Journal of Phonetics (1990) **18**, 299–320

Gestural specification using dynamically-defined articulatory structures

Catherine P. Browman and Louis Goldstein*

*Haskins Laboratories, 270 Crown Street, New Haven, CT 06511, U.S.A. *and Department of Linguistics, Yale University, Box 1504A, Yale Station, New Haven, CT 06520 U.S.A.*

The types of analyses required within a framework that uses dynamically-defined articulatory gestures as the primitive units of phonetic description are outlined. The overall approach is then exemplified using investigations into an overlap hypothesis of reduced vowels, specifically the hypothesis that the difference between the bisyllable "beret" and the monosyllable "bray" can be attributed to a difference in the overlap of the labial closure and tongue rhotic gestures. This overlap hypothesis is supported by a perceptual test of simulations of the utterances as well as by preliminary articulatory analyses of X-ray microbeam data, and leads to some possible overlap typologies for reduced syllables.

1. Introduction

In recent years, we have been pursuing an approach to phonetics and phonology that invokes dynamically-defined articulatory gestures as the basic units. In other papers we have outlined the theoretical motivation for this approach (Browman & Goldstein, 1986*a*), pursued some implications for historical change and casual speech (Browman & Goldstein, 1986*b*, 1987, 1990), discussed how distinctiveness could be captured within gestural structures (Goldstein & Browman, 1986; Browman & Goldstein, 1986*a*, 1990), and explored the relation between a phonology of articulatory gestures and other non-linear phonologies (Browman & Goldstein, 1989). One basic tenet in all these papers has been that much is missed when the line between phonological patterning and physical processes is drawn too firmly. The strong form of our view proposes that phonological structure resides in the organization of the physical actions involved in speaking. Thus, we call the approach we have been pursuing an "articulatory phonology".

However, we also think that characterizing speech in terms of dynamical gestures has much to offer regardless of one's hypotheses about the nature of phonology. Such a characterization uses a form of description that has proven useful in other domains of action (e.g., Cooke, 1980; Kelso, Holt, Rubin & Kugler, 1981; Kelso & Tuller, 1984*a*; Kugler & Turvey, 1987; Saltzman & Kelso, 1987), and thus relates speech activity to more general issues in motor behavior as well as drawing upon principles and techniques from this area. Moreover, it provides a framework that allows analytical and rigorous investigations of articulatory structure to be conducted in a way that makes direct contact with linguistic issues. From this

0095–4470/90/030299 + 22 $03.00/0

perspective, then, the gestural framework could be pursued from within different phonological approaches as a phonetics of dynamically-specified articulatory gestures.

Within the framework being developed, the basic units are dynamically-defined articulatory gestures. These gestures are coordinative structures (Turvey, 1977) modeled in terms of task dynamics (Saltzman, 1986; Saltzman & Kelso, 1987). Task dynamics captures two important properties of gestures. First, the gestures are defined in terms of speech *tasks,* the formation and release of various constrictions such as bilabial closure (for [b]). Such tasks typically involve the coordinated motions of several articulators rather than the independent motions of individual articulators (such as the lower lip, upper lip and jaw, in the example of [b]). Second, the gestures are defined in terms of the underlying *dynamics* that serve to characterize the motions. Such a dynamical description provides a representation that is itself time-free, and yet characterizes the articulatory movements through space and over time, as a function of the system's dynamical parameters. Thus, a dynamical description simplifies the relation between categorical and continuous characterizations of articulation, which is desirable from both a practical and a theoretical perspective. (Further discussions of the application of dynamics, and specifically task dynamics, to speech can be found in Fowler, Rubin, Remez & Turvey, 1980; Kelso & Tuller, 1984*b*; Browman & Goldstein, 1985; Ostry & Munhall, 1985; Saltzman & Kelso, 1987; Vatikiotis-Bateson, 1988; Hawkins, in press.)

To aid in making the gestural framework as rigorous and testable as possible, we are developing a computational system in conjunction with our colleagues Elliot Saltzman and Philip Rubin at Haskins Laboratories. Figure 1 portrays the components of the system: the linguistic gestural model specifies a gestural score (see below) given some input, the task dynamic model generates articulator movements given the gestural score, and the vocal tract model generates an acoustic signal given the articulator movements. The task dynamic and vocal tract models are

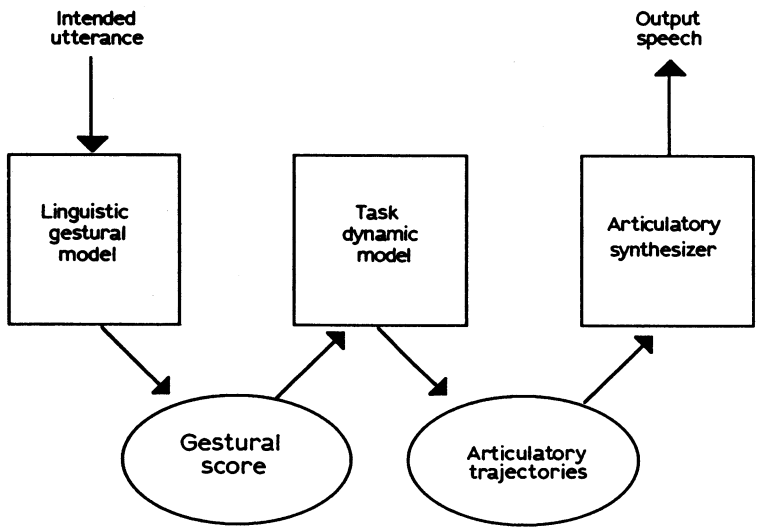

Figure 1. Gestural computational model.

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

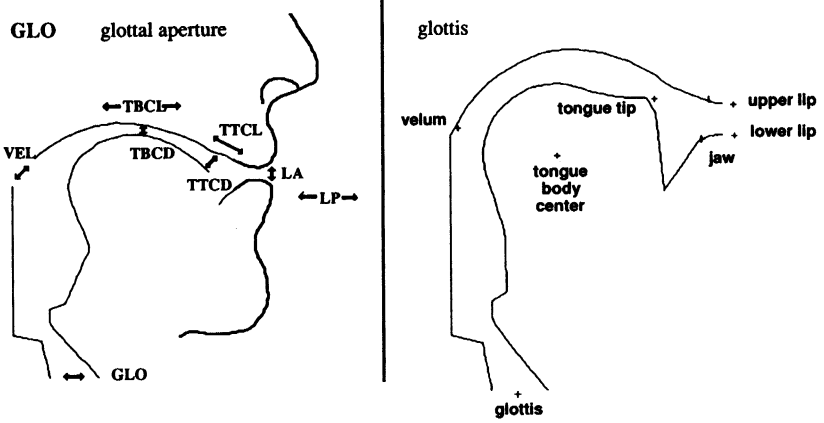

Figure 2. Tract variables and the articulators involved.

fairly completely described in Saltzman & Munhall (1989) and Rubin, Baer & Mermelstein (1981), respectively. The best description of the linguistic gestural model to date is in Browman & Goldstein (1987).

In the current computational model, there are eight variables, called tract variables, that can be used to specify speech tasks. These variables, and the articulators that are coordinated to achieve the speech task for each variable, are shown in Fig. 2, on the left and right respectively. Note that for oral gestures, the tract variables are paired, with a separate tract variable for each of the two dimensions of constriction formation: CL, the constriction location (i.e., the place along the wall of the oral cavity where the constriction is formed), and CD, the constriction degree (i.e., the size of the constriction). The gestures for an utterance are organized by phasing (and other) statements in the linguistic gestural model into a gestural score that contains the activation intervals (domain of active control) for each gesture, and the values of the dynamic parameters for each of the gesture's tract variables. (These parameters are identified and explained in Section 2 below.) Figure 3 shows a sample gestural score. Within each of the boxes in the figure, the values of the dynamical parameters are fixed, and serve to define the particular gesture in question. This gestural score is input to the task dynamic model, which uses the information about gestural activation and parameter values to generate the movements of the tract variables (shown superimposed on the gestural score in Fig. 3).

