, Tongue Tip, Tongue Body, VELum, GLOttis).

Gestures can function as primitives of phonological contrast. That is, two lexical items will contrast if they differ in gestural composition. This difference can involve the presence or absence of a given gesture, parameter differences among gestures, or differences among organizations of the same gestures (discussed further in Section 1.2). This can be illustrated with the aid of displays showing the arrangement of gestural events over time. Lexical items contrast gesturally, first of all, if a given gesture is present or absent (e.g.,'add' vs. 'had', fig. 2a, b; 'add' vs. 'bad', fig. 2a, c; 'bad' vs. 'pad', fig. 2c, d; 'pad' vs. 'pan', fig. 2d, f). We assume that, in speech mode, the larynx is positioned appropriately for voicing unless otherwise instructed. Note that 'had' and 'bad' would typically be considered to differ from 'add' by the presence of a segment, while 'bad' and 'pad', and 'pad' and 'pan', would contrast only in a single feature, voicing or nasality, respectively. Gesturally, all these contrasts are conveyed by the presence or absence of a single gesture. Another kind of contrast is that in which gestures differ in their assembly, i.e., by involving different sets of articulators and tract variables, such as lip closure vs. tongue tip closure (e.g.,'bad' vs.

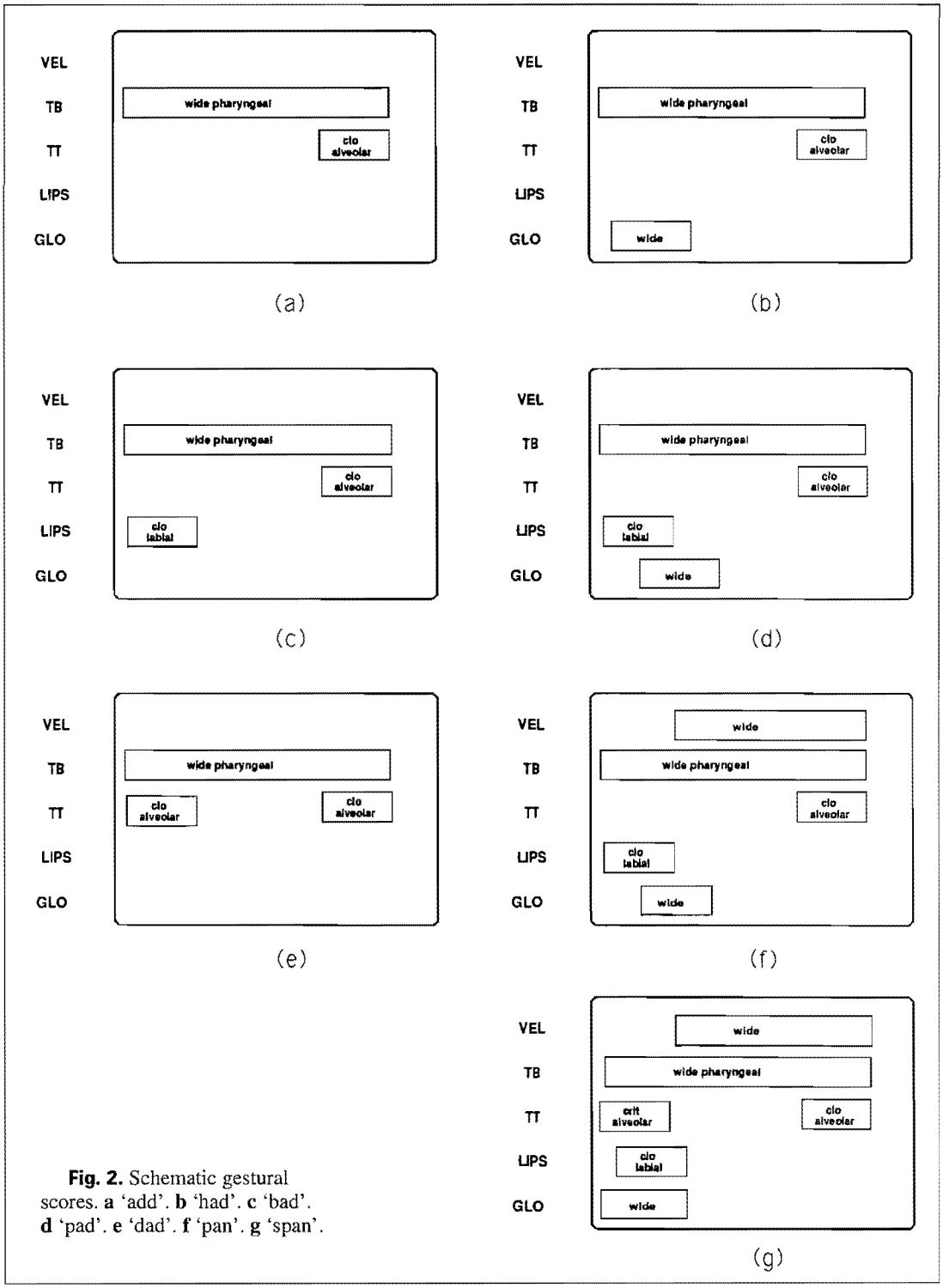

Fig. 2. Schematic gestural scores. **a** 'add'. **b** 'had'. **c** 'bad'. **d** 'pad'. **e** 'dad'. **f** 'pan'. **g** 'span'.

'dad', fig. 2c, e). All these differences are inherently categorically distinct.

Gestures can also differ parametrically, i. e., in the values of the dynamical parameters that define the spatiotemporal structure of the articulatory event, such as a target value for the tongue tip constriction degree that would lead to a complete closure vs. a critical value that would lead to the generation of turbulence (see gestures on TT tier in fig. 2g, e). While such differences are not inherently categorical, we have suggested [Browman and Goldstein, 1991] that distinct ranges of the possible parameter value space (for a given articulator set) will tend to be selected by a language on the basis of quantal articulatory-acoustic relations [e. g., Stevens, 1989] and/or on the basis of adaptive dispersion principles [e. g., Lindblom et al., 1983; Lindblom and Maddieson, 1988; Lindblom and Engstrand, 1989; Diehl, 1989]. In addition to target values for constriction degree, other dynamical parameters serve to distinguish gestures as well, as discussed in Browman and Goldstein [1989, 1990a]: constriction location target, stiffness (possibly, vowels vs. glides), and damping (possibly, flaps vs. stops, in languages where they contrast).

Another major function of a phonological description is to represent natural classes. Since gestures are embedded in the vocal tract, the vocal tract itself acts to organize the gestures into a hierarchical articulatory geometry [Browman and Goldstein, 1989], the levels of which have been shown to represent natural classes by work in feature geometry [e. g., Sagey, 1986]. The major organizational difference between this articulatory geometry and various feature geometries has been that, in the gestural approach, constriction degree (the closest gestural analog to continuancy) is low on the tree, in effect depending from the articulator node and sister to constriction location (place), whereas in feature geometries, continuancy has typically been close to the top of the feature tree. Recent work in feature geometry, however, has begun to lower the position of continuancy or its analogs such as aperture [e. g., Clements, in press]. Indeed, based on generalizations about the phonological behavior of assimilations in a variety of languages, Padgett [1991] proposes that continuancy should be represented as depending from the articulator node, a proposal consistent with the gestural approach. Such a move of course supports the relevance of the gestural unit to the organization of phonological feature geometry.

For the velic and laryngeal subsystems, featural descriptions can sometimes appear very similar to gestural descriptions. Featural descriptions of the velic and laryngeal subsystems usually contain the constriction degree of the particular articulator as an inherent aspect; in these cases, they are very close to a gestural description (for example, [+nasal] corresponds to a velic opening gesture). However, even for the velic and laryngeal subsystems, there are situations in which a featural and a gestural analysis differ. For general discussions of distinctions in voicing and aspiration in the gestural framework, see Browman and Goldstein [1986] and Goldstein and Browman [1986]; this latter paper is part of an exchange with Keating [e. g., 1984, 1990] about the viability of featural and gestural accounts of various voicing phenomena. For a gestural analysis of the category of Hindi stop variously called 'voiced aspirated', 'breathy voiced', or 'murmured', see Schiefer [1989], who compared a gestural account of these stops with a featural account in which the category is treated as a sequence of features [Dixit, 1987; also see Keating, 1990]. Schiefer demonstrated that the sequential differences in these stops fall out naturally within the gestural framework, in which the breathy voice is realized with a single glottal gesture, timed

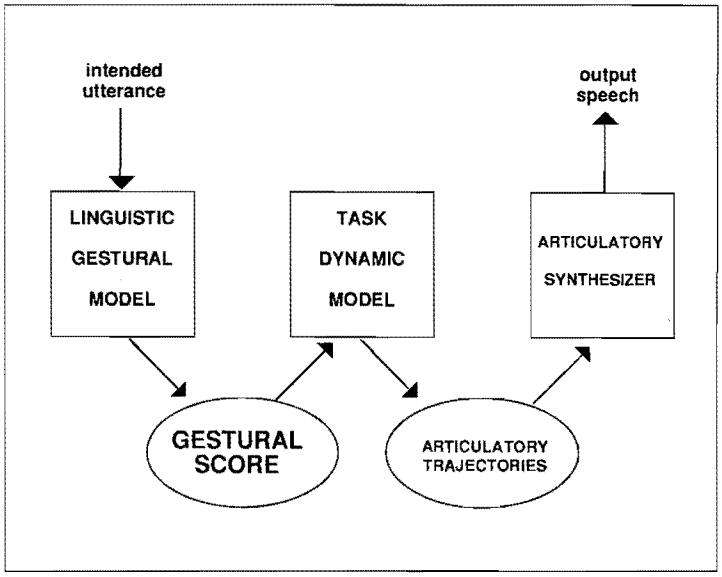

Fig. 3. Gestural computational model.

comparatively late. Since gestures have an extent in time, and describe movements that change in amount of openness at different points during their realization, all the acoustic changes can be accounted for by this single glottal gesture (and its timing with respect to other gestures).

