Available online at www.sciencedirect.com

SCIENCE DIRECT

Journal of Phonetics 31 (2003) 149–180

Journal of Phonetics

www.elsevier.com/locate/phonetics

The elastic phrase: modeling the dynamics of boundary-adjacent lengthening

Dani Byrd[a,*,1], Elliot Saltzman[b,c,1]

[a] Department of Linguistics, University of Southern California, 3601 Watt Way GFS 301, Los Angeles, CA 90089-1693, USA
[b] Haskins Laboratories, New Haven, CT, USA
[c] Boston University, Boston, MA, USA

Received 8 January 2001; accepted 13 November 2002

Abstract

This work examines the relation between phrasal structure and the control and coordination of articulation within a dynamical systems model of speech production. In this context, we review how speakers modulate the spatiotemporal organization of articulatory gestures as a function of their phrasal position. We present computational simulations that capture several important qualitative properties of these phrase boundary effects, such as prosodically-induced local slowing. This slowing is generated by dynamical effects on the activation timecourse of articulatory gestures and is controlled by prosodic gestures or π-gestures, which share much with the familiar dynamical description of constriction gestures. Prosodic gestures, however, function at boundaries purely to temporally stretch or shrink gestural activation trajectories. This modulation of the "clock-rate" that controls the temporal unfolding of an utterance near junctures is such that the clock slows increasingly as the boundary is approached and speeds up again as the boundary recedes. Viewing phrase boundaries as warping the temporal fabric of an utterance represents a promising confluence of the fields of prosody and of speech dynamics.

1. Introduction

Spoken language relies on an elegant and intricate acoustic structure to support communication between speakers and listeners. The temporal orchestration of articulatory activity that produces this acoustic signal is an essential element in the communication process. The complex messages

---

*Corresponding author. Tel.: +1-213-821-1227; fax: +1-213-740-9306.
E-mail address: dbyrd@usc.edu (D. Byrd).

[1] The authors are listed in alphabetical order. They contributed in equal measure.

doi:10.1016/S0095-4470(02)00085-2

and emotions of spoken language must be communicated by the precise choreography of the jaw, tongue, lips, larynx, and respiratory system. Understanding the temporality of speech has proven to be a formidable challenge. For example, it has become increasingly clear that the temporal patterning of a word's articulation and its coordination with nearby words is crucially dependent on the prosodic structure in which the words occur. The phonological specification of a word does not merely *unroll* along a linear timeline. In his classic work, James Gibson commented that…

> "the flow of ecological events is distinct from the abstract passage of time…The stream of events is heterogeneous and differentiated into parts, whereas the passage of time is supposed to be homogeneous and linear. Isaac Newton asserted that "absolute, true, and mathematical time, of itself and from its own nature, flows equably without relation, to anything external." But this is a convenient myth." (Gibson, 1986, p. 100)

While Gibson's focus was primarily on the human perception of events in time, his point can be understood also with respect to the human production of events in time. Below we review a dynamical systems approach to speech production in which the flow of time during an utterance is sensitive to the prosodic structure of that utterance. Thus, rather than a conceptualization in which abstract time is viewed as an immutable flow that is indifferent to the concomitantly evolving event patterns, our position echoes Gibson's view that "abstract time is a ghost of the events of the world" (Gibson, 1975).

2. Why a dynamical systems approach?

Phoneticians have long been confronted with a general puzzle: How can surface variability in spoken language performance be explained in light of what is intuited as a fundamental underlying invariance in control? This puzzle exists because of certain fundamental assumptions. Linguists assume that words are built of a limited number of discrete units that can be combined and recombined, and that these units have no independent meaning but are the building blocks of words. They have been referred to as phonological units. It is this capacity to (relatively) freely combine and organize phonological units that allows words to be coined easily. Further, the existence of a small fixed inventory of phonological units for a given language is presumably one of the factors allowing for the successful acquisition of the complex skill of language use. However, when the units are made manifest in word and sentence production, linguists have recognized that their realization by the articulatory system, and consequent character presented to the auditory system, is highly variable. As phonological units are structured into syllables and phrases, their temporal and spatial characteristics are not constant but, rather, vary systematically as a function of the larger linguistic units in which they are participating. Such variability has been discussed by phoneticians under the headings of coarticulation, coproduction, contextual variability, and prosodic variability.

A dynamical systems framework allows an understanding of the relationship between underlying units of control and surface variability of those units. In the following, we will review the Task Dynamics model of speech production (e.g., Saltzman, 1986; Saltzman & Munhall, 1989; Saltzman & Byrd, 2000a; see also Hawkins, 1992) and then explore how prosodically conditioned surface variability can be modeled within this framework. In particular, we will review phrasal effects on articulatory timing as an example of such dynamically-shaped variability. This focus will allow us to explore a specific view of how underlying temporal characteristics of linguistic units can be modulated for communicative ends in the production of a particular utterance.

2.1. Task dynamics and articulatory phonology

A dynamical systems model is a set of laws or rules that specify the forces that change aspects of a system over time. Such models have been used to capture characteristics of many natural phenomena including the control of skilled human movement. In skilled human movement, many moving body parts must be coordinated such that patterns are stable in space and time and can be repeatedly generated. Dynamical systems models are useful in this regard because such a framework provides a unified account not only of movement forms and their stability properties, but also of the way that these forms become altered as the system's underlying parameter values are changed (e.g., Saltzman & Kelso, 1987). Saltzman (1995) and Browman and Goldstein (1995) discuss the application of these ideas to speech production.

We will frame our further discussion within a particular theory of phonology and a particular dynamical architecture of the speech production system. Specifically, we adopt *Articulatory Phonology* (Browman & Goldstein, 1986, 1990, 1992) as a formal theory of phonological units and their organization, and *Task Dynamics* (Saltzman, 1986; Saltzman & Munhall, 1989) as a quantitative model that implements these phonological units in the speech production system.

Articulatory Phonology views lexical representations as being composed of stable combinatorial units, called gestures, that are articulatory in nature. A gesture is a goal-directed movement of the vocal tract such as the formation of a constriction. Gestures have two functions (Browman & Goldstein, 1992)—they function as units of information, i.e., linguistic contrast, and as units of action, i.e., speech production. This view of word structure and speech production postulates that there is no mediation between the phonological representation and its implementation by the speech production system. The units of contrast/combination and those of articulation are one and the same, and the representation includes all information necessary for the execution. This is possible because of three properties of gestures. First, gestures are *abstract*. They are defined in terms of linguistic tasks (e.g., "achieve a narrow constriction using the tongue tip") not in terms of specific movements (e.g., *not* "move the tongue tip, tongue body, and jaw through particular distances in particular directions"). Second, gestures have inherent spatiotemporal properties—they are instantiated/defined in particular regions of vocal tract space and with particular durational characteristics. Third, gestures are *coordinated*, or *phased*, with respect to one another in a manner that creates spoken utterances that cannot be described as simple ("*beads on a string*") sequences of discrete gestures. Rather, the pattern of relative phasing among gestures involves temporal overlap, and this gestural coproduction causes competition for the articulators, thereby yielding coarticulatory effects that are ubiquitous in speech.