The gestural score serves as an *input* specification, from which the movement of the vocal tract articulators and the resulting acoustic *output* unfold in a lawful fashion, as currently simulated by the task dynamic and vocal tract models. This

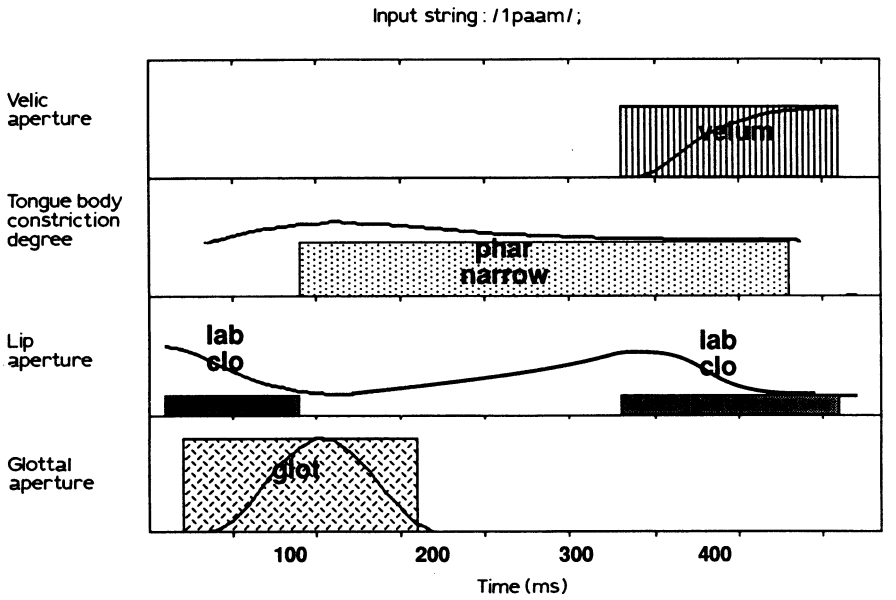

Figure 3. Gestural score and generated tract variable motions for "palm" (pronounced [pɑm]). The input is specified in ARPAbet, so IPA /pɑm/ = ARPAbet input string /paam/. The boxes indicate gestural activation, and the curves the generated tract variable movements. Within each panel, the height of the box indicates the targeted degree of opening (aperture) for the relevant constriction: the higher the box (or curve) the greater the amount of opening.

clear separation between input and output can do useful work. That is, one of the benefits of gestural specification is that certain acoustic (and perceptual) properties of utterances can be accounted for as by-products or side-effects of their gestural organization. An example of a possible "by-product" perceptual effect will be considered in some detail in Section 3, after gestural specification is discussed in Section 2. We will begin, however, by clarifying the distinction between input and output further.

1.1. *Input–output relations*

Consider the case of a single gesture in isolation. Within the model, each gesture is a dynamical control regime that regulates the formation (and/or release) of a characteristic constriction within the vocal tract. The gesture's dynamical parameters include a "target" (equilibrium position) specification of the values for the location and degree of a particular vocal tract constriction (e.g., bilabial). When the gesture's regime is active, the associated articulatory synergy (for a bilabial, the upper lip, lower lip and jaw) displays a characteristic time-varying response to this gestural input (calculated by the task dynamic model). Eventually, assuming the gesture is active long enough, the constriction targets are achieved. The articulator motion determines, in turn, the time-varying shape of the vocal tract and thus the acoustic output (computed by the vocal tract model). In this simple one-gesture universe, a given constriction will always yield the same acoustic output (ignoring the non-involved articulators), and thus the information captured by parameterizing

the input (constriction) will be in 1:1 relation to the output (acoustics). If the nature of multi-gestural structures in speech were such that the gestures were produced in strict, non-overlapping, sequence, the choice of input or output would still not make a lot of difference. By the end of a given gesture, roughly the same acoustics would always be achieved (although the path to get there would depend on the state of the system left by the preceding gesture).

In real speech, of course, gestures overlap in time: they are coproduced (Fowler, 1980). Thus, the acoustic output associated with a given gesture will vary as a function of other concurrently active gestures. The overlapping of invariant articulatory gestures can account for, among other things, the varying acoustic frequency characteristics of stop consonant bursts and formant transitions in the environment of different vowels (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967). A demonstration of how invariant articulations might yield varying acoustics is given in Öhman (1967), who shows midsagittal X-ray profiles in which the tip of the tongue reaches a relatively invariant target position for the consonant in /idi/, /ada/, and /udu/, while the shape of the tongue body and lips at the time of this consonant target are determined primarily by the vowel. That is, an invariant consonant gesture (produced with the tongue tip) is overlapping different neighboring vowel gestures (produced with the tongue body and lips). The acoustic properties of the stop closure and release will reflect the combined simultaneous effects of lips, tongue tip and tongue body gestures. This kind of overlapping production ("coproduction") has been viewed as the cause of other kinds of "coarticulation" (e.g., vowel-to-vowel effects—Fowler, 1981), and of allophonic variation (e.g., nasalization of vowels before nasal consonants—Krakow, 1989).

The above discussion touched on the utility of specifying phonetic information in terms of gestural input. Turning now to the role of the acoustic output, in our view the output associated with a set of gestures is relevant to the "tuning" of the parameter values associated with individual gestures and their organization into assemblies (Browman & Goldstein, 1989, 1990; Goldstein, 1989). This tuning occurs during the development of language within an individual talker, and is also relevant to establishing the patterns of contrastive gestures that languages come to employ. While the development of gross constriction gestures of the lips, tongue tip and tongue dorsum is a universal part of language development (as can be seen in babbling, e.g., Locke, 1983), the language-specific values of constriction location and degree associated with each of these gestures must be acquired by the child from listening to the acoustic output. (For example, the parameters of the tongue tip closure gesture are different in English and French, specifically in constriction location.) The child's job may be made easier by the fact that languages tend to favor certain patterns of parameter values for gestures, as well as certain patterns of gestural organization. Contrastive values for constriction location and degree tend to evolve in such a way that the acoustic properties associated with a given set of parameter values are relatively stable (Stevens, 1972, 1989) and tend to differ sufficiently from the parameter values for other contrasting gestures (Lindblom, MacNeilage & Studdert-Kennedy, 1983). Thus, there may well be systemic preferences for how gestures are parameterized that take into account their output, at least in ideal, careful speech contexts.

Output considerations do not, however, appear to constrain actively the processes of variation in speech production. Once the pattern of gestures for a given language

is acquired, we argue, variation with respect to speaking style and prosodic context follows from very general principles of gestural overlap and magnitude that are blind to their acoustic consequences. The extreme case of this blindness is when gestures increase their overlap to the extent that one becomes completely hidden by others. For example, as discussed in Browman & Goldstein (1987, 1989), increase in overlap between (invariant) input gestures can lead to apparent (i.e., acoustic and perceptual) deletions and assimilations in the output. One gesture can be acoustically hidden by other concurrent ones. An example presented in these papers is the deletion of the final /t/ in "perfect" when in the phrase "perfect memory". X-ray evidence showed that the alveolar closure for the /t/ was still being produced in the fluent phrase, even though its acoustic consequences were completely hidden by the preceding velar closure and the following bilabial closure. From the point of view of the acoustic output, such changes in the production of a word are drastic, deleting all the criterial output properties of segments. However, all the input gestures are present; only their organization has been changed. Moreover, this change is hypothesized to be a very general characteristic of casual, fluent speech—gestures tend to show increasing overlap. These acoustic changes, and many other superficially unrelated ones, follow automatically from the gestural structures and this principle of variation. Thus, by specifying gestural input, rather than acoustic output, certain types of variation can be accounted for in general, explanatory ways.

2. Gestural specification

In the preceding section, we argued that it is possible to gain insight into various phenomena by using a conceptual framework that describes phonetic structure in terms of overlapping input gestures (spatiotemporal articulatory units), and that distinguishes carefully between input and output. In many cases (as in the case of "perfect memory"), relevant observations can be made without detailed analyses of the dynamical characteristics of the gestures or their relationships. Nonetheless, to apply the framework more widely, and to see how (or whether) the quantitative aspects of the articulatory structure of speech and its variability can be accounted for, it is necessary to understand the details of exactly what is involved in specifying gestures using task dynamics.