1.2. Gestural Constellations: Combinations of Overlapping Gestures

As characterizations of physical events, gestures occur in space and over time. This has several implications. Since gestures have internal duration, they can overlap with each other; and since gestures are physical events, they are affected by physical processes occurring during the act of talking. In this section, we will focus on structure – how gestural overlap is used distinctively. Later sections will focus on process – how gestures vary in the act of talking.

The gestures that are employed in a given utterance are organized, or coordinated, into larger structures. We view the organization formed by those particular gestures as constituting the phonological structure of that utterance (or at least part of this structure). Of course, not every utterance in a language has an individual organization – there are general principles that define how classes of gestures are organized, or *phased.* These principles capture the syntagmatic aspect of a language's phonological structure, while the inventory of gestures that can participate in these organizations captures its paradigmatic aspect.

In the linguistic gestural component of the computational model currently being developed at Haskins Laboratories (fig. 3), a first approximation of these phasing principles is used to coordinate the gestures with one another [Browman and Goldstein, 1990b]. This gestural phasing results in a structure called a *gestural score.* A gestural score for the word 'palm' (pronounced [pʰam]) can be seen in fig. 4. This representation displays the duration of the individual gestures as well as the overlap among the gestures. The horizontal extent of a given box indicates the discrete

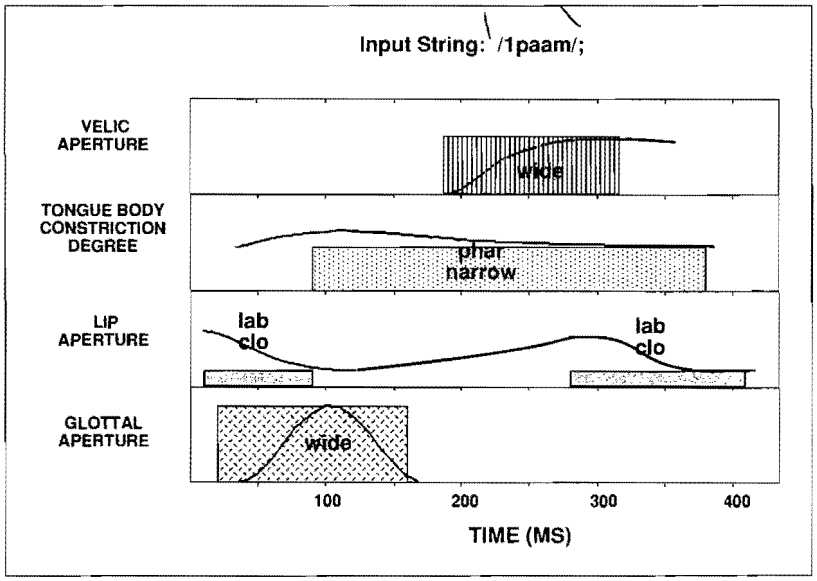

Fig. 4. Gestural score for the utterance 'palm' (pronounced [pʰam]), with boxes and tract variable motions as generated by the computational model. The input is specified in ARPAbet, so IPA /pam/ = ARPAbet /paam/. The boxes indicate gestural activation, and the curves, the generated tract variable movements. Within each panel, the height of the box indicates the targeted degree of opening (aperture) for the relevant constriction: the higher the box (or curve), the greater the amount of opening.

interval of time during which its particular set of values for the dynamic parameters is active. Given overlap, this means that several different gestures – sets of values – can be actively affecting the vocal tract at any particular instant in time. For example, in fig. 4, at time 50 ms, both the labial closure gesture and glottal gestures are active; by approximately time 125 ms, the labial closure gesture is no longer active but the tongue body narrow pharyngeal constriction has been activated for the vowel, so that at that point in time the glottal gesture and tongue body gesture are both active. Thus, with overlap the overall state of the vocal tract is dependent on more than one gesture. Articulatory phonology uses 'tube geometry' to characterize the patterns arising from overlapping combinations of gestures. As proposed by Browman and Goldstein [1989] and further developed by Bird [1990], tube geometry represents the constriction degree effects at each level of the vocal tract (when viewed as a set of linked tubes), and in this way forms the basis for natural classes that have been defined using features such as [sonorant].

As currently implemented in the computational model, the phasing statements coordinate pairs of gestures by specifying a particular dynamically defined point in each gesture that is to be synchronized. A very restricted set of points is used, for consonants generally the achievement of the target or the beginning of movement away from the target, and occasionally the onset of movement towards the target. The importance of these or similar points has been noted by others. For example, Huffman [1990, p. 78] suggested that closure onset and offset are among those 'landmarks... [that] serve as the organizational pivots for articulatory coordination'. Krakow [1989] observed regularities in the timing of the movements of the velum and lower lip with regard to these points (to be further discussed in Section 2.2). Finally, both Kingston [1985, 1990] and Stevens [in press] have emphasized the importance of related points, but defined in the acoustic domain.

Notice (in fig. 4) that gestural scores provide an inherently underspecified representation [e.g. Browman and Goldstein, 1989], in that not every tract variable is specified at every point in time. This is most akin to the restricted underspecification argued for by Clements [1987] and Steriade [1987], among others. Notice also that gestural scores are exclusively tier-based. Hierarchical units such as syllables are currently generally represented by the mechanism of associations (phasing) among individual gestures rather than by hierarchical nodes. The only hierarchical unit for which we currently have evidence is that of the oral gestures in a (syllable-initial) consonant cluster [Browman and Goldstein, 1988]. In these clusters, the oral gestures overlap only minimally rather than maximally, as typically happens when gestures from different articulatory subsystems co-occur (e.g. the oral and glottal gestures in fig. 4).

Much of the richness of phonological structure, in the gestural framework, lies in the patterns of how gestures are coordinated in time with respect to one another. We have used the term *constellations* to refer to such gestural coordinations without prejudging the correspondence between the constellations and traditional units of phonological structure (e.g., segments, syllables). Utterances comprised of the same gestures may contrast with one another in how the gestures are organized, i.e. the same gestures can form different constellations. Contrasts between nasal and prenasalized stops or between post-aspirated and pre-aspirated stops are possible examples of this kind. Considering only pairwise combinations of gestures with a similar extent in time, Browman and Goldstein [1991] have proposed that possible contrasts in organization for these gestures are restricted to three distinct types of temporal overlap: minimal overlap, partial overlap, and complete overlap.

Gestural organization is constrained in more specific, language-dependent ways as well. For example, Browman and Goldstein [1986] proposed two organizational principles governing glottal opening-and-closing gestures occurring in word-initial onsets (for a least a subset of Germanic languages, including English): (1) that glottal peak opening is synchronized to the midpoint of any fricative gestures, and otherwise to the release of any closure gestures (following Yoshioka et al. 1981) and (2) there is at most a single glottal gesture word-initially. [See Kingston, 1990, for a discussion of these issues from a different perspective, and Goldstein, 1990, for an evaluation of the problems therein.] Given these generalizations, word-initial 'sp' and 'p' are both presumed to have a single glottal gesture, as shown in fig. 2f, g (rather than the two glottal gestures for 'sp' expected from a segmental analysis [see e.g. Saltzman and Munhall, 1989]. The (allophonic) difference in aspiration between 'sp' and 'p' then follows automatically from timing principle (1) combined with the fact that gestures are events with temporal extent.

The fact that gestures are events with temporal extent can also eliminate the need for certain phonological adjacency constraints, which can often be seen to follow directly from gestural overlap. For example, much work in feature geometry [e.g., Sagey, 1986; McCarthy, 1988; Clements, in press] constrains assimilation to be the spreading of a feature to an adjacent slot, rather than the replacement of one feature by another. From the point of view of gestural overlap, many cases of 'assimilation' or apparent 'coarticulatory' feature-spreading follow directly from the fact that several gestures are co-occurring, either lexically or through later concatenation or sliding (this will be discussed further in Section 2 and 3) [see also Fowler, 1980; Bell-Berti and Harris, 1981, 1982; Gelfer et al.,

1989; Boyce et al., 1990; Boyce, 1990]. As these authors have also emphasized, there is no need to spread a feature, since gestures already have an inherent extent in time. A related constraint, that 'total place assimilation in consonants will be restricted to immediately adjacent consonants' [Clements, in press, p. 29], also follows directly from gestural overlap. Zsiga [in preparation] discusses a number of cases in which overlap can account for various phonological phenomena (as well as some problem areas for a gestural account). In general, the existence of gestural overlap means that a number of phonological constraints follow automatically rather than having to be stipulated.

The general style of coordination (or phasing) between gestures may also vary from language to language. Smith [1988, 1991] has provided acoustic and articulatory evidence that temporal patterns in Italian and Japanese are affected differently by the change of an intervocalic consonant from singleton to geminate, and Dunn [1990] has found similar evidence in a comparison of Italian and Finnish. Smith found that, in Italian, little effect on the timing of the vowels was seen when consonants differed between singleton and geminate, but in Japanese, the intervowel organization was significantly altered. Such results are consistent with a gestural organization for Italian in which the vocalic gestures are directly phased with each other, and for Japanese in which vocalic gestures are phased only indirectly, by being phased with respect to the intervening consonantal gestures. In turn, such different coordination types are consistent with the characterization of Japanese as mora-timed [e.g., Han, 1962; Port, Dalby & O'Dell, 1987] and Italian as syllable- (or possibly stress-) timed [e. g., Farnetani & Kori, 1986]. The gestural account of such 'rhythmic' differences as being due to a difference in direct or indirect coordination of vowels not only provides a potentially explanatory account of phonological differences, but predicts such phonetic detail as whether the vowels are shortened as intervening consonants are added (or lengthened).