The gestural units of Articulatory Phonology have been computationally implemented using the Task Dynamic model of sensorimotor control and coordination (e.g., Saltzman, 1986; Saltzman & Munhall, 1989; Saltzman & Kelso, 1987). This model addresses the manner in which an underlying invariant dynamical system can give rise to lawful, contextually-conditioned surface variability in observed motion patterns of limbs and speech articulators. For modeling the gestural patterning of speech production, two functionally distinct but interacting levels are defined. The *constriction level* is defined according to both model articulator (e.g., lips and jaw) coordinates and tract-variable (i.e., vocal tract constriction system, e.g., lip aperture and protrusion) coordinates. The *activation level* is defined according to a set of gestural activation coordinates. Gestural units are posited in the form of context-independent sets of intrinsic dynamical parameters (e.g., target and stiffness coefficients) and are associated with corresponding subsets of activation, tract-variable, and model articulator coordinates. Each unit's activation coordinate represents a forcing function whose value reflects the strength with which that gesture can shape or dominate vocal tract movements at any point in time. The tract-variable and model articulator coordinates of each unit specify, respectively, the particular vocal tract constriction (e.g., lip aperture) and articulatory synergy (e.g., lips and jaw) whose behaviors are affected directly by the associated unit's activation. The constriction level defines the dynamics that accounts for the coordination among articulators at a given point in time due to the currently active gesture set. The activation level governs the temporal evolution of the activation trajectories of individual gestures in an utterance as well as the patterns of relative timing or phasing among the gestural units participating in the utterance. In Articulatory Phonology, the activation intervals and parameter sets for a given word are specified according to a *gestural score*. A visual representation of the gestural score for the word /bæn/ is shown in Fig. 1.

2.1.1. Timing and lexical representation

The lexical representation for a word must include the gestures (i.e., atomic phonological units) of which the word is composed. So for the word /bæn/ this would include a lip aperture gesture, a tongue body gesture, a velum gesture, and a tongue tip gesture (see Fig. 1). The first element of lexical contrast is the tract variable affiliation of a gesture; this determines the vocal tract subsystem that will perform the linguistically important constriction. Next, target parameter values are part of the gestural representation of lexical items, since these values form the basis for the system of phonological contrast in a language, for example contrasting constriction locations and degrees. Lastly, the temporal organization of these gestures must be captured with some level of detail in the lexical knowledge. After all, the same gestures occur in both [bæ̃n] and [næb] but with different temporal orderings.

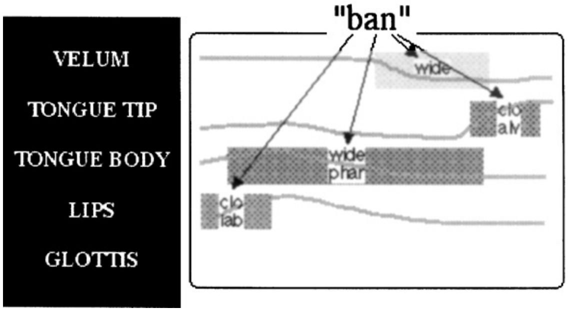


Fig. 1. The gestural score for [bæ̃n]. Boxes indicate activation intervals, notation inside boxes indicates gestural parameter values for constriction location and/or degree, and each tier indicates a particular tract variable. Tract variable trajectories have been overlaid. (Adapted from C. Browman and L. Goldstein.)

Whether something more detailed than canonical precedence relationships of gestures (e.g., order of target achievement) is part of the relative timing specification in a lexical entry remains a difficult question. In our view, a specification of at least the allowable ranges or *windows* of relative timing among the gestures in a given sequence must also be included (see, e.g., Byrd, 1996b; Docherty, 1992; Saltzman & Byrd, 1999, 2000a). This would allow patterns of intergestural timing (e.g., between consonants in sequence) to be somewhat variable (see, e.g., Byrd, 1996a, b; Docherty, 1992; Chitoran, Goldstein, & Byrd, 2002). An alternative view is that of Browman and Goldstein (2000) in which pairs of gestures in a word are specified as participating in an explicit and precise timing or phasing relationship. Phase relations are assigned by rule—for example, consonants in clusters are specified as being phased such that there is no overlap between them, while each onset consonant–vowel relation is specified as being synchronous (fully overlapped). Finally, each of the phase relationships has an associated bonding strength that represents the degree to which that relationship will yield in the face of other conflicting timing relationships in which those gestures participate. That is, timing specifications are subject to a competitive dynamics in which the specifications compete in a weighted fashion to produce the timing observed, in for example CCV sequences.

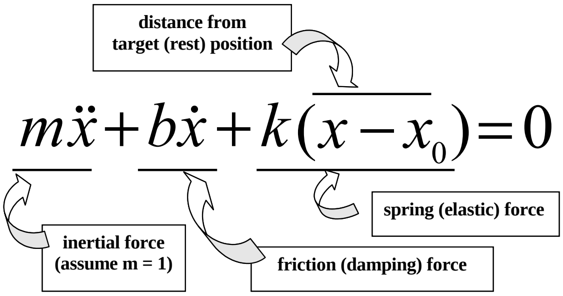
Both of these approaches provide a means for describing how timing might play a critical role in characterizing larger gestural molecules (i.e., phonological units larger than the atomic gesture) such as segments and syllable onsets (Browman & Goldstein, 1989; Byrd, 1996b; Goldstein & Fowler, submitted). Namely, such molecules are conceived of as having particularly cohesive timing relationships (Byrd, 1996b; Browman & Goldstein, 1989; Goldstein & Fowler, submitted). Despite their differences in details of implementation, both of these approaches acknowledge the importance of incorporating information about the temporal relations of gestural units beyond simple canonical order into the lexical representation of a word.

In the remainder of this paper we turn to a specific way in which prosodically-conditioned temporal variability can be conceptualized in a dynamical systems approach by our two-level (activation and constriction) model of gestural patterning.