A task dynamic specification of speech gestures is a constrained, reduced degree-of-freedom description compared with the continuous movement trajectories of multiple articulators that are observable. This can be seen in two ways, one relating to the notion of "task" and one to the dynamical aspects of the specification. First, the concept of the tract variable means that not all the articulators need to be analyzed individually, but rather articulatory actions can be combined into the linguistically significant task variables. Second, a dynamic description provides a way of characterizing all the points in a trajectory using only a few numbers (the values of the dynamic parameters).

Despite the constrained nature of task dynamics, it is often desirable to ease the analysis process by reducing the degrees of freedom further wherever possible. We do this by making simplifying assumptions (such as a constant damping ratio); the analysis of articulatory data in the light of these assumptions appears to lead to acceptable preliminary gestural specifications. We further assume that the results of

a given data analysis apply as generally as possible, and thus treat results of simple analyses as hypotheses to be tested for general applicability. These simplifying assumptions and hypothesized generalizations are sure to be wrong in many respects, and ultimately need to be challenged and modified as appropriate. Nevertheless, we will include them in the overview below so that the reader may better evaluate this approach to reducing the dimensionality of articulatory description. In addition the overview will include indications of some directions of research into gestural specification that seem particularly promising to us.

In Section 2.1, the dynamical control parameters will be described. In Section 2.2, we will look briefly at the coordination of gestures into larger assemblies or constellations in terms of relative phasing parameters. In Section 2.3, some possible types of prosodic gestural variation will be explored.

2.1. *Individual gestures*

2.1.1. *Dynamical specification*

Dynamical specification of a gesture requires first choosing the type of dynamical regime that will govern the motion of a particular tract variable (and its associated articulators). In the current formulation of the task dynamic model, the regime is always specified as a damped mass spring ("point attractor") model with constant mass, as shown in (1):

$$m\ddot{x} + b\dot{x} + k(x - x_0) = 0 \tag{1}$$

where

m = mass (currently fixed at 1.0 in the model)
b = damping
k = stiffness
$\ddot{x}$ = instantaneous tract variable acceleration
$\dot{x}$ = instantaneous tract variable velocity
x = instantaneous position of the tract variable
x_0 = rest position of the tract variable

Specification of a gesture further requires that the values of the dynamical parameters in (1) (for the appropriate tract variable or variables[1]) be set, and that the (temporal) domain of active control for the tract variable(s) be delimited. The three parameters whose values must be specified for each tract variable are (a) the rest position x_0, (b) the stiffness k and (c) the damping ratio $b/(2[mk]^{1/2})$, from which the damping b can be computed. Since oral gestures have CD and CL tract variables (see Fig. 2), these three parameters must be specified for both variables in an oral gesture. Each of the boxes in Fig. 3 is defined by constant values for each of these three parameters. We will briefly explain what the parameters mean, and present some of the issues that must be resolved when using dynamically characterized gestures.

[1] In addition to specifying the particular tract variables associated with the gesture, the relative contributions (called weights) of the associated articulators (see Fig. 2) must also be specified. These relative contributions hold only under "everything else being equal" conditions. The actual articulatory contributions will depend on the ensemble of concurrently active gestures. See Saltzman & Munhall (1989) for discussion.

(a) The *rest position,* x_0, is related to the notion of "target"—it determines the tract variable position towards which the system moves. The system approaches the rest, or target, value most closely, without overshoot, in the case of critical damping. For oral gestures, questions involving specification of x_0 are related to the familiar characterization of phonetic units in terms of manner and place: constriction degree (CD) and constriction location (CL) respectively (cf. Fant, 1960; Stevens & House, 1955; Stevens, 1989). The x_0 value for the gestures in the current model are based on available articulatory descriptions, tuned so that the vocal tract model output sounds appropriate.

(b) It is in questions of stiffness and damping ratio that a dynamic approach is distinguished from other approaches. *Stiffness, k,* determines (in conjunction with damping ratio) the durational characteristics of tract variable motion associated with the gesture: the stiffer the tract variable, the less time it will take to achieve its rest position, everything else being equal. In an undamped system, stiffness determines the frequency of oscillation. (It is important not to confuse the durational effects of stiffness with acoustic duration. As will be discussed in Section 2.3, stiffness is only one of several possible determinants of acoustic duration.) Stiffness does not have a long history as a phonetic descriptor; therefore, basic questions about it still need to be addressed. In the current formulation of the linguistic gestural model, only two underlying stiffness values are used, both derived from articulatory analyses—one for consonants and one for vowels. While this works quite well as a first approximation, further work is required to determine whether stiffness covaries with x_0, and whether the stiffnesses of the two related tract variables (CD and CL) should differ from each other. Another question concerning stiffness involves the possibility that it could form the basis for natural classes. For example, gestures for glides might differ from those for vowels primarily in their stiffnesses (glides being stiffer); similarly, gestures for stops (and affricates) might be stiffer than those for fricatives.

(c) *Damping ratio* determines what happens when the tract variable approaches its rest position—whether it overshoots this value (underdamping), approaches it as a limit without overshooting (critical damping), or never approaches it very closely at all (overdamping). In the current model, all non-laryngeal gestures are assumed to be critically damped. However, this assumption needs to be further investigated. As with stiffness, it is also of interest to ask if the damping ratio helps define phonetic natural classes. For example, it is possible that flaps might be less highly damped than other gestures.

2.1.2. *Parameter estimation*

In order to address the above issues, it is necessary to estimate the values of the parameters from analyses of observed articulatory data. However, in so doing, it is important to realize that such analyses can only provide approximations to the gestural specification since gestures are comparatively abstract—they are not the articulatory movements themselves, but rather the functions underlying the observed movements. In some cases, the relationship between the observed movements and the underlying gestural regimes will be particularly opaque, such as when two simultaneously active gestures are affecting the same tract variables, e.g., the velar closure gesture and the vowel gesture in "key". In this example, both gestures are defined in terms of TBCL and TBCD, and are also partially overlapping in time.

Thus, the observed tract variable motions will be affected by both gestures, making it difficult to separate out the contributions of the individual gestures. This suggests a strategy of first analyzing utterances in which coactive gestures involve distinct tract variables (e.g. bilabial consonants and unrounded vowels), and then generalizing to the more difficult cases. A further contributor to the difficulty of relating observed motions to underlying gestural regimes lies in the fact that the task dynamic framework currently provides no analytical procedure for dealing with the effects of physical mass and biomechanical constraints. However, bearing these caveats in mind, let us see how the various parameters can be estimated.

For x_0, a "target" value for location and degree of a constriction can, at least in principle, be estimated from examining articulatory data (such as lateral X-rays or X-ray microbeam data), no matter how difficult this is in practice. The stiffness and damping ratio of gestures must be determined by a mathematical analysis of articulatory movement data, preferably movement data from which it is possible to calculate an approximation to tract variable motion. If the damping ratio of the movements being analyzed is known (e.g., if there is reason to think that it is close to zero), then it is possible to estimate the stiffness as a function of the movement duration, or alternatively by the ratio of peak velocity to maximum displacement. The first technique has been employed by Browman & Goldstein (1985), the latter by Kelso, Vatikiotis-Bateson, Saltzman & Kay (1985), Vatikiotis-Bateson (1988), and Beckman, Edwards & Fletcher (in press), among others. If the damping ratio of the data is unknown, then it is possible to estimate parameter values for both stiffness and damping ratio using parameter estimation methods. One such method is used in a program currently under development at Haskins Laboratories (McGowan, Smith, Browman & Kay, 1988, 1990); this program assumes that a sequence of observed values was generated by a damped mass–spring system, and computes a least-squares estimate of the parameter values that could give rise to that sequence.

In order to employ these parameter estimation techniques, it is necessary to choose some stretch of time of a tract variable's motion during which the tract variable is assumed to be under active control of the dynamical system being fitted. This amounts to a hypothesis about the domain of active gestural control. Articulatory analyses have typically assumed that displacement extrema (in the articulator's or tract variable's time function) demarcate the edges of active gestural control (e.g., Kelso *et al.*, 1985; Vatikiotis-Bateson, 1988). Recently this assumption has begun to be questioned (see Browman & Goldstein, 1985; McGarr, Löfqvist & Story, submitted; Smith, Browman, McGowan & Kay, submitted). In the current formulation of the linguistic gestural model, active gestural control is bounded by the edges of comparatively flat displacement "plateaus" (with the plateaus themselves being "uncontrolled").