2. Contrast and Allophonic Variation

We often refer to a gestural analysis as an analysis of the 'input', and more traditional analyses as analyses of the 'output', where input and output refer to descriptions of the (local) articulatory gestural organization and resulting global vocal tract shape/acoustics, respectively. Traditional segmental analyses are descriptions of the combined effects of the (overlapping) gestures in a gestural constellation, and therefore are typically descriptions of the acoustics and therefore the 'output', in our terminology. Even featural descriptions often refer to attributes of segments, and are therefore often 'output' descriptions. This is the source of the differences in description between the gestural approach, on the one hand, and segments and/or features on the other hand. An example of the descriptive differences has already been alluded to, re the voicing and aspiration issue [Keating, 1984, 1990; Goldstein and Browman, 1986; Browman and Goldstein, 1986; Schiefer, 1989]. In this section, we will present a number of examples of gestural analyses of cases that have traditionally been analyzed in segmental and/or featural terms as different kinds of allophonic variation, showing that the gestural analyses capture a wider range of behavior, and do so by using general principles rather than special category-changing rules. At the same time, the underlying 'input' structures capture contrast in a simple fashion.

Traditionally, the complement to contrast has been seen as identity. That is, two primitive phonological units either contrast or they

are considered to be identical. Where this identity is at odds with phoneticians'/phonologists' percept of speech, this led historically to positing a single underlying phonemic (or phonological) unit, with distinct allophonic units in a more narrow phonetic representation [Anderson, 1974]. The same phoneme is 'spelled' as categorically distinct allophones in different environments. However, when articulatory gestures are used as phonological primitives, much of the variation that was traditionally captured by a distribution of distinct allophonic units can, instead, be captured either by quantitative variation in the 'input' parametric specification of a given gesture, or as a direct 'output' consequence of overlap of invariant gestural units.

Generalizations: there are cases in which a gestural analysis reveals generalizations that have been missed in traditional allophonic descriptions. For example there are cases in which two very different allophonic rules (when couched in terms of segments and features) must be posited to describe what is quantitative variation in one and the same gesture in the same contexts. Further, there are cases in which particular prosodic contexts (e. g. stress and syllable positions) show a very similar influence on gestures of different types (oral and laryngeal, for example), or on their organization. We will discuss such cases below.

Relation between allophonic and other variation: there is much systematic, quantitative variation of speech gestures that has never been captured in a narrow allophonic transcription of the conventional sort, and could not be easily described in this way [e. g., differences in the magnitude and duration of stop consonant gestures in different prosodic environments – Browman and Goldstein, 1985; Kelso et al., 1985; Munhall et al., 1985]. As will be argued below, there is no principled difference between this kind of variation and the kind that has been annotated in a narrow transcription. In fact, we will examine cases in which the same parameter of variation has been treated as allophonic in some contexts and (implicitly) as quantitative in others. Moreover, as others have argued [e. g., Pierrehumbert, 1990; Sproat and Fujimura, 1989], this intermediate allophonic representation does not contribute in a useful way to the complete description of the variability. It is either unnecessary, or gets in the way of stating the generalizations. Thus, it seems that many allophonic differences are just quantitative differences that are large enough that phoneticians/phonologists have been able to notice them, and to relate them to distinctive differences found in other languages.

In this section, then, we will see that the very same syntagmatic organization will give rise to superficially different kinds of 'output' variation such as 'coarticulation' and allophonic differences, depending on the nature of the particular gestures in the organization (Sections 2.1, 2.2). In addition, we will see that general patterns of quantitative variation in gestural parameters can also give rise to a variety of superficially unrelated 'output' consequences (Section 2.3).

2.1. 'Coarticulation' of Consonants and Vowels

In the phasing rules that are currently implemented in our model, oral constriction gestures are divided into two functional classes: vocalic and consonantal [Browman and Goldstein, 1990b]. The distinction reflects the intrinsic differences between the two classes of gestures in their dynamical parameters. The consonantal gestures typically have a greater degree of constriction and a shorter time constant (higher stiffness) than the vocalic gestures. Syllable-sized organizations are defined by phasing (oral) consonant and vowel gestures with respect to one another. The basic re-

lationship is that initial consonants are coordinated with vowel gesture onset, and final consonants with vowel gesture offset (the specific points being coordinated also differ in the two cases). This results in organizations in which there is substantial temporal overlap between movements associated with vowel and consonant gestures, as was seen in the gestural score of fig. 4.

When the same (invariant) consonant gesture is coproduced with different overlapping vowel gestures (e. g., in [ada] vs. [idi]), the articulator motions produced by the task dynamic model will differ, reflecting the vowel gestures' demands on the articulators that they share in common with the consonant. As discussed in Saltzman and Munhall [1989], the nature of this variation produced by the model will differ depending on whether the overlapping gestures are defined with respect to the same or distinct tract variables. In the case of distinct tract variables (e. g., TT for [d] and TB for vowels), the consonant gesture will achieve its invariantly specified tract variable (TT) target regardless of what vowel is overlapping, although the particular contribution of articulators used to achieve this target (jaw, tongue body, tongue tip) will differ depending on the vowel. Thus, the overall shape of the vocal tract produced during the tongue tip closure will differ in [ada] and [idi]. As shown in Saltzman and Munhall [1989], this difference corresponds to that seen in Ohman's [1967] X-rays. The different articulatory trajectories will produce different acoustic formant frequency transitions for the two stops, but apparently no difference in the consonant's percept [Liberman et al., 1967; Fowler, 1980; Liberman and Mattingly, 1985].

In the case where consonants and vowels share the same (TB) tract variables (e. g., the consonant [g] as in [aga] or [igi]), the consonant and vowel gestures cannot both simultaneously achieve their targets, since they are attempting to move exactly the same structures to different positions. As a result, the location (but not degree) of constriction achieved for the consonant will vary as a function of the overlapping vowel [Saltzman and Munhall, 1989]. Again, this is consistent with the X-ray data of Ohman [1967]. In this case, however, the difference is perceptible (at least to phoneticians), and has sometimes been represented by distinct 'front' and 'back' allophones.

These examples of consonant/vowel overlap illustrate two important points about gestural structures. First, they show how, as invariantly specified phonological units, gestures can give rise to context-dependent articulatory and acoustic trajectories, without having to posit any 'implementation rules' for converting specific invariant (phonological) units into variable (physical) parameters. The variation follows directly from the definition of the units as parameterized task-dynamical systems, their phonological organization (pattern of overlap), and the general principles of how overlapping units blend. The same gestural structures simultaneously characterize phonological properties of the utterance (contrastive units and syntagmatic organization) and physical properties. Second, this example suggests how the very same syntagmatic structure (pattern of overlap) can yield different kinds of variation (allophonic vs. just 'articulatory-acoustic') as a function of the particular gestures involved – in particular, whether those gestures use the same or different articulator sets.

2. 2. Higher-Level Units in Velic and Oral Subsystems

Recently, the differing intergestural organization found in different (syllable) positions has been investigated in detail for two different gestural constellations in English: nasal consonants [Krakow, 1989] and /l/ [Sproat and Fujimura, submitted]. Both are constellations

comprising two gestures: a nasal consonant includes oral constriction and velic lowering gestures; /l/ includes tongue tip constriction and tongue body retraction gestures. Comparison of the data from these two papers reveals important similarities in how gestural organization varies as a function of position, despite differences in the traditional descriptions. For nasals, the traditional account characterizes syllable position differences by spreading the relevant feature ([nasal]) to the preceding vowel in the syllable-final case [e.g., Ladefoged, 1982], while for /l/, the position differences in certain dialects of English are handled by positing different allophones ['clear' vs. 'dark', differing in the feature [back], e.g., Keating, 1985] in initial and final position. However, as we saw with consonant-vowel overlap, this turns out to be an example in which the syntagmataic organization of the gestures is the same in these two cases, an aspect missed by the allophonic and featural descriptions.

Krakow's [1989] results show a clear difference in coordination between word-initial nasals (e.g.,'see more') and word-final nasals (e.g.,'seem ore'). In the word-initial case, the end of the velum lowering movement is roughly synchronous with the end of the lip-closing movement. The gestures appear to be phased so that the effective achievement of their targets coincide. For the word-final case, however, the end of velum lowering occurs substantially earlier (100–350 ms) than the end of lip movement. In fact, the end of velum lowering appears to coincide with the *beginning* of the lip-closing movement in this case. Syllable-position effects are similar to these word-position effects.

Sproat and Fujimura [submitted] found that the tongue body retraction (TB) and tongue tip raising (TT) movements for English /l/ also differ in their coordination as a function of word position. In word-initial position (e.g., 'B. Likkovsky'), the extremum of the TB movement follows the TT extremum slightly, while in the word-final position (e.g., 'Beel, equate') the TB extremum occurs substantially earlier than that for TT. Sproat and Fujimura manipulated the strength of the prosodic boundary following noninitial /l/, from none (e.g., 'Beelik') to an intonation break (e.g., 'Beel, equate'), and concluded that there is continuous variation in the relative timing of the two movements as a function of the boundary strength. However, examination of the relative timing data for prevocalic /l/ shows that, in general, truly word-final /l/s show the pattern with TB leading (with the magnitude of the lead affected by the strength of the following syntactic boundary), while the non-word-final cases (initial, medial, and medial before morphological boundaries) show either simultaneity or a slight lagging of TB.