2.1.2. Constriction level: the virtues of gestural point attractors

Gestures are characterized qualitatively as having *point attractor* dynamics at the constriction level. Point attractor models with varying degrees of complexity have been used to characterize many skilled human movements, for example, reaching (for reviews see, e.g., Shadmehr, 1995; Mussa-Ivaldi, 1995; Flash & Sejnowski, 2001). Point attractors have important properties which seem to be characteristic of speech behavior. For example, they approach their target (or "equilibrium position") regardless of their initial position or unexpected perturbations. This means that linguistic tasks (such as making a closure for a stop) can be specified as attractors that move the vocal tract articulators toward the appropriate position for that task regardless of their current (or past or future) positions. If two attractors (or phonological tasks) are simultaneously exerting forces on the vocal tract, some accommodation will be reached.

Let us consider a very simple point attractor model: the damped mass–spring system. Such a system has been used to model the formation and release of vocal tract constrictions, and a familiarity with the parameters of such a system will be important later when we explore



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

Fig. 2. The equation of motion for a damped mass–spring system, or a simple *point attractor*.

prosodically-conditioned variability in the realization of gestures. The equations that describe the dynamics of damped mass–spring systems include force components associated with four parameters. The target parameter, x_0, specifies the position of the system's mass when the spring is at its rest length or equilibrium length. In Articulatory Phonology, the target position is the constriction location or constriction degree for the particular vocal tract subsystem on which that gesture is defined (e.g., for the tongue tip for [d] or [s]). The mass parameter, m, is assumed to uniformly equal 1 for all gestures. The *stiffness* parameter, k, represents the elasticity of the system and is proportional to intrinsic gestural "speed." A gesture with a lower stiffness (perhaps a vowel) will have an intrinsically longer duration than a gesture with a higher stiffness (perhaps a consonant). Finally, the damping parameter, b, is typically set to be "critical" in a gestural model indicating that a gesture does not display a succession of oscillatory overshoots and undershoots on approach to its target but rather moves toward it smoothly. Using Newton's second law, $F=ma$, these parameters can be incorporated into an expression of the forces characterizing a damped mass–spring system, or a simple *point attractor*, in the linear equation of motion shown in Fig. 2.

Point attractor behavior characterizes not just many physical systems but behavioral/cognitive systems as well. In fact, Plaut (1995) comments that "virtually all existing attempts to model cognitive and neuropsychological phenomena have relied on "point" attractors" (p. 541). Thus, within Task Dynamics, the mass–spring model is used not as a model of a "physical" system but of the cognitive control of abstract linguistic units. In particular, constriction gestures are defined in tract-variable space and not at the level of the individual articulators. For example, the lip closing gesture for a [b] is defined in a relatively abstract lip aperture space where the goal is to achieve lip closure and is not defined in a more concrete articulator space as separate movement goals for the upper lip, lower lip and jaw. Setting x_0 for this tract-variable point attractor to zero lip aperture[2] will cause the articulators to move until the target ($x_0 = 0$) is achieved or until they are called away by another overlapping (i.e., competing) gesture. (For a detailed description of how task-space-defined gestures are quantitatively implemented at the articulator level, see Saltzman & Munhall, 1989.)

---

[2] Actually, a negative lip aperture is a more realistic target since the lips compress on closure.

Such gestural point attractors illustrate one way in which underlyingly invariant dynamics can give rise to contextually-conditioned surface variations in performance.[3] For example, a damped mass–spring system (i.e., gesture) with an invariant target for [b] will yield contextually-variable positions of the lips and jaw for the [b] in [bi] and [bæ]. These variations are due to the temporal overlapping or coproduction of the gestures for [b] and the following vowel, and reflect a blending at the articulatory level of gestural influences associated with attaining each gesture's invariant tract-variable goals. The temporal overlap of gestures is determined by the relative timing of gestural activation trajectories, which are shaped at the activation level of the model.

2.1.3. Activation level: the shaping of activation trajectories

In approaches relying on gestural scores for timing, the evolution of activation trajectories has neither an intrinsic dynamics of its own nor is it coupled dynamically to the ongoing trajectories at the constriction level. Thus, the activation trajectories for words are preplanned according to rules of the phonology and unidirectionally coupled into the constriction level. In recent work (Saltzman, Löfqvist, & Mitra, 2000; Saltzman, 1999), we have explored how activation dynamics might be defined both independently of any other dynamical level and also in a manner that incorporates ongoing modulatory feedback from the constriction level. Roughly, the dynamical system for activation can be conceptualized in terms of a "clock" whose timeflow is mapped into a pattern of sequentially ordered and temporally overlapping gestural activations and which is sensitive to the ongoing state of the constriction-level variables. In this view, the clock provides a temporal context for an utterance by defining a phase-angular flow relative to which gestural timing is specified. When the clock rate is uniform throughout its cycle, clock time (phase) is linearly related to the flow of "real-time" as measured, for example by a stopwatch held by an external observer. This kind of uniform clock rate can be conceived in terms of a global clock-frequency parameter that remains constant over a period of observation. However, experimental *phase-resetting* results (Saltzman, Löfqvist, Kay, Kinsella-Shaw, & Rubin, 1998), in which mechanical perturbations delivered to the articulatory periphery act to transiently speed the flow of the utterance, not only require the existence of such a clock but also show that the clock rate is not necessarily uniform throughout an utterance. More specifically, this result is consistent with the hypothesis that the perturbation is fed back to the underlying clock whose rate is transiently increased through a corresponding increase in the value of the clock's frequency parameter. The presence of such feedback modulation from the perturbed articulatory periphery has led to a related proposal that clock rate is continuously, albeit subtly modulated by ongoing unperturbed articulatory state (Saltzman et al., 2000). Additionally, it has been proposed by several investigators that articulatory clock rate is also modulated according to an utterance's segmental and syllabic structure (Bailly, Laboissière, & Schwartz, 1991; Barbosa & Bailly, 1994; Laboissière,

---

[3] We have suggested that gestural parameter values (as well as the tract variable on which a gestures is defined) form the basis for stable phonological representation. It is worth noting that "stable" in this context does not imply absolute invariance. We suggest that experience and recency play an important role in "tuning" parameter values of gestural representations of words. Gestural parameter values are determined by the language user by an ongoing process of continual updating. This is not a new idea—though we have posed it specifically in terms of gestural parameter tuning—and has been suggested with respect to VOT drift by Sancier and Fowler (1997). Sancier and Fowler describe experimental results indicating that relative timing of gestures (VOT for a Portuguese–English bilingual) is tuned in an on-going fashion as a function of recent experience (e.g., language setting).