The question of the domain of gestural activity raises issues that must eventually be resolved in tandem with other specification issues. For example, while the current computational model uses step functions to indicate regions of gestural control, it has been suggested that ramped onsets and offsets should be used instead (Perrier, Abry & Keller, 1988). In such an approach, the generated shape of the tract variable motion would be influenced by the ramping function as well as the other parameters. Such additional degrees of freedom would be undesirable because of the additional complexity, but may prove necessary in the end to model tract

variable motions accurately. Another question about the domain of gestural activity is perhaps of more immediate linguistic interest. In the current model, the three superficially distinct components of a gesture—formation of constriction, a "holding" phase (see McGarr, Löfqvist & Story, submitted), and release—are generated using separate domains of activation. However, it might be preferable to use a single domain of activation to encompass them all (or some intermediate possibility). The choice here interacts with the choice of damping ratio, in that a single critically damped regime could not generate all three components. Ultimately, the solution to this may need to also consider a wider range of dynamical regime types. For example, a periodic attractor (see Abraham & Shaw, 1982) might be a more appropriate regime for combining constriction and release components.

2.2. *Gestural constellations*

Since a typical utterance consists of more than a single gesture, the relations among gestures must be characterized in addition to the individual gestures themselves. Ultimately some kind(s) of dynamical self-organizing principles may be found that would serve to narrow the range of coordinative possibilities to a few distinct modes (e.g., Kay, Kelso, Saltzman & Schöner, 1987; Turvey, Rosenblum, Kugler & Schmidt, 1986). Preliminary attempts to find distinct coordinative modes among speech gestures have shown similarities to other motor tasks. For example, Kelso, Saltzman & Tuller (1986) described a phase transition in the coordination of syllable-final bilabial closure and glottal opening gestures in /ip/. As repetition rate increased, there was an abrupt transition from syllable-final coordination to syllable-initial coordination that was similar to the kind of phase transition observed in repetitive finger-waggling (Haken, Kelso & Bunz, 1985). Such research may provide avenues for future characterizations of gestural relations. At present, however, we find it necessary to pursue simpler approximations to the question of gestural organization in order to get an empirical handle on it.

In the linguistic gestural component of the present computational system, gestural coordination is specified in terms of the relative phasing of gestures (Browman & Goldstein, 1987; see Kelso & Tuller, 1987, and Nittrouer, Munhall, Kelso, Tuller & Harris, 1988, for general discussions of using phasing in the specification of speech organization). Two gestures are coordinated by specifying two points, one in each gesture, that must coincide temporally. The points (in a gesture) are defined in terms of the phase of a "virtual" cycle whose duration (i.e., period) is determined by only the stiffness (natural frequency) ascribed to the gesture. In addition to phasing, this virtual cycle is used in the specification of gestural activation: a gesture is defined as beginning at the zero degree point of this virtual cycle, and it remains active until some later phase in the virtual cycle. Thus, as the stiffness of a gesture changes, the amount of time it is activated will automatically change as well as its temporal relation (as a whole) with other gestures, including those it is phased with respect to.

Given this general approach to specifying how the gestures constituting an utterance are coordinated with respect to each another, a number of issues arise. First, for a given pair of gestures that are to be coordinated, the actual phase of the synchronized points must be determined. Second, not every gesture in an utterance is phased with respect to every other gesture, so the sets of gestures to be

coordinated must be determined. Finally, in some cases phasing may occur with respect to larger gestural collectives rather than with respect to individual gestures. We now discuss these issues in turn.

The decision as to which phases to synchronize (between two gestures) can be made by examining movement data and observing what pattern of synchronization seems most characteristic (or most invariant) across multiple tokens of the particular gestural structure (see Tuller, Kelso & Harris, 1982, for an example of such an approach). An important question is whether there is a limited subset of points that languages use for phasing. In investigations to date using this model (see, for example, Browman & Goldstein, 1987), satisfactory results have been obtained by using only a few different points: the achievement of target and the onset of movement (either towards or away from a target), where points are based on intervals of active control defined using the edges of extrema plateaus rather than single extrema points. This is also consistent with the observations of Krakow (1989), who found that the phasing of the velum-lowering gesture for nasal consonants with respect to the oral constriction is "bistable"—it is phased either with respect to oral gesture onset or with respect to achievement of target (depending on syllable position). Much remains to be done on this important question, however, Moreover, it might ultimately be preferable to state an overall phasing between two gestures rather than to synchronize particular points. This could perhaps be done in terms of a coupling function between the two control regimes (e.g., Kay *et al.*, 1987).

A related issue is the choice of which gesture to coordinate with each other. Browman & Goldstein (1987) proposed that vocalic gestures are phased with respect to preceding (syllable-initial) consonantal gestures and that (syllable-final) consonantal gestures are phased with respect to preceding vocalic gestures. The phasing of V to C and C to V appears to work for English, at least as a first approximation, but the gestures that are coordinated may differ in different languages—for example, some languages may coordinate vocalic gestures directly. It is possible that the choice of gestures to be coordinated may be correlated with the prosodic nature of the language. For example, Smith (1988) has proposed (on the basis of acoustic evidence) that coordination might be C–V in languages such as Japanese that have been described as mora-timed but V–V in languages such as Italian that have been described as syllable-timed.

In the discussion so far, it has been assumed that individual gestures are phased with respect to other individual gestures. It may, however, be the case that some gestures are organized into larger collectives or constellations for the purposes of coordination. It is to be hoped that such constellations would correspond to linguistically significant units such as segments, syllable onsets, or syllables—but this need not be the case. Note that the investigation of the relation of gestures to linguistically significant units can proceed in two (related) ways: some particular unit can be assumed, and gestural correlates searched for; or articulatory movements can be parsed and gestural units established on the basis of criteria such as cohesiveness and variability.

An excellent example of a study proceeding from linguistic unit to articulation can be found in Sproat & Fujimura (1989), in which both light and dark /l/ were discovered to be complex segments—gestural constellations—differing primarily in the relative timing of the tongue tip and tongue body gestures. An example of a

study arriving at units beginning from an analysis of movement data can be found in Browman & Goldstein (1988), who suggested that gestures in the syllable onset form a unit for the purposes of coordination with the syllable nucleus, whereas coda gestures do not but rather are timed individually. Further questions can be posed relating, for example, syllable structure to variability in phasing and the amount of overlap between gestures, with tautosyllabic gestures expected to show more overlap and less variability than heterosyllabic gestures.

2.3. *Prosodic variation of gestures*

Thus far, the discussion has focused on how the canonical forms of utterances are specified using gestures. However, it is also important to understand how the proposed gestural structures may be flexibly adjusted to yield the differences in articulation (and acoustics) that are observed in different prosodic environments. Since the specification includes both the parameters of individual gestures and the phasing parameters that define intergestural organization, variation in either (or both) of these parameter sets is available as a way of characterizing prosodic differences.

One of the major aspects of prosodic variation involves temporal variation: for example, the acoustic duration of stressed syllables is typically longer than that of unstressed syllables, and phrase-final syllables are often lengthened acoustically relative to other syllables. In the dynamic approach, quantitative temporal information is provided not by specifying time directly but by specifying the parameters of the gestural regimes and their phasing. The pattern of relationships among the abstract dynamical coefficients of the model's equations automatically generates the quantitative temporal information as an inherent feature of the motion of the articulators. Thus, variation in acoustic duration could be a result of changes either in gestural parameters or in intergestural phasing. Consider the example "add". Changing the stiffness of the vocalic gesture for [æ] would (everything else being equal) change the amount of time required for the articulators to move to the configuration for [æ], and hence the acoustic duration associated with it. Changing the relative phasing between the vocalic gesture for [æ] and the following alveolar closure gesture for [d] would also change the acoustic duration of the [æ].

Given these two aspects of gestural specification, either of which will be associated with changes in acoustic duration, it is of interest to ask how these mechanisms are related to different types of prosodic variation. Stress, for example, has been associated with a decrease in the stiffness of the stressed gestures (Browman & Goldstein, 1985; Kelso *et al.*, 1985). However, McGarr *et al.* (submitted) have suggested that the difference between stressed and unstressed vocalic gestures lies in the duration of a comparatively steady state position, rather than in the movement towards this position as would be expected with stiffness changes. Such a pattern would be consistent with a change in the phasing between the vowel and the following consonant, with the longer steady state portion resulting from the consonant's delay (decreased overlap). Alternatively, or in addition, the steady state portion might involve an increase interval of active control of the vowel gesture itself, or of its "holding" phase (if separate).