There is an apparent similarity, then, in behavior of the gestures forming the constellations for nasals and /l/. Both constellations exhibit changes in relative timing as a function of word (or possibly syllable) position. In both cases, nonfinal position shows the gestures more nearly synchronous than in final position, and in both cases, it is the gesture with the narrower oral constriction (lip closure for the nasals, TT raising for /l/) that lags substantially in final position. In the case of the nasals, there is evidence for a specific shift in phasing: the end of velum lowering is coordinated with the end of lip closing for initials, but the beginning of the lip closing movement for finals.

It would strengthen the parallelism if evidence for such a shift also existed for /l/. Sproat and Fujimura did not examine this directly, although there is some indirect evidence in their data for such a shift. In final position, the TB gesture offset (as measured by movement extrema) precedes the TT gesture offset substantially. If the TB gesture off-

set were, in fact, being coordinated with TT gesture onset, as the analogy with the nasal behavior would predict, then as the TT movement increases in duration (e.g., before different boundaries), the measured offset-to-offset lag between the gestures should increase proportionally. Sproat and Fujimura measured the acoustic duration of the preboundary rime (which presumably is related to the acoustic duration of the /l/, and hence to the movement duration of TT); a clear correlation between this duration and the offset-to-offset lag for final /l/ can be observed in their figure 3. This parallels a correlation between lip closure duration and offset-to-offset lag found by Krakow [1989] for the final nasals. Moreover, the points in Sproat und Fujimura's figure corresponding to nonfinal /l/ show TT leading, and do not appear to show any correlation between the magnitude of the TT lead and /l/ duration. This lack of correlation with duration would be expected if the offsets were being coordinated in this case, and such a lack of correlation is also found for nonfinal nasals.

The parallelism of nasals and /l/ reveals organizational patterns that are similar across subsystems and correlated with position in the word (or syllable). Viewing these behaviors gesturally suggests a (speculative) possible wider generalization, namely that there is a single syllable-final organizational pattern in which the wider constrictions always precede narrower constrictions [reminiscent of the sonority hierarchy; cf. also the related hypotheses of Sproat and Fujimura, submitted, and Mattingly, 1981]. The same pattern would then be invoked for the (vocalic) tongue body and (consonantal) tongue tip gestures in 'add', the two /l/-related tongue gestures in syllable-final /l/, and the velic and lip (or tongue) gestures in syllable-final nasals. Parallelism between the velic and oral subsystems has been noted elsewhere as well. For example, Browman [1992] showed how, if syllable-final vowel nasalization were treated as a long velic gesture, then similarities in behavior between syllable-final nasals and long oral gestures, i.e. geminates, on a gating task [Lahiri and Marslen-Wilson, 1992] could be explained.

The similarities across subsystems revealed in these studies are generalizations only in a gestural approach, and not in the more traditional analyses of these variations as being different in kind (in the nasal and /l/ example, as feature-spreading and different feature values, respectively). While the articulatory and acoustic consequences differ depending on the particular gestures involved, in a gestural approach these consequences do not need to be explicitly controlled, as they are automatic consequences of the syntagmatic organization and the particular gestures involved.

2.3. Glottal Gestures: Positional (and Other) Variants

We have seen in previous subsections how what is traditionally described as contextual or allophonic variation can result automatically from the fact of overlap between invariant gestural units (e.g., overlap between consonants and vowels), or from differences in the characteristic patterns of overlap of gestures in syllable-initial and -final positions. In addition, some kinds of allophonic variation can be shown to result from quantitative variation in a gesture's dynamic parameters as a function of prosodic variables such as stress and position. Gestures shrink in space and in time in some contexts. This latter kind of variation is quite constrained – it scales the metric properties of a gestural event, but does not alter the composition of articulatory components out of which it is assembled.

2.3.1. Aspiration in English. A relevant example involves voiceless stops in English. Traditionally, these have been described as

having aspirated and unaspirated allophones in different environments. Kahn [1976], for example, defines the environment that selects the aspirated allophone as 'exclusively syllable-initial', with the unaspirated allophone occurring elsewhere. Kahn's rule assigns the feature [+spread glottis] in these aspirated environments, with [–spread glottis] generally being used for unaspirated allophones. This distinction is not an accurate characterization of the aspiration differences in English; nor is it either accurate or desirable to use a categorical rule to describe the aspiration of stops in English.

In many of the environments in which the output appears to be unaspirated, there is in fact a glottal opening-and-closing gesture present in the input. That is, presence or absence of aspiration in the output is generally not a discrete function of whether or not the glottis is spread, but rather is either a function of the timing of the glottal gesture with an associated oral gesture or a (gradient) function of the magnitude of the glottal gesture. The first cause of lack of aspiration in the output occurs in initial [s]-stop clusters, as mentioned in Section 1, in which lack of aspiration automatically results from the pattern of overlap among the contrastive gestures. As noted previously, English has a constraint that at most one glottal opening (spreading) gesture can occur in word-initial position. When this single gesture is associated with a fricative gesture, whether as a singleton or as a member of a sequence of oral gestures, the peak glottal opening is phased to the middle of the fricative gesture (probably its peak displacement). In the case of an [s]-stop cluster, this means that the glottis is already narrowed by the time the stop is released, which results in a 'short lag' in the onset of voicing following release (VOT). This is the basis for the description of stops in such clusters as voiceless unaspirated [Lisker and Abramson, 1964].

The second cause of lack of aspiration in the output is the gradient reduction of glottal magnitude due to differences in stress and position. In analyses such as Kahn's, stress and position allophones are represented categorically. Voiceless stops are unaspirated in word-medial position before unstressed vowels (e. g. 'rapid') because they are 'ambisyllabic' rather than exclusively syllable-initial and therefore are represented as [–spread glottis]. However, voiceless stops are aspirated ([+spread glottis]) in the same position when before stressed vowels because they are considered to be syllable-initial. Single stops in word-initial position before either stressed or unstressed vowels are also aspirated and represented as [+spread glottis]. This categorical approach to aspiration is not supported by a recent study by Cooper [1991], who used transillumination to measure glottal aperture in four environments: initial vs. medial, before stressed and unstressed vowels.

Examining these four environments in two-syllable reiterant speech utterances (/pipip/, /titit/, and /kikik/), Cooper found, first of all, that there was a glottal spreading gesture in all four environments, contrary to the prediction that the unaspirated environment is [–spread]. Secondly, he found effects of both stress and word position on the magnitude of the glottal spreading gesture (in both space and time), with initial position and stress favoring larger gestures. Thus, the medial unstressed position showed the smallest glottal spreading gesture overall. From a gestural point of view, there is nothing special or categorically different about the medial unstressed case – it is simply the environment that shows the most gestural reduction because of the combined effect of stress and position. In an analysis such as Kahn's in which the medial unstressed case is viewed as an allophone categorically distinct from the form occurring in the other three environments, one would expect to observe qual-

itatively distinct laryngeal behavior in the medial unstressed case. This expectation is not borne out by Cooper's data. A weaker prediction of the categorical view is that there should be a robust interaction between stress and position factors, such that stress has a large effect medially, but little or no effect initially. This weaker prediction is also not borne out – the utterances with /t/ and /k/ generally show no interaction at all (although an interaction is observed for /p/). Cooper's own conclusion, based on additional experiments not summarized here, is that stress and word position, rather than syllable structure and aspiration category, are the relevant variables that regulate laryngeal behavior of voiceless consonants in English.

Voicelessness in final position differs from that in other positions. In final position (word or possibly syllable), a glottal spreading gesture for English stops is usually not observed [e.g., Lisker and Baer, 1984]. However, the muscular activity normally associated with spreading gestures (increased activity of the posterior crico-arytenoid muscles, suppression of the interarytenoid) is found for such final stops in Lisker and Baer's data [and also in Hirose and Gay, 1972], although reduced in magnitude. This is consistent with a gestural reduction analysis: final position represents the most extreme case of reduction. However, analysis of final position is complicated by the fact that a constriction of the false (ventricular) folds is sometimes observed in this position [Fujimura and Sawashima, 1971; Manuel and Vatikiotis-Bateson, 1988]. It is presumably this constriction that led Kahn to posit yet a third allophone for voiceless stops ([+constricted glottis]) in final position. Since the relation between this constriction and the muscular control of the glottis (proper) has not been explicitly investigated, it is not clear how to relate this constriction to the glottal spreading gesture.

2.3.2. Aspiration and 'h'. As reported above, positional and stress allophones of English voiceless stops result from quantitative variation in gesture magnitude (with the possible exception of the final ventricular constriction). Since the unit of reduction is the gesture, the gestural analysis predicts that similar patterns of reduction should be found, regardless of whether they have been analyzed as a segment ('h') or a feature ([+spread]). Pierrehumbert and Talkin [1992] have recently measured amount of reduction in glottal abduction for 'h' in various prosodic contexts, using acoustic analysis to estimate the actual abduction. While most of their focus was on more global prosodic structure (phrasal accent and intonation boundaries), they also found reduction effects due to word stress and position generally similar to those found by Cooper [1991] (although as noted above, Cooper's data show some degree of influence of the supralaryngeal gesture on the laryngeal gesture). In a nongestural approach, the similarity in behavior of 'h' and [+spread] is not captured, since unlike aspiration in stops, the variation in 'h' is not usually represented at all, even by distinct allophonic units (except where the reduction is so extreme that it is sometimes analyzed as deleted, for example in 'vehicle'). In a gestural approach, however, the same reduction process gives rise to both kinds of variation.