Schwartz, & Bailly, 1991; O'Dell & Nieminen, 1999; Port & Cummins, 1992; Saltzman et al., 2000; Vatikiotis-Bateson, Hirayama, Honda, & Kawato, 1992; for a review of related time-warping approaches in the acoustic domain see van Santen, 1997). In this paper, we explore the hypothesis that clock rate is modulated by the phrasal structure of utterances according to the action of what we have labeled prosodic gestures (π-gestures). We will return to the details of our π-gesture hypothesis in Section 4, where we describe the manner in which these gestures might account for durational effects associated with phrase boundaries.

At first blush, our adoption of a clock to pace the flow of gestural activation appears to place this model in the category of the so-called *extrinsic* timing models. In these types of model, exemplified by Klatt (1976) or Ferriera (1993) and reviewed in Fowler (1980), time and temporal properties are not part of the purely symbolic representation of linguistic units. Rather, the temporality of a unit is only provided in performance through the operation of an executor that recruits the units in a particular serial order and timing pattern, with reference to a timekeeper or clock that is extrinsic to the units themselves. In contrast, *intrinsic* timing models are characterized by the incorporation of temporality into the definition of the units themselves. In the gestural approach, this property is a consequence of specifying gestures as dynamical systems that by definition incorporate laws governing behavior over time. Although we will invoke a clock to control the flow of gestural activation, we consider our model to fall into the intrinsic timing category. This is because the activation level dynamics of the clock and the constriction level dynamics of the gestural units are bidirectionally coupled and hence form a single higher-order dynamical collective.

3. Temporal patterning in speech production

Let us now turn to a categorization of four types of timing phenomena in speech—global, transgestural, intergestural, and intragestural—and how they are handled within the dynamical systems approach that we have outlined.

> Global timing refers to the temporal properties of an entire utterance, e.g., overall speaking rate, governed in part by sociolinguistic, dialectal, and individual factors.
>
> Transgestural timing refers to modulations of the timing properties of all gestures active during a localized portion of an utterance; i.e., local accelerations and decelerations.
>
> Intergestural timing refers to the relative coordination among gestures; and
>
> Intragestural timing refers to the temporal properties of a given gesture, e.g., the time from gestural onset to peak velocity or to target attainment.

It would be convenient to imagine that there is a different specific dynamical source for each of these phenomena. In the case of global timing, for example, it is reasonable to attribute global timing to the value of the central clock's frequency parameter. However, there do not appear to be single dynamical sources that account for the other three categories of timing phenomena. Rather, as is the nature of any multidimensional and multilayered complex system, the macro-level behaviors emerge from the confluence of activity distributed throughout the system.

Specifically, we hypothesize that these timing phenomena have their origins not only in the constriction-level dynamics of individual gestures (e.g., relating to a gesture's stiffness and target parameters), but also in the activation-level dynamics that shape each gesture's activation waveform and the relative timing among an utterance's set of gestural activations. For example, at least two dynamical sources potentially could give rise to transgestural timing phenomenon, that is, temporally local modulations of the durational properties of all concurrently active gestures. First, previous work (Beckman & Edwards, 1992; Edwards, Beckman, & Fletcher, 1991; Byrd & Saltzman, 1998) has described how changes in the gestures' intrinsic constriction-level dynamics, for example in gesture stiffness at a phrase boundary, might give rise to a local slowing of gestures in the vicinity of the boundary. However, a second possibility, which for reasons described below we consider to be more felicitous, views transgestural timing phenomenon as primarily resulting from the transient modulation of clock rate at the system's activation level.

Intergestural and intragestural timing phenomena are even more multifaceted. One way of conceiving the shaping of individual gestural activation trajectories as well as the relative timing among a set of gestures is to consider how the state of the activation-level's clock is mapped onto gestural activations. Imagine that in a word a particular gesture's onset, rise to an interval of maximal activation, and fall to offset are characterized by an associated set of lexically-specified gestural phase angle parameters on the clock. The trajectory shapes of individual gestural activation waveforms as well as the relative timing patterns among an utterance's gestural set will be shaped by the relation between the clock's evolving phase angle and the phase angle parameter values of the gestural set. For example, as the clock crosses a given gesture's onset phase, the gesture will "turn on," i.e., its activation value will begin to rise from zero; while the clock is within the gesture's maximum phase range, the gesture will be maximally activated; when the clock moves out of this region, activation will decrease; and when the clock passes the gesture's offset phase angle, the gesture's activation value will reach zero and the gesture will "turn off."[4]

Relatedly, intergestural timing for a given gestural set will be in part determined by the relations among the phase angle parameter values associated with each gesture in the set. It will also be affected by the rate of the clock as it moves through the angles of these parameter sets. It is worthwhile to consider the phenomenon of intergestural truncation (Harrington, Fletcher, & Roberts, 1995; Edwards, Beckman, & Fletcher, 1991) from this perspective. Truncation and its converse, de-truncation, refer to the kinematic consequences of the sliding in time of gestural activation waveforms relative to one another. For two gestures sharing articulators, when the leading edge of a following gesture slides backward into the trailing edge of a preceding gesture, the kinematic trajectory of the first gesture will be truncated—i.e., its duration will shorten and spatial extent will be smaller. Conversely, when the second gesture slides downstream away from a preceding gesture, the first will be detruncated—i.e., realized more fully, being larger and longer.

---

[4] Interestingly, it seems reasonable to hypothesize that the phase angle parameters of the gestural sets of lexical items are language specific. For example, the anticipation intervals of gestures may be defined as the time between gestural onset and target attainment. These intervals have been shown to display different degrees of temporal elasticity across languages (e.g., Abry & Lallouache, 1995; Saltzman et al., 2000). By temporal elasticity, we refer to the relationship between a given gesture's activation interval and a preceding interval before target attainment in which there are no conflicting gestures. In English, for example, anticipation intervals are relatively fixed regardless of the length of the preceding no-conflict intervals; in French, however, the anticipation periods stretch in proportion to the lengths of the no-conflict intervals.

We can identify several means of characterizing sliding in this model. One source involves changes in the phase angle parameter sets of the gestures in question, while another would involve changes in the rate of the clock as it passes through the parameter sets of the gestural set. This serves as an example of how multiple dynamical sources can affect intergestural timing phenomena.

Finally, intragestural timing phenomena likewise can be traced to multiple dynamical sources in this model. For example, the degree of symmetry of a gesture's velocity trajectory is defined by the time from a gesture's kinematic onset to its peak velocity divided by the time from kinematic onset to offset. We have shown previously that this measure is sensitive to the relationship between the value of a gesture's constriction level stiffness parameter and the rise-time of the gesture's activation (Byrd & Saltzman, 1998). In turn, the activation rise time is affected by both the phase angle parameter set of the gesture and the rate of the clock as it flows through these phase angles.