Beckman *et al.* (in press) have investigated these various factors in English. Unlike previous studies, Beckman *et al.* explicitly compared the two types of

gestural variation. They analyzed jaw movements in syllables that were either phrase-final or non-phrase-final, and also either accented or unaccented, in order to investigate possible gestural correlates of both phrase-final lengthening and accentual lengthening. They found that the increases in acoustic duration associated with final lengthening and with accent were articulatorily quite different, and suggested that these differences might be modeled by the two types of gestural variation. Specifically, subjects slowed down the movement of the jaw into the phrase-final closure, without concomitant changes in amplitude of movement. However, accented jaw movements had larger amplitudes compared with unaccented movements, and the movement itself lasted longer but was not slower (i.e., did not have smaller peak velocity). They suggested that the phrase-final lengthening might be modeled by reducing the stiffness of the final closure (at least at normal speech rates), whereas the accent effect might be modeled by increasing the intergestural spacing (phasing) for the accented syllables as opposed to unaccented syllables.

It is difficult to compare the results of the various studies on stress and accent, since the criteria being used to identify the hypothesized domain of gestural control are often different, or not clearly laid out. More research is clearly needed to resolve the apparent conflicts. However, the possibility that two different dynamic mechanisms might be associated with two different prosodic phenomena is quite intriguing. That is, the dynamic characterization might reveal physical differences between different prosodic phenomena that are lost in a description based on acoustic duration, in which the consequences of the different dynamic mechanisms merge.

3. Reduced syllables

We are now in position to examine in more detail the kind of analyses that can be performed in a gestural framework. The primary example we will discuss in this section involves reduced syllables in English. The "vowel" portion of reduced syllables in English is difficult to characterize, and typically shows great contextual variation. For example, consider the first syllable in the word "beret". It is sometimes produced in such a way as to be transcribed with the vowel [ə] (schwa) preceding the [ɹ] of the following syllable. In other cases, it may be represented as a syllabic "r": [ɻ] or [ɚ]. Finally, in casual speech, it may cease to be syllabic altogether, the tendency to do so being a "graded" one, dependent on a number of contextual factors (e.g., Dalby, 1984).

This variation can be thought of in one of two ways. First, the speaker might be selecting one of the three distinct variants, with the choice of variants being a probabilistic function of the style of speaking and the local environment. Alternatively, the speaker might be maintaining a single input gestural structure, with the different variants resulting from the independently motivated casual speech processes—increase in overlap and reduction in magnitude (Browman & Goldstein, 1987). From this latter perspective, the different variants are not selected from the lexicon, but rather represent different acoustic and perceptual consequences of completely continuous variation in talking processes—increasing overlap and reduction. To see how this would work, a gestural specification for such reduced "nuclei" that could yield this kind of behavior needs to be identified.

A candidate gestural structure is one in which the nucleus of a reduced syllable

does not have any explicit vowel gesture associated with it. In this structure, reduced syllables would be characterized by an organization of the consonants preceding and following the nucleus such that the consonants show no overlap and thus produce an acoustic interval for the nucleus that is gesturally unspecified. (Note that this structure is the gestural equivalent of identifying a unit by a skeletal X-slot—timing information—but no melodic information.) According to this hypothesis, the shape of the vocal tract during the interval between the two non-overlapping gestures would be determined both by the positions in which the articulators were left by the preceding gesture and by the movement of the articulators to their own specific neutral positions when they are not involved in any active gesture.

Given such a gestural structure, utterances such as "beret" and "bray" each contain the same gestures. The distinction between them is in the organization of the initial consonant gestures (bilabial closure and rhotic). In "bray", the bilabial closure and rhotic gestures should be tightly (and closely) organized with respect to one another, and to the vowel, according to the C-center hypothesis (for syllable onsets) outlined in Browman & Goldstein (1988). In "beret", we hypothesize that the C-center organization does not hold for these gestures, but rather the bilabial closure is set off from the rhotic and vowel gestures, showing no overlap with them. In casual speech, then, it would be possible for the degree of overlap to increase to the value more usually seen for "bray". This would then be perceived as a vowel deletion, or a loss of syllabicity.

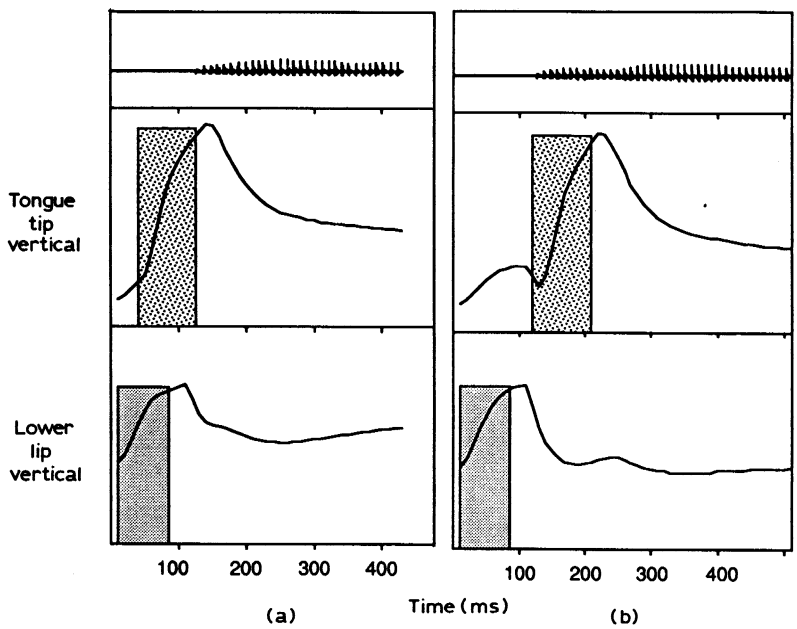

Figure 4. Gestural scores and articulator motions for the initial tongue tip rhotic and bilabial closure gestures in "beret", for the ends of the overlap continuum. To facilitate comparison with X-ray data, vertical motions of the articulators are displayed, rather than the generated tract variable motions. Thus, the higher the curve, the higher the articulator in space. Boxes indicate gestural activation. (a) Maximum overlap (40 ms); (b) maximum separation (40 ms).

To test the plausibility of this hypothesized difference in gestural structure, a series of gestural scores was constructed that differed only in the amount of overlap between the initial bilabial and rhotic gestures. If the gestural organizations for "bray" and "beret" differ only in overlap, then we would expect listeners' percepts to shift from one to the other when this parameter is varied. The scores were generated, not by analyzing articulations of "bray" and "beret", but by using the existing generalized statements in the linguistic gestural model (with one exception to be discussed below) that had been derived from other articulatory analyses. The statement specifying the phase relation between the bilabial closure gesture and the rhotic gestures was then manipulated to produce the stimulus series. Figure 4 shows partial gestural scores for the two ends of the overlap continuum. The boxes show the activation intervals for the rhotic and bilabial closure gestures, and the superimposed curves show the vertical motions of the tongue tip and lower lip.

The rhotic was generated as a "complex" gestural constellation consisting of two simultaneous gestures, one controlling the tongue tip tract variables to produce a retroflex constriction on the hard palate, and the other controlling the tongue body tract variables to produce an upper pharyngeal constriction. In fact, Lindau (1978) argues that it is the tongue body gesture that constitutes an articulatory invariant for American English /r/. The tongue tip gesture may be replaced in some speakers and in some environments by a "bunched" tongue body gesture—a constriction formed by the tongue body at the margin between the hard and soft palates (Delattre & Freeman, 1968). Figure 5 shows the midsagittal outline of the vocal tract model when both of the gestures in the rhotic constellation have reached their targets. The complete gestural scores contained no overlap between the rhotic and the vocalic gesture. This differs from what would be produced by our generalized phasing statements and is unlikely to be correct, but we did not want to introduce any additional assumptions about how the tongue body gesture for the rhotic would blend with the tongue body gesture for the vowel (see Browman & Goldstein, 1989; Saltzman & Munhall, 1989 for a discussion of within-tract-variable blending).