There is also a symmetry in final position between voiceless stops and 'h' in English. In final position, glottal spreading gestures are reduced to the limiting case of no observable opening. This is exactly the environment in which 'h' does not occur in English. In a gestural framework, this distributional fact follows from the facts of reduction noted in voiceless stops. That is, words cannot have a contrastive glottal spreading gesture in final position, because such gestures are reduced to zero in final position, regardless of whether

the glottal spreading gesture co-occurs with an oral closure or not [contrast between final voiced and voiceless stops is possible only because this contrast involves other differences such as vowel length – Lisker, 1974 – which can themselves be analyzed as overlap differences between consonant and vowel gestures, Fujimura, 1981]. In more traditional approaches, this relationship between the distribution of 'h' and the allophones of voiceless stops is not captured.

2.3.3. Generalizations across Glottal and Oral Gestures. If the variation in the glottal gesture due to position and stress is in fact due to a general process, then such variation should be observed in other gestures occurring in similar environments. Similarities in the behavior of glottal and oral movements due to position and stress differences have indeed been observed.

The behavior of tongue tip movements is known to be affected by stress and position. For example, flapping of alveolar closures in English tends to occur in medial unstressed environments [Kahn, 1976], where we have seen that there is also substantial reduction in glottal spreading. If we assume that a flap is a reduced tongue tip closure gesture, reduced in time and possibly also in displacement, then the tongue tip and glottal gestures are behaving similarly. Apparent counter-examples are the medial unstressed alveolar stops that have not been considered to be flaps (e.g., in 'after'). Since glottal gesture reduction applies in 'after' – the 't' isn't aspirated – one would expect a reduced alveolar gesture here as well. However, these cases can be handled very nicely when input and output descriptions are properly distinguished. Although the alveolar in 'after' is not considered to be a flap, it is possible that the alveolar closure *is* reduced in this context (input), but that the percept of a flap (output) depends on having an open vocal tract both before and after the reduced tongue tip movement. This analysis is related to that of Banner-Inouye [1989], who analyzes flapping in English autosegmentally as resulting from spreading of 'open aperture' ([-cons]), from either side onto the timing slot associated with a coronal consonant. The phenomenon of flapping is thus analyzed by her as a short (single timing slot) open-closed-open contour that results from spreading in English. In the gestural framework, the reduction (making the movement short) would occur regardless of what other gestures are involved, but the description (or percept) of the resulting structure as a flap would depend on an open-closed-open acoustic contour (i.e., the structure in 'butter' but not 'after'). That is, the reduction process would always reduce the oral gesture in this environment, but the contour that is perceived as a flap would simply be one of the possible output consequences, depending on the appropriate set of gestures.

There are also potential parallels between glottal spreading and tongue tip closure gestures in final position. As we shall see in the next section, final alveolar closure gestures are subject to a variable amount of reduction in final position, including the failure to achieve any tongue tip contact. This is, of course, reminiscent of the frequent failure to see any actual glottal opening finally. When such reduced final alveolars coincide with the ventricular constriction discussed above, this produces the structure that has traditionally been described as the glottal stop [ʔ] allophone of /t/. The confluence of these events can be seen in the fibroscopic and palatographic data of Manuel and Vatikiotis-Bateson [1988].

Other oral constriction gestures also exhibit patterns of reduction similar to those exhibited by the glottal spreading and alveolar closure gestures. For example, bilabial closure gestures show effects of stress [e.g.,

Kelso et al., 1985; Beckman et al., 1992] and stress/position [Browman and Goldstein, 1985; Smith et al., submitted], similar to those shown by glottal gestures. These papers show substantial reduction of labial gestures in non-initial reduced syllables (initial reduced syllables were not examined). Thus, the reduction processes associated with stress and position in English for glottal gestures appear to be general, operating on tongue tip and labial gestures occurring in the same environments. Note again that while the variation in the dynamics of the tongue tip gesture has been represented as allophonic, the variation in the lip gesture has not been. Yet both seem to be instances of a very general reduction process, one that also operates on glottal gestures.

In addition to looking at similarities in the environments in which different kinds of reduction occur, it is possible to focus on the form of the reduction itself, as observed in the dynamic properties of the gestures. Munhall et al. [1985] have demonstrated similarities in the velocity profiles of movements of the glottis and the tongue dorsum (in /k/). In addition, the quantitative changes in the kinematic properties (i.e. displacement, duration, peak velocity) for different stress conditions were shown to be similar for the tongue dorsum and glottal movements.

In summary, allophonic variation associated with prosodic variables such as position and stress has been shown, in many cases, to be a constrained quantitative and gradient variation, rather than a categorical variation. Viewing such variants as gradient changes within a gestural framework captures similarities in behavior across position and stress and across different featural and segmental characterizations of glottal spreading gestures, and also captures similarities in behavior across different articulatory subsystems.

3. Variation during the Act of Talking

In this section, we examine some of the consequences of using the gestural approach to analyze phonological and phonetic variation that can be attributed to processes occurring during the physical act of talking. This variation arises from two interlocking sources, one gradient and one categorical. Beginning with a contrastive canonical gestural structure, processes occurring during the act of talking will cause gradient changes that can ultimately be perceived as a categorically different gestural structure. This is due, among other things, to the fact that the acoustic (as well as articulatory) consequences of a given invariantly specified gesture will differ depending on what other gestures are concurrently active [Browman and Goldstein, 1990a, b]. The following examples will show how the constrained processes available in the gestural view provide a unified and explanatory view of a variety of superficially different kinds of phonetic and phonological alternations.

3.1. Speech Production Errors; Connected Speech

One aspect of the act of talking that appears to be well handled by a gestural account is that of speech production errors. Mowrey and MacKay [1990] recorded muscle activity for [l] during experimentally induced speech errors in tongue twisters such as 'Bob flew by Bligh Bay'. In one session, about a third of the 150 tokens showed anomalous muscle activity, such as insertion of [l] activity in 'Bob' or 'Bay' and diminution of [l] activity in 'flew' or 'Bligh'. Only five of these tokens, however, involved all-or-none behavior; most of the activity was gradient. That is, the magnitude of activity in both the inserted and 'original' [l] fell on a continuum. Some of the errors were small enough so that they were not audible. The timing of the inserted activity was, how-

ever, localized and consistent. Such errors, in which the positioning (organization) is categorical but the magnitude is gradient, can be handled very naturally in a gestural framework.

Another aspect of the act of talking that is well handled in the gestural framework involves alternations that occur in connected speech. As shown in some of the data summarized below, in connected speech the patterns of gestural overlap may vary. In particular, factors associated with increased fluency (e.g., increased rate, more informal style) result in increasing the temporal overlap among gestures. Additionally, prosodic boundaries may influence the degree of overlap between neighboring gestures that belong to successive words. We have hypothesized that this kind of variation can result in changes that have traditionally been described as 'fast speech' alternations of various sorts, and have presented articulatory evidence for this [Browman and Goldstein, 1990a, b]. However, it is important to note that such gestural sliding is endemic in talking [e.g., Hardcastle, 1985], and not limited to the cases that have been noted as alternations. Thus, this is another situation (like those discussed in Section 2), in which some, but not others, of the results of a single gradient process have been noted in phonetic transcriptions. In a gestural account, a single generalization (increase in overlap) characterizes all these cases.

Evidence for increased overlap as rate increases has been presented for consonant and vowel gestures [Gay, 1981; Engstrand, 1988] and for the laryngeal gestures for two voiceless consonants in contiguous words [Munhall and Löfqvist, 1992]. Hardcastle [1985] has presented evidence for variation in gestural overlap as a function of prosodic boundary strength as well as rate. Using electropalatography, he measured overlap in time between the dorsal closure for /k/ and the onset of the tip/blade contact for a following /l/. The /kl/ sequences employed included word-initial clusters and examples in which the /k/ and /l/ were separated by various boundaries (syllable, word, clause, and sentence). Sentences were read at fast and slow rates. In general, the amount of overlap was consistently greater at the fast rate than at the slow rate. The effect was observed in all phonological and syntactic contexts, but was largest at the clause and sentence boundaries. Here, slow rates often showed long 'separation' intervals between the gestures (rather than overlap), while fast rates tended to show considerable overlap, often greater than that seen in the within-word or within-phrase cases. Thus, both rate and prosodic boundaries influence gestural overlap.

In this example, variation in gestural overlap did not produce changes that have been described as connected speech alternations. However, we have proposed [Browman and Goldstein, 1990b] that there are circumstances in which increased overlap would result in such alternations. One such circumstance we refer to as gestural 'hiding'. This occurs when gestures employing distinct tract variables (cf. Section 2.1) increase their overlap to such an extent that even though all the relevant constrictions are formed, one of them may be acoustically (and perceptually) hidden by another overlapping gesture (or gestures). X-ray evidence for this hiding analysis was provided in Browman and Goldstein [1990b]. For example, two productions of the sequence 'perfect memory' were analyzed, one produced as part of a word list (and thus with an intonation boundary between the two words), the other produced as part of a fluent phrase. In the fluent phrase version, the final [t] of 'perfect' was not audible, and it would be conventionally analyzed as an example of alveolar stop deletion in clusters [e.g., Guy, 1980].

However, the articulator movements suggested that the alveolar closure gesture (for the [t]) still occurred in the fluent version, with much the same magnitude as in the word list version that had a clearly audible final [t]. The difference was that in the fluent version, the alveolar closure was completely overlapped by other stop gestures – the closure portion by the preceding velar closure ([k]), the release portion by the following labial closure (for the [m]). Thus, from the point of view of an articulatory phonology, all the phonetic units (gestures) were present in both versions. The difference between the list and fluent forms was due to variation in the gradient details of overlap, a process for which there is independent evidence. In other contexts, for example when a velic lowering gesture co-occurred with the hidden gesture, hiding produced apparent assimilations, rather than deletions. Thus, in the phrase 'seven plus' produced at a fast rate, the final consonant of 'seven' was audibly [m], but evidence for an alveolar closure was still present. Only a single gesture was hidden (the oral alveolar closure gesture) and not a segment-sized constellation of gestures. It is precisely this fact that leads to the percept of assimilation rather than deletion in this kind of example.