In summary, all four types of timing phenomena we outlined can be accounted for with reference to the dynamics of a two-level model incorporating both activation-level and constriction-level dynamics. In general, however, with the exception of global timing, the temporal properties of speech kinematics are multifaceted in their dynamic origins.

4. Boundary-adjacent lengthening

In the last two decades, linguistic descriptions of prosodic structure and of its relation to syntactic structure have flourished (e.g., Selkirk, 1984; Pierrehumbert & Beckman, 1988; Nespor & Vogel, 1986; Shattuck-Hufnagel & Turk, 1996; Beckman, 1996; Steedman, 1991, 2000). Here we will use the terms prosody and prosodic structure to refer to linguistic structure above the level of the gesture, segment, or syllable—specifically *accentual prominence* and *phrasal organization*. Examples of these are:

Accentual prominence:	There's no SMOKING on the plane…(but you can drink.) There's no smoking on the PLANE…(but you can smoke if you take the train.)
Phrasal organization:	When teenagers drive quickly they get tickets. (organization 1: When teenagers drive quickly, they get tickets.) (organization 2: When teenagers drive, quickly they get tickets.)

Prosody serves multiple purposes. Phrasing serves to group informational units together into appropriate "chunks" for cognitive processing for a speaker. It may also aid listener processing due to the salient temporal and pitch modulations that occur at phrase edges (see, e.g., Cutler, Dahan, & van Donselaar, 1997; Koiso, Shimojima, & Katagiri, 1998; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Sanderman & Collier, 1997; Schafer, 1997; Schafer & Speer, 1997; Schepman & Rodway, 2000). Prominence or accent directs the attentional resources of the listener to new and/or important information in the discourse (see, e.g., Cutler et al., 1997; Birch & Clifton, 1995). It may also reflect speaker strategies for focusing on this important information in production.

The use of intonation in encoding prosodic information has been studied for many decades. In more recent years, speech scientists have begun to establish a picture of how phrasal and accentual structure modulate the supralaryngeal articulatory production of words. In the remainder of this work, we will be considering specifically variation in the temporal patterning and spatial patterning of articulation at phrase boundaries.

Three phenomena in particular have been described as occurring at phrase edges—in simplistic terms, gestures get longer, larger, and farther apart. In addition to being motivated by speakers' cognitive needs, the instantiation of phrase boundaries is potentially an important communicative event signaling the informational structure of an utterance to listeners. Each of these effects can be viewed as increasing the perceptual salience of (i.e., "marking") a phrase boundary for a listener. Lengthened gestures allow longer perceptual exposure to the acoustic cues of these gestures and longer intervals during which to process this information. Lessened overlap prevents constrictions from obscuring one another and provides information in the acoustic signal at any particular point that (more) uniquely cues the listener to the presence of a single particular gestural unit rather than providing (more) parallel, coproduced information about multiple gestures. Under certain circumstances, even increases in spatial magnitude might provide more robust acoustic cues to recovery.

Let us first consider the previous articulatory findings regarding boundary-adjacent lengthening. At phrase edges, both initial and final, articulatory gestures have been observed to undergo a local lengthening (increased duration) (Beckman, Edwards, & Fletcher, 1992; Byrd & Saltzman, 1998; Byrd, 2000) and, sometimes, slowing (decreased peak velocity) (Cho, 2001, submitted). Changes in peak velocity of phrase initial closing gestures are observed by Cho (2001, submitted) who finds slower peak velocities for gestures following stronger boundaries. Lengthened lingua-palatal contact for closure seals of stops in initial position is observed by Fougeron (2001) and Cho and Keating (2001) for French and Korean, respectively. Fougeron (2001) remarks that "the occlusion is initiated and even achieved…during the preceding pause" (p. 123). In situations without a silent interval between phrases, lengthening of articulation in phrase-initial position contributes to the acoustic lengthening (e.g., "final lengthening") that has long been recognized at the end of a phrase. In particular, if a phrase initial constriction takes longer to achieve, greater acoustic final lengthening is likely to result. Signature changes in the shape of the movement trajectory of these lengthened gestures (e.g., longer times to peak velocity) indicated that a lowering of the stiffness parameter for gestures in the vicinity of the phrase edge—a controlled local slowing—might be a possible source of the lengthening (Beckman et al., 1992; Byrd & Saltzman, 1998; Byrd, 2000). However, other works have suggested that stiffness modulation is insufficient to characterize the full range of kinematic changes at boundaries (Saltzman & Byrd, 2000a,b; Cho, 2001, submitted). Rather than focusing solely on manipulating the stiffness parameter values of individual gestures, the approach we will outline to controlling this local slowing is by modulation of an utterance's temporal fabric in the vicinity of phrasal boundaries.[5] This idea was presaged by Gaitenby in her classic 1965 paper "The Elastic Word" in which she described the (acoustic) stretching and shrinking of words in continuous speech and suggested that prepausal lengthening be considered as a speech rate phenomenon.

---

[5] In fact, such strategies appear to be at work in another domain of skilled human action—musical performance. For example, Repp (1998) notes that at any particular point in a performance the music's structure may be more or less temporally stretchable.

4.1. Prosodic gestures

How can these phrase boundary effects on articulatory timing be captured within a dynamical model? Or more generally, how can apparently atemporal symbolic linguistic structure exert in a principled fashion the effects needed to explain the observed patterning of low-level inherently temporal action units? In Byrd, Kaun, Narayanan, and Saltzman (2000) and Byrd (2000), we described a conceptual approach to boundary-adjacent slowing. We suggested the view that phrase boundaries are instantiated by a π-gesture (or prosodic-gesture) that acts transgesturally to slow all concurrently active constriction gestures in proportion to the activation level of the π-gesture. Just as articulatory gestures have inherent durational properties and are temporally coordinated and potentially overlapping with other gestures, π-gestures also have an extent in time and overlap with vocal tract constriction gestures. A schematic gestural score for two gestures spanning a phrase boundary that is instantiated via a π-gesture is shown in Fig. 3. This might, by way of concrete example, be the opening (release) gesture for the tongue tip and the following bilabial closing gesture for the underlined consonants in the utterance "*ByeDad.Pickmeupatsix.*" But it should be emphasized that we consider our simulations to be generic in the sense that the modeled gestures are not meant to represent any specific consonant or vowel and their corresponding dynamical characteristics; rather, our gestural scores are meant to represent the activity of a set of abstract gestural point attractors, in order to keep our model as generalizable as possible at this early exploratory point.