At one endpoint of the overlap continuum [Fig. 4(a)] there is 40 ms overlap of the control regimes for the rhotic and bilabial closure gestures, while at the other endpoint [Fig. 4(b)] the gestures' control regimes are separated by 40 ms. As a way of visualizing the effect of differential overlap, Fig. 6 shows the midsagittal outline

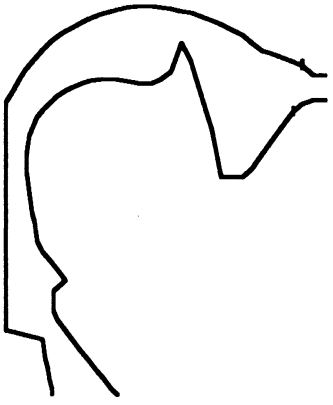

Figure 5. Midsagittal vocal tract model shape for [ɹ].

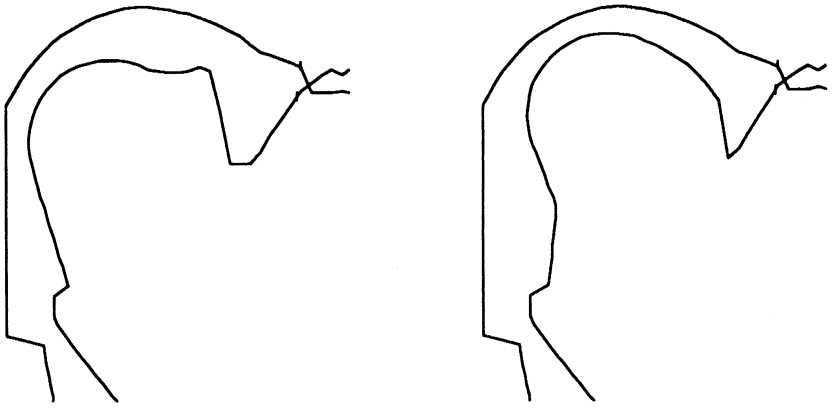

Figure 6. Midsagittal vocal trace model shapes when the bilabial closure gesture is turned off (a) for gestural score with maximal overlap (b) for gestural score with maximal separation.

of the vocal tract model when active control for the bilabial closure gesture turns off, for the two endpoint stimuli. For the maximal overlap configuration [Fig. 4(a)], Fig. 6(a) shows that the tongue shape has already begun to look like that for the rhotic at the point at which active control for the bilabial closure gesture turns off. However, for the maximal separation configuration [Fig. 4(b)], Fig. 6(b) shows there is no r-shape visible when the bilabial gesture turns off.

A total of nine gestural scores were created, with the phasing for the intermediate stimuli chosen so that, going from extreme overlap to extreme separation, the overlap decreased by exactly one synthesis frame (10 ms) for each step. The nine gestural scores were input to the task dynamic and vocal tract models to generate synthetic speech stimuli. The maximally separated endpoint stimulus was played informally to listeners to satisfy ourselves that the various gestures could be accurately perceived in a completely open-response format. Then a stimulus tape was created by randomizing ten repetitions of each of the nine stimuli; six listeners were asked to identify each token in this set as "bray" or "beret".

Figure 7 shows the percentage of "bray" or "beret" responses for the stimuli, totaled over all six subjects. As a group, the listeners switched from 67% "bray" responses at 10 ms overlap to 83% "beret" responses at the next step (which had zero overlap). The first "beret" point was at 0 ms overlap for four of the six subjects, and one 10 ms step earlier for the other two subjects. Five of the six subjects switched responses in a single 10 ms step: the average "bray" response from the frame just before the individual's crossover was 80%, while the average "beret" response one frame later was 80%. Thus, there was an abrupt switch from a percept of one syllable to a percept of two syllables caused solely by changing the amount of overlap. A first conclusion, then, is that changing overlap alone is effective in distinguishing pairs like "bray" and "beret".

A more surprising aspect of the results involved the location of the perceived boundary. For the first stimulus perceived as "beret", there was 0 ms overlap between the bilabial closure control regime and the rhotic control regimes. Thus, "beret" responses were made to stimuli with no overlap between control regimes

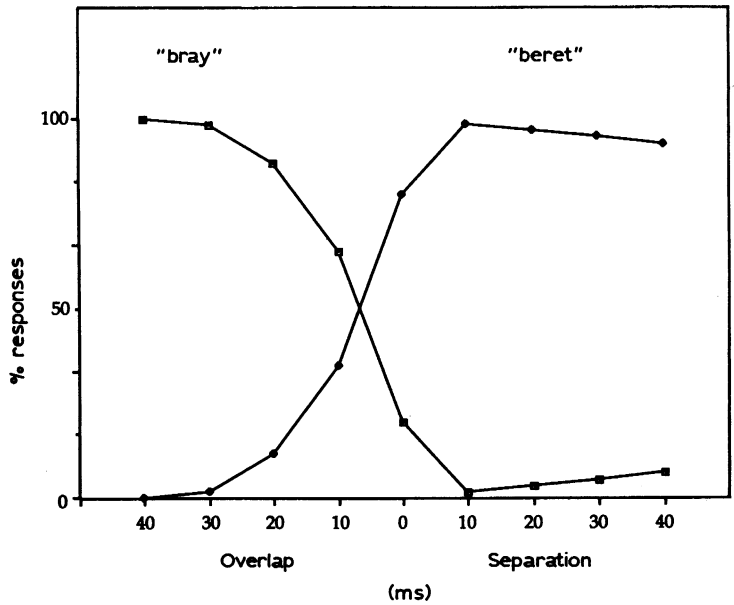

Figure 7. Percentage of "bray" *vs.* "beret" responses, totaled across all subjects, as overlap between bilabial closure and rhotic gestures decreases. 100% = 60 responses.

and "bray" responses were made to stimuli that did show overlap of control regimes. It may be, of course, that this was serendipitous. It is also true that the cross-over stimulus was the middle one in the continuum, as is to be expected in such experiments. Nonetheless, coincidence of the theoretical notion of overlap with the particulars of the results is encouraging, and lends plausibility to the hypothesis that the underlying difference in gestural structures for pairs like "bray" and "beret" could reside in the phasing of the consonant gestures. As noted earlier, this hypothesis would automatically account for the perceived loss of syllabicity in casual speech: with increased overlap, the gestural structure for "beret" would be the same as that for "bray".

While these results with model-generated speech are encouraging, tokens of natural speech must be examined to see whether their articulation is consistent with this proposed gestural difference. We have recently begun to collect X-ray microbeam data relevant to this issue at the NIH facility at the University of Wisconsin. Preliminary analyses appear consistent with the hypothesis of differing gestural overlap. As can be seen in Table I, the pairs "braid"/"bereted" and "prayed"/"parade" differ in the length of the interval between the articulatory bilabial release and the achievement of the rhotic target. (The bilabial release was defined to be the point at which the lower lip lowering out of the bilabial closure first increased to a velocity of 0.1 mm/s. The achievement of target for the rhotic was defined as the point at which tongue tip raising into the rhotic first decreased to a velocity of 0.1 mm/s.)

The difference in the bilabial-rhotic interval for the mono- and bisyllables analyzed is consistent with a difference in overlap. However, before claiming that the sole difference between such pairs resides in the overlap, it would be necessary

TABLE I. Interval between articulatory bilabial release and rhotic target

		ms				ms		
		n	mean	s.d.		*n*	mean	s.d.
Accented	Braid	4	13	12	Bereted	6	116	13
	Prayed	5	−1	21	Parade	6	158	17
Unaccented	Braid	5	−1	11	Bereted	3	73	25
	Prayed	5	−3	17	Parade	5	90	20

both to determine the gestural onsets and to show that there is no extra tongue body movement (for schwa) in the bisyllables. Although the analysis has not proceeded to the point that definite conclusions can be made, the representative tokens in Fig. 8 suggest that the difference between the mono- and bisyllables might indeed be ascribable solely to a difference in overlap.