In analyzing casual speech alternations as resulting from gestural overlap, we were led to make the strong hypothesis [Browman and Goldstein, 1990b] that *all* examples of fluent speech alternations are due to two gradient modifications to gestural structure during the act of talking – (a) increase in overlap and (b) decrease in gesture magnitude. (The latter modification is related to the gestural modifications as a function of prosodic structure discussed in Section 2). A typical example of magnitude reduction might be the pronunciation of the medial (velar) consonant in 'cookie' as a fricative rather than as a stop [Brown, 1977]. Under this hypothesis, casual speech variation is quite constrained: all the lexical phonological units are present, though they may be decreased in magnitude and overlapped by other gestures. Gestures are never changed into other gestures, nor are gestures added.

3.2. Assimilation of Final Alveolars

A related hypothesis has been proposed by Nolan [1992], based on analyses of apparent assimilations of single final alveolar stops to following labial and velar stops (e.g., /t/ → [k] in '...late calls...'). Using electropalatographic contact patterns, he found that the final alveolars were present, but reduced in degree to a variable extent, in the forms that were perceived as assimilated [see also Barry, 1985; Kerswill, 1985, for examples of such 'residual' tongue tip gestures]. Moreover, even in cases in which no alveolar electropalatographic contact was observed, the assimilated forms were perceptually distinguishable from forms with no lexical alveolar stop gesture at all (e.g., assimilated 'bed' vs. 'beg'). These findings led Nolan to propose that 'differences in lexical phonological form will always result in distinct articulatory gestures.' From the point of view of articulatory phonology, this constraint follows quite naturally – the phonological form *is* an organization of gestural events.

Nolan's experiments on the class of final alveolar assimilations focussed on the role played by the reduction of the tongue tip gesture. In addition to reduction, however, the overlap between that gesture (reduced or not) and the following stop gesture may play a role in perceived assimilations. The role of overlap in the acoustics and perception of similar assimilations was investigated by Byrd [1992]. Using the computational gestural model discussed in Section 1.2, Byrd generated utterances with a continuum of overlap for each of the phrases 'bad ban' and 'bab dan' by

systematically varying the overlap between the alveolar and bilabial closure gestures. She found an asymmetry between the perceptions of the gestures in word-final position. When the first word ended in [d], the word-final alveolar was perceived as being assimilated to the following [b] when overlap increased substantially. However, with the same amount of overlap, the word-final [b] was not assimilated, and in fact, the following word-initial [d] in such cases tended to be perceived as being assimilated to the [b]. (An asymmetry in the same direction, although less extreme, was found when subjects listened to the first word extracted.) Byrd related this perceptual asymmetry in favor of the labial closure to the VC and CV formant transitions produced by synchronous (overlapping) labial and alveolar closure gestures. In general, such formant transitions were more similar to those produced by labial stops alone than those produced by alveolar stops alone. Thus, the effect of overlap tended to obscure final alveolars but not final labials. This could contribute to the tendency in English for final alveolar stops (but not final labials or velars) to assimilate to following stops [Gimson, 1962].

The simulation results of Byrd suggest that formant frequency transitions into final alveolar stops should vary as a function of the following stop (as long as they are at least partially overlapping). This hypothesized acoustic 'context effect' was confirmed in an investigation of natural speech by Zsiga and Byrd [1990]. They examined formant frequency transitions into the medial closure in phrases like 'bad pick', 'bad tick', and 'bad kick' produced at different rates. The major finding was that formant transitions shifted away from those expected for an alveolar stop towards those expected for the following consonant – either a labial stop, as in 'bad pick', or a velar stop, as in 'bad kick'. In the case of the following labial, the effects on formant transitions agreed with those observed in Byrd's simulations of 'bed ban' in which the labial closure gesture overlapped the alveolar gesture – both F_2 and F_3 were lower at the offset of the first vowel for 'bad pick' than for 'bad tick'. The magnitude of these effects was generally smaller than that found in Byrd's complete synchrony condition, which is consistent with the fact that final alveolar consonants in this natural speech experiment were actually perceived as such and were not assimilated to the following labials or velars. In general, perceptual assimilation should occur only when the effects of gradient overlap and reduction exceed some perceptual threshold.

A second finding of Zsiga and Byrd's was that, for utterances where the second word in the phrase began with a velar stop (e.g., 'bad kick'), a systematic relation was observed between temporal and spectral properties as rate was varied. When rate variation resulted in a decrease in the total duration of the medial closure, there was also an increase in the velar effects seen in the formant transitions. This relation can be simply accounted for by assuming that these cases involve increased overlap between the tongue tip and tongue body gestures.

Finally, an ongoing experiment by A. Suprenant is explicitly testing the relative contributions of overlap and gestural magnitude to the percept of final stops. The experiment employs tokens of utterances like 'My po*t* *p*uddles' collected at the X-ray microbeam facility at the University of Wisconsin. These tokens show variation in both the magnitude of tongue tip raising for the final [t] in 'pot' and in the temporal overlap of that gesture and the lip closure gesture of the following word. Listeners are presented with these sentences in a speeded 'detection' task. Preliminary results suggest that detection of 't' is a function both of its magnitude and amount of overlap with the following consonant.

3.3. Reduced Syllable Deletion

Assimilations (and deletions) of stop consonants represent only one kind of fluent speech alternation. Another example that follows directly from changes in gestural overlap is deletion of schwa in reduced syllables. For example, in a word like 'beret', the vowel in the first syllable, either [ə^{r}] or [ɻ] may be apparently deleted in continuous speech, producing something transcribed as [bɹeɪ]. The tendency for deletion has been shown to be a 'graded' one, dependent on a number of contextual factors [e.g., Dalby, 1984]. We have demonstrated [Browman and Goldstein, 1990a] that the concomitant shift in syllabicity could be the perceptual consequence of an increase in overlap between the initial labial closure gesture and the tongue gestures for the 'r'. This was shown by using the computational gestural model to generate a continuum in which the degree of overlap or separation between the control regimes for the labial closure and the 'r' varied in small steps. In the canonical organization for 'beret', the labial and 'r'gestures did not overlap at all. This meant that the labial gesture was released before the 'r' was formed. This differed from the canonical organization for 'bray', in which the gestures were partially overlapping [like the velar and 'l' gesture in the clusters illustrated in Hardcastle, 1985]. When listening to items from the continuum in a forced-choice test, subjects responded with 'bray' to items in which labial and 'r' gestures overlapped, and 'beret' to items in which they did not overlap.

Thus it is possible to view reduced syllable deletion as resulting from an increase in gestural overlap in fluent speech. This treatment is attractive for two reasons. First, it treats deletion as resulting from the same general process that gives rise to other (superficially unrelated) alternations. Second, it leaves us with the claim that all phonetic units constituting a lexical item are still present in fluent speech; only the overlap has changed, in a predictably gradient way. This seems to be a more natural treatment than one which would assume that an important structural unit (a syllable) is suddenly and completely eliminated in fluent, connected speech.

Another important aspect of this treatment of reduced syllables is the fact that the lexical difference between 'bray' and 'beret' was modeled only in terms of the coordination of labial closure and 'r' gestures. There was no explicit tongue gesture for a schwa. This hypothesis was sufficient to generate gestural scores that produced speech with the appropriate perceptual properties for both 'bray' and 'beret'. In addition, the overlap of the vertical components of their articulatory trajectories was consistent with tokens of this distinction collected using the X-ray microbeam system at the University of Wisconsin [Browman and Goldstein, 1990a].

However, in another investigation of reduced syllables [Browman and Goldstein, 1992], data analysis and modeling revealed that an explicit tongue gesture for a schwa *was* required in utterances of the form ['pVpəpVp], although the target of the required gesture was completely colorless in that it was the average of the tongue body positions for all full vowels for that speaker. Therefore, at the very least, development of a more complete typology of the gestural structure of reduced syllables is needed, and is currently being pursued, to evaluate the phonological and morphological conditions for schwas of various kinds, both in English and other languages. With respect to deletion processes, however, we should note that even if there is a tongue gesture associated with a particular schwa, increase in overlap between consonants on either side of it could result in hiding that gesture. Thus, even if an active schwa gesture is required in a word like 'difficult', increase in overlap so that the labioden-

tal fricative and the velar stop partially overlap could result in hiding of this gesture.

In summary, increase in overlap among gestures in fluent speech is a general gradient process that can produce apparent (perceived) discrete alternations. The examples above were describable as consonant deletions, consonant assimilations, and vowel deletions; another possible example is that of epenthetic stops in English [e.g., Anderson, 1976; Ohala, 1974], as discussed in Browman and Goldstein [1990b]. However, the fact that stop epenthesis in words like 'tense' is not found in some dialects of English [South African: Fourakis, 1980] raises the larger issue of variability of fluent speech alternations across dialects and languages. That is, if the process of increase in overlap is a completely general property of talking, why does it create epenthetic stops in one dialect but not another? We have suggested [Browman and Goldstein, 1989] that such dialect/language differences may arise from differences in the canonical patterns of coordination in the different languages. Two kinds of coordination differences are relevant here. First, languages may differ in the amount of canonical overlap between two gestures. For example, sequences of stops in English are canonically partially overlapping [Catford, 1977], whereas sequences in Georgian, for example, are canonically non-overlapping, i.e., are released stops [Anderson, 1974]. We would expect that an amount of increase in overlap that produces hiding in English would not necessarily do so in a language such as Georgian. Second, two gestures may be directly phased with respect to one another in one language, but only indirectly phased in another language (as discussed in Section 1.2). It is possible that gestures that are directly phased will be more likely to retain their canonical organization in connected speech.