The π-gesture's maximum level of activation is determined by the prosodic boundary strength (boundary strength could, for example, be viewed as the number of aligned domain edges). Prosodic gestures are notably different from constriction gestures (whose dynamics are that of point attractors in tract-variable space) in that π-gestures have no independent articulatory realization and are only realized *vicariously* via their effect on the constriction dynamics. In our earlier work (Byrd et al., 2000), we suggested that the π-gesture affected the stiffness parameter values of the concurrently active constriction gestures. The stronger the boundary, the greater the degree of stiffness lowering and, hence, local slowing that occurs for the boundary-adjacent articulatory gestures with which the π-gesture overlaps. In the work below, we show the shortcomings of this stiffness-modulation approach and propose an alternative—namely, that the π-gesture locally slows the clock that controls the timeflow of an utterance.

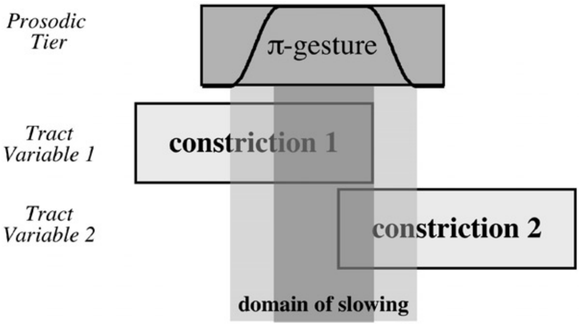


Fig. 3. A schematic gestural score for two gestures spanning a phrasal boundary instantiated via a π-gesture.

A traditional view might see prosodic events as linked to and *precisely coextensive with* segmental events. The π-gesture approach still anchors prosodic events (nonconstriction events) to segmental events (constriction events), but allows a flexible temporal relationship such that the prosodic events may overlap the segmental events and need not precisely share an edge with an individual gesture, segment, syllable or phrase. In contrast, for example, while Fujimura's (2000) C/D model also recognizes the influence of boundaries of different strengths, his boundary pulses are locked to phrase edges and also act on preplanned activation trajectories in a rather different fashion than our π-gestures.

The nature of the temporal coordination of π-gestures with constriction gestures (and intonational gestures) and the durational extent of π-gestures, which could conceivably be related to boundary strength, is a critical topic for continued research. It appears that maximal phrasal effects in the articulatory domain are observed close to the juncture but that lesser effects extend out from the juncture, perhaps in the same degree regardless of boundary strength (see Cambier-Langeveld, 1997). For example, Cho (2001, submitted) found that in an (unaccented) sequence of (lip) *opening*][*closing–opening* the temporal characteristics (i.e., lengthening) of the pre and postboundary opening were "much the same" (though only the closing gesture immediately following the boundary had a lower peak velocity and larger movement at stronger boundaries).

In summary, our proposal of π-gestures represents a first step in approaching a dynamical implementation of phrasal structure within Articulatory Phonology. We extend Browman and Goldstein's idea that the fundamental units in speech production are dynamical, serving both informational and implementational ends simultaneously, to include phrasal structure—in a *turtles-all-the-way-down* (*up*) approach. Further, the separation of prosodic units and articulatory units in the speech production process allows the prosodic pattern to be abstracted from any particular utterance and draws an equivalence between the same prosodic structure when it is realized on different utterances.[6] This approach entails several predictions for speech patterning, which are reviewed in the remainder of Section 4. Section 5 goes on to describe simulations of the kinematic consequences of π-gestures and evaluates the results in terms of experimental data from the literature. At the end of Section 5, we present an exploration of the model's parameter space. Finally, Section 6 concludes by considering outstanding challenges for the approach.

4.1.2. Implications of π-gestures as phrasal junctures

The π-gesture approach to implementing phrase boundaries in a dynamical model of speech production makes several predictions.

(1) It predicts that constriction gestures of whatever sort will undergo lengthening if they are active during the domain of π-gesture activation (i.e., if they overlap the π-gesture).

This prediction receives some support in Byrd (2000), which demonstrates that vowels seem to be lengthened just as consonants (Byrd & Saltzman, 1998; Beckman et al., 1992). It is also supported for lengthening following a boundary by Fougeron (2001) which found lengthening for /t, k, n, l, s, i, ɑ̃/ in French.

Further, based on the studies reviewed in Section 4 regarding boundary-adjacent articulatory lengthening of phrase initial gestures as well as phrase final gestures, we assume that, in English at

---

[6] We thank Mary Beckman for this observation.

least, the π-gesture spans the juncture to an extent and is anchored in some sense to it. Lastly, we make the parsimonious assumption of only a single π-gesture at a juncture. Given these assumptions, further predictions follow.

> (2) The degree of slowing will be greatest as the π-gesture's maximum activation is approached at the phrase edge.

This prediction receives some support in Byrd and Saltzman (1998) and Byrd (2000), who find that in [...C_1V_1#C_2V_2...] sequences, V_1 and C_2 are affected more than C_1 and V_2. Berkovits' (1993a,b, 1994) acoustic studies demonstrate a pattern of progressively greater lengthening within a final syllable such that final consonants lengthen more than preceding vowels. Such data are consistent with the hypothesis that peak activation of the π-gesture occurs at the edge of the phrase and activation wanes more remotely from this edge. This also fits comfortably with Edwards et al.'s observation that "final lengthening has a smaller, but still discernable, effect on the penultimate gesture of the phrase" (1991, p. 381).

Next, because π-gestures occur at junctures, phrasal lengthening effects are predicted to be local to junctures. That is, it is predicted that:

> (3) Effects will be limited to gestures near the domain edge and will not occur at gestures quite remote from it.

In her acoustic study, Cambier-Langeveld (1997) showed that while the amount of lengthening (corresponding here to π-gesture magnitude) is greater for larger boundaries, the *domain* of lengthening (corresponding here to π-gesture duration) does not change. Fletcher (1991) in interpreting her acoustic data from an experiment on final lengthening in French similarly comments that "[f]inal lengthening of accented syllables appears to involve a neutral lengthening isolated to the latter part of the syllable..." (p. 211). She contrasts this with accentual lengthening, which appears throughout the syllable. In their articulatory study, Edwards et al. (1991) describe final lengthening as a "targeted slowing down at a phrase edge...that is local to the final gesture in the phrase." While we suggest that a π-gesture will predominately effect edgemost constriction gestures, its effect will be felt on any of the gestures with which it is coarticulated; under the assumption that the π-gesture is anchored to the prosodic juncture, these will be gestures closest to the phrase edge. However, because π-gestures must be coordinated in time with constriction gestures, a more complex range of empirical behavior is in principle possible (assuming that our currently parsimonious theoretical assumptions are relaxed, obviating prediction three). For example, this theoretical approach does not preclude the possibility that in some languages the π-gesture might be attracted, perhaps due to stress, to a syllable earlier in the phrase-final word (see for example Shattuck-Hufnagel & Turk, 1998). In this case, the maximal slowing effect will be observed on these gestures since they are now in a coproduction relationship with the π-gesture.