Figure 8 shows the movement of pellets placed on the tongue dorsum, lower lip, and tongue tip for "braid" (solid lines), overlaid with data for "bereted" (dotted lines). Both words were produced in the phrase "I say ____ today", with "braid" (or "bereted") accented. Note that the lower lip is relatively high during the rhotic, indicating that a rounding gesture accompanies the tongue tip gesture, as is regularly seen for American English /r/ (Delattre & Freeman, 1968). Two separated labial raising movements can be observed for "bereted". The two utterances clearly differ in the relative timing of the rhotic tongue tip raising motion and the initial bilabial gesture (the points used in the measurements for Table I are marked with arrows). Moreover, the tongue dorsum movements in the two utterances are quite similar. If "bereted" were to contain a separate vowel gesture (a schwa) that is absent in "braid", we would expect to see a difference between the two utterances in the behavior of this pellet (see Browman & Goldstein, in press).

However, another articulatory study does not support the strong form of the claim that perceived instances of reduced syllables are solely the result of non-overlap of successive consonant gestures. Browman & Goldstein (in press) examined the nature of medial schwa vowels in utterances of the form /'pV1pəPV2pə/ to test the hypothesis that there would be no explicit vocalic gesture associated with the schwa, and therefore that the tongue would move in a continuous fashion from the position for V1 to that for V2, passing through some intermediate position during the acoustic interval for the schwa. This was tested both by statistical analyses of X-ray microbeam data and by simulations using various hypothesized gestural scores. The analyses showed that the strong form of the hypothesis could not be maintained for these utterances, at least in the environment where V1 = V2 = /i/ or V1 = V2 = /u/. In these cases, the position of the tongue during the schwa was lowered compared with the high vowels on either side, whereas the hypothesis predicted that the tongue should remain high.

While the strong form of the gestural overlap hypothesis was not supported by this study, viewing schwa in terms of gestural overlap led to an interesting and viable form of conceptualizing the data. The target for the X-ray tongue pellets for the schwa turned out to be completely "colorless": it was the mean of the targets for all the full vowels. Overlapping V2 (but not V1) during the entire time domain of this colorless schwa proved to correctly capture patterns of systematic articulatory

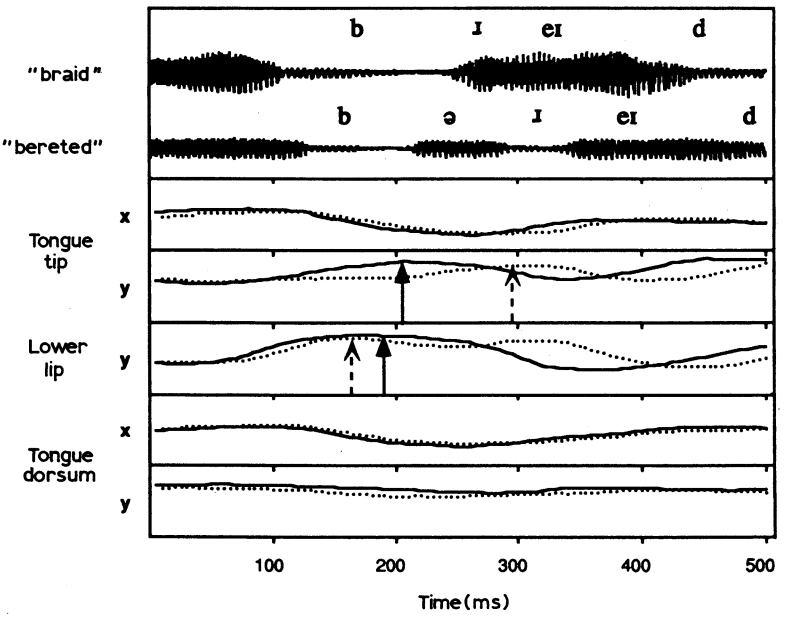

Figure 8. X-ray data for "braid" (——) and "bereted" (· · · ·), showing horizontal and vertical movement of pellets on tongue tip and tongue dorsum and vertical movement of pellet on lower lip. The vertical extent of each panel is 30 mm. The arrows mark the achievement of target for the rhotic (tongue tip) and the onset of the opening movement for the bilabial (lower lip). (Solid arrow for "braid"; dashed arrow for "bereted".) The two sets of data are lined up such that the achievement of target for the bilabial closures coincide.

variation during the schwa. The identity of the schwa, then, was very weak, both in its colorless nature and in its being completely overlapped by a full vowel. Moreover, it was possible to obtain a percept of schwa in a simulation of /'pipəpipə/ in which there was no active schwa gesture. This gestural score differed from the gestural score that encoded the data analyses only in having the schwa gesture removed and the consonants on either side phased closer together so that the acoustic schwa duration was shorter. For acoustics generated from this gestural score, a schwa percept was obtained in an informal listening test, even though V1 and V2 were /i/.

Although gestural overlap played an important role in both these studies, and although the schwa was weak in the second study, the results of the two studies with regard to the hypothesis of a totally unspecified reduced syllable nucleus were conflicting. The conflict might be resolved in several different ways. It might be, for example, that further investigations would discover that all reduced syllables contain a gesture for the nucleus. In such a case, an increase in overlap between the surrounding consonants could still be the source of the different variants of "beret". Another possibility is that reduced syllables in different phonetic and/or morphological environments (e.g., stop–stop *vs.* stop–liquid, or *Rosa's vs. roses*) might be associated with different gestural structures. On the basis of the above results, we would expect that reduced syllables in stop–stop environments and also in words such as *Rosa('s)* might contain a vocalic gesture for schwa, while those in stop–liquid environments and words such as *roses* might have no separate gesture.

This further suggests that investigations conducted within the gestural framework might lead to interesting typologies of reduced syllables, and more generally of processes involving changes in syllabicity.

One such possible typology involves the development of epenthetic vowels in languages. Matson (in preparation) has hypothesized that such vowels develop from the perception of the interval between two consonant gestures, as the consonants spread apart (in time) and overlap less. The fact that these intervals may later be identified with one of the full vowel gestures of the language could be considered a subsequent, "listener-based", sound change, along the lines of Ohala's (1981) model. Matson proposes that different types of epenthesis occur depending on whether the separating consonants are in the same syllable or not. On the one hand, if the consonants are either heterosyllabic or "unsyllabifiable", she finds that languages insert a constant vowel, and that such vowels are universally non-low, as would be expected of an interval resulting from consonant gesture separation, in which the tongue body would frequently be high due to the surrounding consonant constrictions. On the other hand, if the separating consonants are tautosyllabic, she finds, following Steriade (1989), examples in which the epenthetic vowel is identical with the (original) syllable nucleus. This is predicted by a gestural analysis, because the separation (sliding) in this case uncovers the overlapping vowel gesture that underlies the entire original syllable. Steriade (1989) further shows that slightly different amounts of sliding in these cases can yield the metatheses that are found, in place of epenthesis, in related languages. While the details of such a typology are still sketchy, it further demonstrates the kinds of questions one can ask within a framework in which gestural overlap is directly characterized and input structures are clearly distinguished from output consequences.

Our thanks to Caroline Smith and Elliot Saltzman for criticizing earlier versions of this paper, to Joshua Katz for helping analyse the microbeam data, and to Diana Matson for aid in data collection and figure preparation. This work was supported by NSF grant BNS 8820099 and NIH grants HD-01994 and NS-13617 to Haskins Laboratories.

References

Abraham, R. H. & Shaw, C. D. (1982) *Dynamics—the geometry of behavior.* Santa Cruz, CA: Aerial Press.

Beckman, M. E., Edwards, J. & Fletcher, J. (in press) Prosodic structure and tempo in a sonority model of articulatory dynamics. *Papers in laboratory phonology II* (G. Docherty & D. R. Ladd, editors). Cambridge: Cambridge University Press.

Browman, C. P. & Goldstein, L. (1985) Dynamic modeling of phonetic structure. In *Phonetic linguistics* (V. Fromkin, editor), pp. 35–53. New York: Academic Press.

Browman, C. P. & Goldstein, L. (1986*a*) Towards an articulatory phonology, *Phonology Yearbook*, **3**, 219–252.

Browman, C. P. & Goldstein, L. (1986*b*) Dynamic processes in linguistics: casual speech and historical change, *PAW Review*, **1**, 17–18.

Browman, C. P. & Goldstein, L. (1987) Tiers in Articulatory Phonology, with some implications for casual speech, *Haskins Laboratories Status Report on Speech Research*, **SR-92**, 1–30. [Also in *Papers in Laboratory Phonology I: Between the grammar and the physics of speech* (J. Kingston & M. E. Beckman, editors), pp. 341–376. Cambridge: Cambridge University Press (1989).]