4. Developmental Data

Developmental studies show that a child's first words are stored and retrieved not as phonemes but as holistic patterns of 'articulatory routines' [e.g., Fry, 1966; Ferguson and Farwell, 1975; Locke, 1983; Studdert-Kennedy, 1987; Vihman, 1991]. Recent research has suggested that the basic units of these articulatory routines are discrete gestures that emerge prelinguistically (during babbling), and which can be seen as early 'gross' versions of the gestures that adults use [e.g., Browman and Goldstein, 1989; Studdert-Kennedy, 1987; Studdert-Kennedy and Goodell, in press]. Further development can be viewed as differentiation (in terms of parameter values), and coordination of these basic gestures. For example, other recent studies [Nittrouer et al., 1989; Fowler et al., 1991] have shown that coordination into segment-sized units (one kind of constellation) only appears gradually during the course of language acquisition, which not only supports the contention that phonemes are not present in a child's first words, but also suggests that higher-level units are formed out of smaller units during the course of language development. If so, then articulatory phonology would provide a very appropriate approach to child language, and its use would facilitate the study of language development both theoretically and methodologically, since both child and adult utterances can be described in terms of the same basic primitives of gestures.

Fowler et al. [1991] studied experimentally induced speech production errors in CVC utterances by children and adults, using phonetic transcriptions by trained listeners to indicate the existence of an error. The purpose of the study was to test the hypothesis that organization into phonological structures smaller than the level of the lexical item only appears gradually during the course of language learning. Fowler et al. [1991] found that younger chil-

dren were much more prone to blend features in their errors than were adults, as in the error 'bam till' from the utterance 'pam dill'. Adults were correspondingly more likely to retain higher-level organization, whether segmental or subsyllabic, that is to produce the error 'dam pill' from the utterance 'pam dill'. Thus, in this experiment with single-segment onsets, onset (or segment) exchanges increased with age (4- and 5-year-olds 33 %, 8-year-olds 44 %, and adults 74 %), while feature blends decreased (4- and 5-year-olds 33 %, 8-year-olds 18 %, and adults 8 %).

The results of Fowler et al. support the hypothesis that lexical organization intermediate between the levels of the feature (or gesture) and the word develops as part of learning the language. However, the results do not distinguish between a featural analysis and a gestural analysis. Another study, that of Studdert-Kennedy and Goodell [in press], supports the gesture as the unit out of which words are formed as the child develops language. This study focussed on another kind of 'error', the differences between the child's pronunciation and the canonical adult one. The utterances of a child in transition from babble to speech (91–106 weeks) were recorded. The errors in these utterances were argued to arise either from 'paradigmatic confusions among similar gestures ... or from syntagmatic difficulties in coordinating the gestures that form a particular word' [p. 20].

If gestures originate as prelinguistic units of action, and gradually develop into the units of contrast, as argued by Studdert-Kennedy [1987] and Browman and Goldstein [1989], then it is possible to see a continuity of development in language. If these gestures then serve as the primitives that are further coordinated in the language-learning process, such continuity includes higher-level phonological units as well as the fundamental contrastive units.

Acknowledgments

Our thanks to Elliot Saltzman and Michael Studdert-Kennedy for critiquing a preliminary version of this paper, and to Lisa Zsiga and Simon Levy for help in manuscript preparation. This work was supported by NSF grant BNS 8820099 and NIH grants HD-01994 and DC-00121 to Haskins Laboratories.

References

Anderson, S. R.: The organization of phonology (Academic Press, New York 1974).

Anderson, S. R.: Nasal consonants and the internal structure of segments. Language *52:* 326–344 (1976).

Banner-Inouye, S.: The flap as a contour segment. UCLA Work. Pap. Phonet. *72:* 40–81 (1989).

Barry, M.: A palatographic study of connected speech processes. Cam. Pap. Phonet. Ling. *4* (1985).

Beckman, M. E.; Edwards, J.; Fletcher, J.: Prosodic structure and tempo in a sonority model of articulatory dynamics; in Docherty, Ladd, Papers in laboratory phonology II, pp. 68–86 (Cambridge University Press, Cambridge 1992).

Bell-Berti, F.; Harris, K. S.: A temporal model of speech production. Phonetica *38:* 9–20 (1981).

Bell-Berti, F.; Harris, K. S.: Temporal patterns of coarticulation: lip rounding. J. acoust. Soc. Am. *71:* 449–454 (1982).

Bird, S.: Constraint-based phonology; PhD diss. Edinburgh 1990.

Bird, S.; Klein, E.: Phonological events. J. Ling. *26:* 33–56 (1990).

Boyce, S. E.: Coarticulatory organization for lip rounding in Turkish and English. J. acoust. Soc. Am. *88:* 2584–2595 (1990).

Boyce, S. E.; Krakow, R. A.; Bell-Berti, F.; Gelfer, C.: Converging sources of evidence for dissecting articulatory movements into core gestures. J. Phonet. *18:* 173–188 (1990).

Browman, C. P.: Comments on chapter 9; in Docherty, Ladd, Papers in laboratory phonology II, pp. 257–260 (Cambridge University Press, Cambridge 1992).

Browman, C. P.; Goldstein, L.: Dynamic modeling of phonetic structure; in Fromkin, Phonetic linguistics, pp. 35–53 (Academic Press, New York 1985).

Browman, C. P.; Goldstein, L: Towards an articulatory phonology. Phonol. *3:* 219–252 (1986).

Browman, C. P.; Goldstein, L.: Some notes on syllable structure in articulatory phonology. Phonetica *45:* 140–155 (1988).

Browman, C. P.; Goldstein, L.: Articulatory gestures as phonological units. Phonology *6:* 201–251 (1989).

Browman, C. P.; Goldstein, L.: Gestural specification using dynamically-defined articulatory structures, J. Phonet. *18:* 299–320 (1990a).

Browman, C. P.; Goldstein, L.: Tiers in articulatory phonology, with some implications for casual speech; in Kingston, Beckman, Papers in laboratory phonology I: between the grammar and physics of speech , pp. 341–376 (Cambridge University Press, Cambridge 1990b).

Browman, C. P.; Goldstein, L: Gestural structures: Distinctivenes, phonological processes, and historical change; in Mattingly, Studdert-Kennedy, Modularity and the motor theory of speech perception, pp. 313–338 (Erlbaum, Hillsdale 1991).

Browman, C. P.; Goldstein, L: 'Targetless' schwa: An articulatory analysis; in Docherty, Ladd, Papers in laboratory phonology II, pp. 26–56 (Cambridge University Press, Cambridge 1992).

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

Byrd, D.: Perception of assimilation in consonant clusters: a gestural model. Phonetica *49:* 1–24 (1992).

Catford, J. C.: Fundamental problems in phonetics (Indiana University Press, Bloomington 1977).

Clements, G. N.: Toward a substantive theory of feature specification. Proc. of NELS 18, vol. 1, pp. 79–89 (1987).

Clements, G. N.: Place of articulation in consonants and vowels: a unified theory; in Laks, Rialland, L'architecture et la géométrie des représentations phonologiques (Editions du CNRS, Paris, in press).

Cooper, A.: An articulatory account of aspiration in English; PhD diss., Yale University, New Haven (1991).

Dalby, J. M.: Phonetic structure of fast speech in American English; PhD diss., Indiana University, Bloomington (1984).

Diehl, R.: Remarks on Steven's quantal theory of speech. J. Phonet. *17:* 71–78 (1989).

Dixit, R. P.: Mechanisms for voicing and aspiration: Hindi and other languages compared. UCLA Work. Pap. Phonet. *67:* 49–102 (1987).

Dunn, M. H.: A phonetic study of syllable structure in Finnish and Italian. 26th Meet. Chicago Ling. Soc. (1990).

Engstrand, O.: Articulatory correlates of stress and speaking rate in Swedish VCV utterances. J. acoust. Soc. Am. *83:* 1863–1875 (1988).

Farnetani, E.; Kori, S.: Effects of syllable and word structure on segmental durations in spoken Italian. Speech Commun. *5:* 17–34 (1986).

Ferguson, C. A.; Farwell, C. B.: Words and sounds in early language acquisition. Language *51:* 419–439 (1975).

Fourakis, M. S.: A phonetic study of sonorant-fricative clusters in two dialects of English. Res. Inst. Phonet. Ind. Univ. *1:* 167–200 (1980).

Fowler, A.; Brady, S.; Curley, S.: The phoneme as an emergent structure; evidence from speech errors. SRCD Meet., 1991.

Fowler, C. A.: Coarticulation and theories of extrinsic timing control. J. Phonet. *8:* 113–133 (1980).

Fowler, C. A.; Rubin, P.; Remez, R. E.; Turvey, M. T.: Implications for speech production of a general theory of action; in Butterworth, Language production (Academic Press, New York 1980).

Fry, D. B.: The development of the phonological system in the normal and the deaf child; in Smith, Miller, The genesis of language: a psycholinguistic approach, pp. 187–206 (MIT Press, Cambridge 1966).

Fujimura, O.: Elementary gestures and temporal organization – What does an articulatory constraint mean?; in Myers, Laver, Anderson. The cognitive representation of speech, pp. 101–110 (North-Holland, Amsterdam 1981).

Fujimura, O.; Sawashima, M.: Consonant sequences and laryngeal control. Annu. Bull. Res. Inst. Logoped. Phoniat. Univ. Tokyo *5:* 1–6 (1971).

Gay, T.: Mechanisms in the control of speech rate. Phonetica *38:* 148–158 (1981).