Next, given the above assumptions, this approach predicts that:

> (4) The dynamical source of intragestural duration and intergestural timing effects is the same at right and left domain edges (i.e., phrase-final and phrase initial edges) (though these effects may well be different in degree or in kinematic characteristics depending on the coordination of the π-gesture with constriction gestures).

While not many studies have simultaneously compared the articulation of segmentally identical stimuli in initial and final positions, Byrd and Saltzman (1998), Byrd et al. (2000), and Cho (2001, submitted) provide some support for this in that articulatory lengthening was observed for both phrase-initial and phrase-final positions. It is noteworthy that the temporal changes, which we hypothesize to arise from π-gestures, can have indirect consequences on spatial magnitudes of constriction gestures. We return to this topic in Section 5.3.1.

Lastly,

> (5) Boundaries of different strengths (or different categories of boundaries: intonational phrase, intermediate phrase, etc.) are also expected to not be distinct in *type* of effect, only in degree.

Again, Byrd and Saltzman (1998) find some support for this in that lengthening occurred at both large and small boundaries in their study, though in different amounts. Similar results are reported by Cho (2001, submitted).

5. Modeling prosodic gestures

In this section we evaluate the utility of the prosodic gesture model of boundary adjacent lengthening by quantitatively modeling the effects of π-gestures on articulatory patterning and evaluating the simulated results relative to the kinematic findings of larger, longer, and less overlapped gestures. First, we discuss the technical assumptions underlying the simulations. Next, we briefly demonstrate why stiffness-lowering as a mechanism to produce final lengthening, such as the general approach explored by Byrd and Saltzman (1998) and Beckman and Edwards (1992), may be less than ideal for capturing the boundary-adjacent phenomena that have been observed. Finally, we present an account of π-gestures implemented via clock modulation that yields a more adequate account of the transgestural effects observed at boundaries.

5.1. Simulation foundations

In the discussion to follow, we focus on the primary oral constriction and release gestures involved in the production of segments near a boundary, though as we point out above, all tract variable gestures coproduced with the π-gesture are predicted to be affected. The simulations implement (1) two constriction gestures in the same or different tract-variables, (2) the neutral gestures associated with the releases of these constriction gestures, and (3) an overlapping prosodic gesture defined within a prosodic tier distinct from the model's set of constriction-related tract variables and centered between the two constriction gestures. In all simulations presented, we have assumed for purposes of simplicity, a one-to-one mapping of articulators to tract variables, i.e., there is only one articulator associated with each tract variable. Under this assumption, the system's equations of motion for a single constriction gesture defined in a given tract-variable are as follows:

$$\ddot{x} = \ddot{x}_c + \ddot{x}_d \tag{1a}$$

where

$$\ddot{x}_c = -a_c(t)b_c\dot{x} - a_c(t)k_c[x - a_c(t)x_{o,c}] \tag{1b}$$

$$\ddot{x}_d = (1 - a_c(t))[-b_d\dot{x} - k_d(x - x_{o,d})] \tag{1c}$$

In Eq. (1), x, and $\dot{x}$ denote the tract variable's position and velocity, respectively, and $\ddot{x}$ denotes the total tract-variable acceleration; $\ddot{x}_c$ is the acceleration due to the constriction gesture; and $\ddot{x}_d$ is the acceleration due to the tract variable's neutral (d for "default") gesture. The role of the neutral gesture in the full task dynamic model (Saltzman & Munhall, 1989) is to bring the vocal tract articulators back to a neutral position when they are not otherwise being actively controlled by a constriction gesture. (Without a neutral attractor, articulators could simply be "stuck" in a constricted posture if not called away by another gesture.) Given our assumption in the present simulations of a one-to-one mapping between articulators and tract variables, the neutral gesture is defined directly at the tract-variable level. In these simulations, the release of gesture 1 is governed by the neutral attractor, whose activation strength trajectory is simply the complement of the preceding constriction's activation strength. (While in the current implementation, releases are governed by the neutral attractor and are affected by a π-gesture in the same way as active constriction gestures, in fact, ultimately, releases may more appropriately be variously defined—e.g., as neutral attractors, as active gestures (Browman, 1994), or as uncontrolled aerodynamic events—and ultimately, these events might behave differently under the influence of π-gestures.)

Gestural activation for the constriction gesture, $a_c(t)$, controls the approach of the current parameter values for both constriction and neutral gestures (see Eq. (1)) toward their underlying or canonical values using multiplicative gating. For example, if an active constriction gesture's underlying stiffness value is k_c, the tract-variable's current working value is $k_c^* = a_c(t) \cdot k_c$. Similarly, if the neutral gesture's underlying stiffness value is k_d, its current working value is $k_d^* = (1 - a_c(t)) \cdot k_d$.

Until relatively recently, activation trajectories have been modeled as step functions (i.e., as on/off "boxes" in gestural scores; see Fig. 1). However, in our present work we use activation functions with gradual rises and falls,[7] defined by Eq. (2) and illustrated in Fig. 4 (see Byrd & Saltzman, 1998; also Kröger et al., 1995):

$$a(t) = \begin{cases} 0, \text{ if } (t < t_{on,1}) \\ 0.5\{1 - \cos(\omega_r[t - t_{on,1}])\}, \text{ if } (t_{on,1} \leqslant t < t_{on,2}) \\ 1, \text{ if } (t_{on,2} \leqslant t < t_{off,1}) \\ 0.5\{1 + \cos(\omega_f[t - t_{off,1}])\}, \text{ if } (t_{off,1} \leqslant t < t_{off,2}) \\ 0, \text{ if } (t \geqslant t_{off,2}) \end{cases} \tag{2}$$

As can be seen in Fig. 4, an activation waveform is described as a half-cosine rise followed by a flat plateau followed by a half-cosine fall. The half-cosine function is a simple monotonically

---

[7] It has been recognized for quite some time that using step-function waveshapes to define gestural activations (or GO signals, forcing functions, or equilibrium point trajectories) is an oversimplification (see, e.g., Bullock and Grossberg, 1988; Coker, 1976; Kröger, Schröder, & Opgen-Rhein, 1995; Ostry, Gribble, & Gracco, 1996; Perrier, Ostry, & Laboissière, 1996). For example, activation step functions are unable to generate gestural velocity profiles with appropriate degrees of temporal asymmetry (e.g., Bullock & Grossberg, 1988; Byrd & Saltzman, 2000).