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

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

Browman, C. P. & Goldsein, L. (1990) Gestural structures: Distinctiveness, phonological processes, and historical change. In *Modularity and the motor theory of speech perception* (I. G. Mattingly & M. Studdert-Kennedy, editors), pp. 313–338. Hillsdale, NJ: Erlbaum.

Browman, C. P. & Goldstein, L. (in press) "Targetless" schwa: An articulatory analysis. In *Papers in Laboratory Phonology II* (G. Docherty & D. R. Ladd, editors). Cambridge: Cambridge University Press.

Cooke, J. D. (1980) The organization of simple, skilled movements. In *Tutorials in motor behavior* (G. E. Stelmach & J. Requin, editors), pp. 199–212. Amsterdam: North-Holland.

Dalby, J. M. (1984) Phonetic structure of fast speech in American English. Unpublished doctoral dissertation, Indiana University.

Delattre, P. & Freeman, D. C. (1968) A dialect study of American *r*'s by x-ray motion picture, *Linguistics*, **44**, 29–68.

Fant, G. (1960) *Acoustic theory of speech production*. The Hague: Mouton.

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

Fowler, C. A. (1981) A relationship between coarticulation and compensatory shortening, *Phonetica*, **38**, 35–50.

Fowler, C. A., Rubin, P., Remez, R. E. & Turvey, M. T. (1980) Implications for speech production of a general theory of action. In *Language production* (B. Butterworth, editor), pp. 373–420. New York: Academic Press.

Goldstein, L. (1989) On the domain of the quantal theory, *Journal of Phonetics*, **17**, 91–97.

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

Haken, H., Kelso, J. A. S. & Bunz, H. (1985) A theoretical model of phase transitions in human hand movements, *Biological Cybernetics*, **51**, 347–356.

Hawkins, S. (in press) An introduction to task dynamics. In *Papers in laboratory phonology II* (G. Docherty & D. R. Ladd, editors). Cambridge: Cambridge University Press.

Kay, B. A., Kelso, J. A. S., Saltzman, E. L. & Schöner, G. (1987) Space-time behavior of single and bimanual rhythmical movement: Data and limit cycle model, *JEP: Human Perception and Performance*, **13**, 178–192.

Kelso, J. A. S. & Tuller, B. (1984*a*) A dynamical basis for action systems. In *Handbook of cognitive neuroscience* (M. Gazzaniga, editor), pp. 321–356. New York: Plenum.

Kelso, J. A. S. & Tuller, B. (1984*b*) Converging evidence in support of common dynamical principles for speech and movement coordination, *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, **246**, R928–R935.

Kelso, J. A. S. & Tuller, B. (1987) Intrinsic time in speech production: theory, methodology, and preliminary observations. In *Motor and sensory processes of language* (E. Keller & M. Gopnik, editors), pp. 203–222. Hillsdale, NJ: Lawrence Erlbaum.

Kelso, J. A. S., Saltzman, E. L. & Tuller, B. (1986) The dynamical perspective on speech production: data and theory, *Journal of Phonetics* **14**, 29–59.

Kelso, J. A. S., Holt, K. G., Rubin, P. & Kugler, P. N. (1981) Patterns of human interlimb coordination emerge from the properties of nonlinear limit cycle oscillatory processes: Theory and data, *Journal of Motor Behavior*, **13**, 226–221.

Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E. L. & Kay, B. (1985) A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling, *Journal of the Acoustical Society of America* **77**, 266–280.

Krakow, R. A. (1989) The articulatory organization of syllables: a kinematic analysis of labial and velar gestures. PhD. dissertation, Yale University.

Kugler, P. N. & Turvey, M. T. (1987) *Information, natural law, and the self-assembly of rhythmic movement*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Lindau, M. (1978) Vowel features, *Language*, **54**, 541–563.

Lindblom, B., MacNeilage, P. & Studdert-Kennedy, M. (1983) Self-organizing processes and the explanation of phonological universals. In *Explanations of Linguistic Universals*. (B. Butterworth, B. Comrie, & O. Dahl, editors), pp. 180–203. The Hague: Mouton.

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

McGarr, N. S. Löfqvist, A. & Story, R. (submitted) Jaw kinematics in hearing-impaired speakers, *Journal of the Acoustical Society of America*.

McGowan, R. S., Smith, C. L., Browman, C. P. & Kay, B. A. (1988) Extracting dynamic parameters from articulatory movement, *Journal of the Acoustical Society of America*, **83**, S113. (Paper presented at the 115th meeting of ASA, Seattle.)

McGowan, R. S., Smith, C. L., Browman, C. P. & Kay, B. A. (1990) Methods for least-squares parameter identification for articulatory movement and the program PARFIT, *Haskins Laboratories Status Report on Speech Research*, **101/102**, 220–230.

Nittrouer, S., Munhall, K., Kelso, J. A. S., Tuller, B. & Harris, K. S. (1988) Patterns of interarticulator phasing and their relation to linguistic structure, *Journal of the Acoustical Society of America*, **84**, 1653–1661.

Ohala, J. J. (1981) The listener as a source of sound change. In *Papers from the parasession on language and behavior* (C. S. Masek, R. A. Hendrick, & M. F. Miller, editors), pp. 178–203. Chicago: Chicago Linguistic Society.

Öhman, S. E. G. (1967) Numerical model of coarticulation, *Journal of the Acoustical Society of America*, **41**, 310–320.

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

Perrier, P., Abry, C. & Keller, E. (1988) Vers une modelisation des mouvements du dos de la langue, *Bulletin du Laboratoire de la Communication Parlée*, **2**, 45–63.

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

Saltzman, E. (1986) Task dynamic coordination of the speech articulators: A preliminary model. In *Generation and modulation of action patterns*, Experimental Brain Research Series 15 (H. Heuer & C. Fromm, editors), pp. 129–144. New York: Springer-Verlag.

Saltzman, E. & Kelso, J. A. S. (1987) Skilled actions: A task dynamic approach, *Psychological Review*, **94**, 84–106.

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

Smith, C. (1988) A cross-linguistic contrast in consonant and vowel timing, *Journal of the Acoustical Society of America*, **86**, S84. (Paper presented at the 116th meeting of the ASA, Honolulu.)

Smith, C., Browman, C. P., McGowan, R. & Kay, B. (submitted) Extracting dynamic parameters from speech movement data, *Journal of the Acoustical Society of America*.

Sproat, R. & Fujimura, O. (1989) Articulatory evidence for the non-categoricalness of English /l/ allophones. Presented at the LSA Annual Meeting, Washington DC, December.

Steriade, D. (1989) Gestures and autosegments: comments on Browman and Goldstein's "Gestures in Articulatory Phonology". In *Papers in laboratory phonology I: Between the grammar and the physics of speech* (J. Kingston & M. E. Beckman, editors), pp. 382–397. Cambridge: Cambridge University Press.

Stevens, K. N. (1972) The quantal nature of speech: Evidence from articulatory-acoustic data. In *Human communication: A unified view* (E. E. David & P. B. Denes, editors), pp. 51–66. New York: McGraw-Hill.

Stevens, K. N. (1989) On the quantal nature of speech, *Journal of Phonetics*, **17**, 3–45.

Stevens, K. N. & House, A. S. (1955) Development of a quantitative description of vowel articulation, *Journal of the Acoustical Society of America*, **27**, 484–493.

Tuller, B., Kelso, J. A. S. & Harris, K. S. (1982) Interarticulator phasing as an index of temporal regularity in speech, *JEP: Human Perception and Performance*, **8**, 460–472.

Turvey, M. T. (1977) Preliminaries to a theory of action with reference to vision. In *Perceiving, acting, and knowing* (R. Shaw & J. Bransford, editors), pp. 221–265. Hillsdale, NJ: Lawrence Erlbaum Associates.

Turvey, M. T., Rosenblum, L. D., Kugler, P. N. & Schmidt, R. C. (1986) Fluctuations and phase symmetry in coordinated rhythmic movements, *JEP: Human Perception and Performance*, **12**, 564–583.

Vatikiotis-Bateson, E. (1988) *Linguistic structure and articulatory dynamics: a cross-language study*. Bloomington, Indiana: Indiana University Linguistics Club.