Gelfer, C. E.; Bell-Berti, F.; Harris, K. S.: Determining the extent of coarticulation: effects of experimental design. J. acoust. Soc. Am. *86:* 2443–2445 (1989).

Gimson, A. C.: An introduction to the pronunciation of English (Arnold, London 1962).

Goldstein, L.; Browman, C. P.: Representation of voicing contrasts using articulatory gestures. J. Phonet. *14:* 339–342 (1986).

Guy, G. R.: Variation in the group and the individual: the case of final stop deletion; in Labov, Locating language in time and space, pp. 1–36 (Academic Press, New York 1980).

Han, M.: The feature of duration in Japanese. Study Sounds *10:* 65–80 (1962).

Hardcastle, W. J.: Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. Speech Commun. *4:* 247–263 (1985).

Hawkins, S.: An introduction to task dynamics; in Docherty, Ladd, Papers in laboratory phonology II, pp. 9–25 (Cambridge University Press, Cambridge 1992).

Hirose, H.; Gay, T.: The activity of the intrinsic laryngeal muscles in voicing control: electromyographic study. Phonetica *25:* 140–164 (1972).

Huffman, M. K.: Implementation of nasal: timing and articulatory landmarks. UCLA Work. Pap. Phonet. *75* (1990).

Kahn, D.: Syllable-based generalizations in English phonology. (Indiana Univ. Linguistics Club, Bloomington 1976).

Keating, P. A.: A phonetic and phonological representation of stop consonant voicing. Language *60:* 286–319 (1984).

Keating, P. A.: CV phonology, experimental phonetics, and coarticulation. UCLA Work. Pap. Phonet. *62:* 1–13 (1985).

Keating, P. A.: Phonetic representations in a generative grammar. J. Phonet. *18:* 321–334 (1990).

Kelso, J. A. S.; V.-Bateson, E.; Saltzman, E.; Kay, B.: A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling. J. acoust. Soc. Am. *77:* 266–280 (1985).

Kerswill, P. E.: A sociophonetic study of connected speech processes in Cambridge English: an outline and some results. Camb. Pap. Phonet. Exp. Ling. *4* (1985).

Kingston, J.: The phonetics and phonology of the timing of oral and glottal events; PhD diss., University of California, Berkeley (1985).

Kingston, J.: Articulatory binding; in Kingston, Beckman, Papers in laboratory phonology I: between the grammar and physics of speech, pp. 406–434 (Cambridge University Press, Cambridge 1990).

Krakow, R. A.: The articulatory organization of syllables: a kinematic analysis of labial and veliar gestures; PhD diss., Yale University, New Haven (1989).

Ladefoged, P.: A course in phonetics; 2nd ed. (Harcourt Brace, Jovanovich, New York 1982).

Lahiri, A.; Marslen-Wilson, W.: Lexical processing and phonological representation; in Docherty, Ladd, Papers in laboratory phonology II, pp. 229–254 (Cambridge University Press, Cambridge 1992).

Liberman, A. M.; Cooper, F. S.; Shankweiler, D. P.; Studdert-Kennedy, M.: Perception of the speech code. Psychol. Rev. *74:* 431–461 (1967).

Liberman, A. M.; Mattingly, I. G.: The motor theory of speech perception revised. Cognition *21:* 1–36 (1985).

Lindblom, B.; Engstrand, O.: In what sense is speech quantal? J. Phonet. *17:* 107–121 (1989).

Lindblom, B.; MacNeilage, P.; Studdert-Kennedy, M.: Self-organizing processes and the explanation of phonological universals; in Butterworth, Comrie, Dahl, Explanations of linguistic universals, pp. 181–203 (Mouton, The Hague 1983).

Lindblom, B.; Maddieson, I.: Phonetic universals in consonant systems; in Hyman, Li, Language, speech, and mind, pp. 62–78 (Routledge, London 1988).

Lisker, L.: On time and timing in speech; in Sebeok, trends in linguistics, vol. 12, pp. 2387–2418 (Mouton, The Hague 1974).

Lisker, L.; Abramson, A. S.: A cross-language study of voicing in initial stops: Acoustical measurements. Word *20:* 385–422 (1964).

Lisker, L.; Baer, T.: Laryngeal management at utterance-internal word boundary in American English. Lang. Speech *27:* 163–171 (1984).

Locke, J. L.: Phonological acquisition and change (Academic Press, New York 1983).

Manuel, S. Y.; V.-Bateson, E.: Oral and glottal gestures and acoustics of underlying */t/* in English. J. acoust. Soc. Am. *84:* S84 (1988).

Mattingly, I. G.: Phonetic representation and speech synthesis by rule; in Myers, Laver, Anderson, The cognitive representation of speech, pp. 415–420 (North-Holland, Amsterdam 1981).

McCarthy, J. J.: Feature geometry and dependency: a review. Phonetica *45:* 84–108 (1988).

Mowrey, R. A.; MacKay, I. R. A.: Phonological primitives: Electromyographic speech error evidence. J. acoust. Soc. Am. *88:* 1299–1312 (1990).

Munhall, K.; Löfqvist, L.: Gestural aggregation in speech: laryngeal gestures. J. Phonet. *20:* 111–126 (1992).

Munhall, K. G.; Ostry, D. J.; Parush, A.: Characteristics of velocity profiles of speech movements. J. Exp. Psychol.: Hum. Perc. Perf. *11:* (1985).

Nittrouer, S.; Studdert-Kennedy, M.; McGowan, R.: The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables as spoken by children and adults. J. Speech Hear. Res. *32:* 120–132 (1989).

Nolan, F.: The descriptive role of segments: Evidence from assimilation; in Docherty, Ladd, Papers in laboratory phonology II, pp. 261–280 (Cambridge University Press, Cambridge 1992).

Ohala, J. J.: Experimental historical phonology; in Anderson, Jones, Historical linguistics, pp. 353–389 (North Holland, Amsterdam 1974).

Ohman, S. E. G.: Numerical model of coarticulation. J. acoust. Soc. Am. *41:* 310–320 (1967).

Padgett, J.: Stricture in feature geometry; PhD diss., University of Massachusetts, Amherst (1991).

Pierrehumbert, J.: Phonological and phonetic representation. J. Phonet. *18:* 375–394 (1990).

Pierrehumbert, J.; Talkin, D.: Lenition of /h/ and glottal stop; in Docherty, Ladd, Papers in laboratory phonology II, pp. 90–117 (Cambridge University Press, Cambridge 1992).

Port, R. F.; Dalby, J.; O'Dell, M.: Evidence for mora timing in Japanese. J. acoust. Soc. Am. *81:* 1574–1585 (1987).

Sagey, E. C.: The representation of features and relations in non-linear phonology; PhD Diss., MIT, Cambridge (1986).

Saltzman, E.: Task dynamic coordination of the speech articulators: a preliminary model; in Heuer, Fromm, Experimental brain research series 15, pp. 129–144 (Springer, New York 1986).

Saltzman, E.; Kelso, J. A. S.: Skilled actions: a task dynamic approach. Psychol. Rev. *94:* 84–106 (1987).

Saltzman, E. L.: Munhall, K. G.: A dynamical approach to gestural patterning in speech production. Ecol. Psychol. *1:* 333–382 (1989).

Schiefer, L.: 'Voiced aspirated' or 'breathy voiced' and the case for articulatory phonology. Forschungsber. Inst. Phonet. sprachl. Kommun. Univ. München *27:* 257–278 (1989).

Smith, C.: A cross-linguistic contrast in consonant and vowel timing. J. acoust. Soc. Am. *86:* S84 (1988).

Smith, C.: The timing of vowel and consonant gestures in Italian and Japanese. Proc. 12th Int. Congr. Phonet. Sci. *4:* 234–237 (1991).

Smith, C.; Browman, C. P.; Mc Gowan, R.; Kay, B.: Extracting dynamic parameters from speech movement data (submitted).

Sproat, R.; Fujimura, O.: Articulatory evidence for the non-categoricalness of English /l/ allophones. LSA Ann. Meet., Washington 1989.

Sproat, R.; Fujimura, O.: Allophonic variation in English /l/ and its implications for phonetic implementation (submitted).

Steriade, D.: Redundant values. CLS *23:* 339–363 (1987).

Stevens, K. N.: Phonetic evidence for hierarchies of features; in Keating, Papers in Laboratory Phonology III (Cambridge University Press, Cambridge, in press).

Stevens, K. N.: On the quantal nature of speech. J. Phonet. *17:* 3–45 (1989).

Studdert-Kennedy, M.: The phoneme as a perceptuomotor structure; in Allport, MacKay, Prinz, Scheerer, Language perception and production, pp. 67–84 (Academic Press, London 1987).

Studdert-Kennedy, M.; Goodell, E. W.: Gestures, features and segments in early child speech; in de Gelder, Morais, Language and literacy: comparative approaches (MIT Press, Cambridge MA, in press).

Turvey, M. T.: Preliminaries to a theory of action with reference to vision; in Shaw, Bransford, Perceiving, acting and knowing: toward an ecological psychology (Erlbaum Associates, Hillsdale 1977).

Vihman, M. M.: Ontogeny of phonetic gestures: speech production; in Mattingly, Studdert-Kennedy, Modularity and the motor theory of speech perception, pp. 69–84 (Erlbaum, Hillsdale, NJ 1991).

Yoshioka, H.; Lofqvist, A.; Hirose, H.: Laryngeal adjustments in the production of consonant clusters and geminates in American English. J. acoust. Soc. Am. *70:* 1615–1623 (1981).

Zsiga, E. C.: Gradient rules in phonology and phonetics; Yale University (in preparation).

Zsiga, E. C.; Byrd, D.: Acoustic evidence for gestural overlap in consonant sequences. J. acoust. Soc. Am. *88:* S82 (1990).