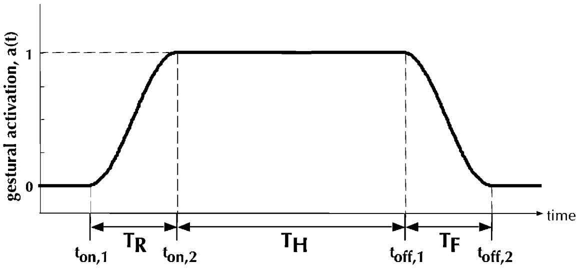


Fig. 4. The activation function used in the gestural simulations to follow.

increasing/decreasing function characterized by a single frequency parameter determining the rise/fall time of the leading/trailing edge of the waveform.

As outlined above, we model the influence of a prosodic boundary by incorporating a prosodic gesture within a prosodic tier that is distinct from the set of tract-variable tiers currently used to define constriction gestures in the model. Each prosodic gesture is affiliated with a phrase edge. Since we wish to examine the effects of π-gestures on both pre and postboundary gestures local to the juncture, we define a π-gesture whose peak is centered "at the boundary," i.e., centered between the constriction gestures, and that, for simplicity's sake, has symmetrical leading and trailing edges. (In Section 5.4, we explore the consequences of relaxing these assumption regarding a π-gesture's shape and relative timing.)

For parsimony's sake and without evidence to the contrary, we choose to adopt the same activation waveform shapes for π-gestures as for constriction gestures, as defined in Eq. (2). The choice of a half-cosine function for the rise and fall portions of activation waveforms (e.g., Byrd & Saltzman, 1998) has been a conservative one, and has allowed us to simulate details of gestural kinematics that would have been impossible to generate using step-function waveshapes (Byrd & Saltzman, 1998).

5.2. Transgestural stiffness modulation

Before turning to simulations of clock slowing due to a π-gesture, it is useful to examine the primary proposal currently in the literature for modeling boundary-adjacent lengthening within the task dynamic framework. This is the stiffness modulation account suggested by Beckman and colleagues (e.g., Beckman & Edwards, 1992) and Byrd & colleagues (e.g., Byrd et al., 2000; Byrd, 2000). We will see that while adequate in certain respects, this approach is less than entirely satisfactory.

In the stiffness-modulation simulation, prosodic gestures affect articulatory movements by affecting the stiffness parameter values of all concurrently active constriction and neutral gestures. In particular, for each constriction and neutral gesture that is co-active with a π-gesture, stiffness (k_c or k_d) is reduced (to k_c^* or k_d^*, respectively) in proportion to the strength of the π-gesture as follows:

$$k^* = (1 - \alpha \cdot a_\pi) \cdot k \tag{3}$$

where α is the boundary or prosodic strength of the π-gesture ($0 \leqslant \alpha \leqslant 1$), and a_π is the activation of the π-gesture ($0 \leqslant a_\pi \leqslant 1$).

Fig. 5 illustrates the movement trajectories for two overlapping gestures with *no* π-gesture, i.e., phrase-medially. The heavy arrows in the bottom panel indicate the duration of the opening movement for gesture one and the closing movement for gesture two, measured between velocity zero-crossings.

In a second simulation, shown in Fig. 6, we see the effect of a stiffness-modulating π-gesture on the movement trajectories for two gestures spanning a phrase boundary. In this simulation, the π-gesture is centered between the two constriction gestures and has an arbitrary prosodic strength of 0.3. Again, the heavy arrows indicate movement durations adjacent to the boundary.

The inset on the right of Fig. 6 demonstrates that, as expected, boundary adjacent lengthening occurs under the influence of the π-gesture. Also, a comparison of Figs. 5 and 6 shows that the π-gesture reduces the peak velocities of these gestures, as expected with modulating stiffness. Finally, notice that the slowing is greater "phrase" finally than initially, a result that is in accord with the empirical literature which focuses on phrase-final effects. In our simulations, this pattern of asymmetric lengthening is due to the fact that the stiffness-lowering π-gesture does not change the activation timing of constriction gesture onsets or their offsets (these offsets coincide with onsets of the associated neutral release gestures; see Eq. (1). As a result, stiffness modulation has only a slight effect on the observed onset and offset of phrase-initial closing associated with the second constriction gesture. The onset of this gesture is caused by the target shift away from

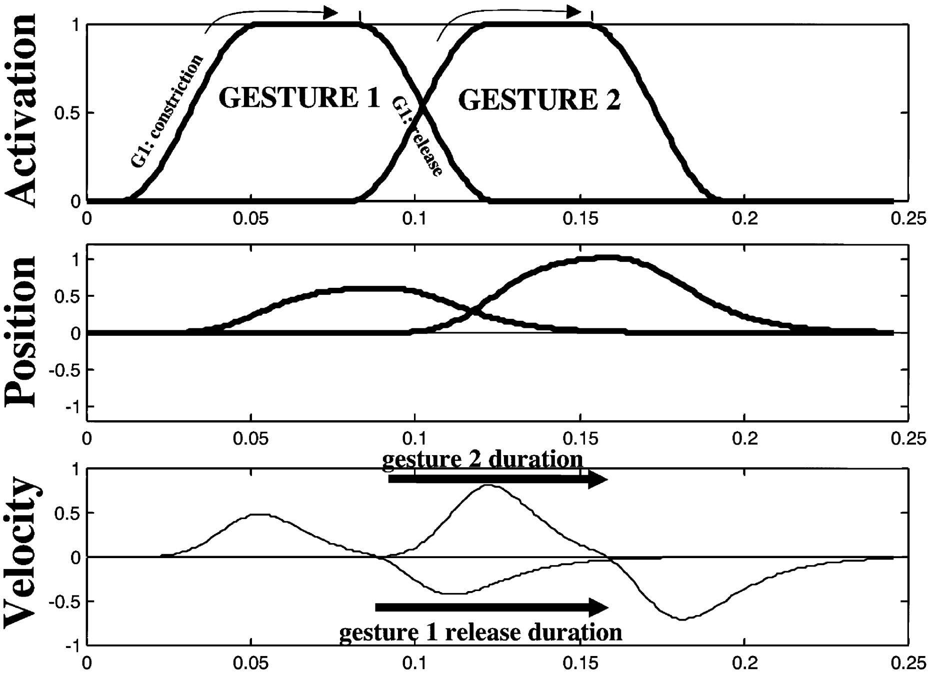


Fig. 5. Movement trajectories for two overlapping gestures with *no* π-gesture, i.e., phrase-medially. Heavy arrows in the bottom panel indicate the duration of the opening movement for gesture one and the closing movement for gesture two, measured between velocity zero-crossings. (For display purposes, all velocity trajectories have been multiplied by scaling factors of 0.025.)

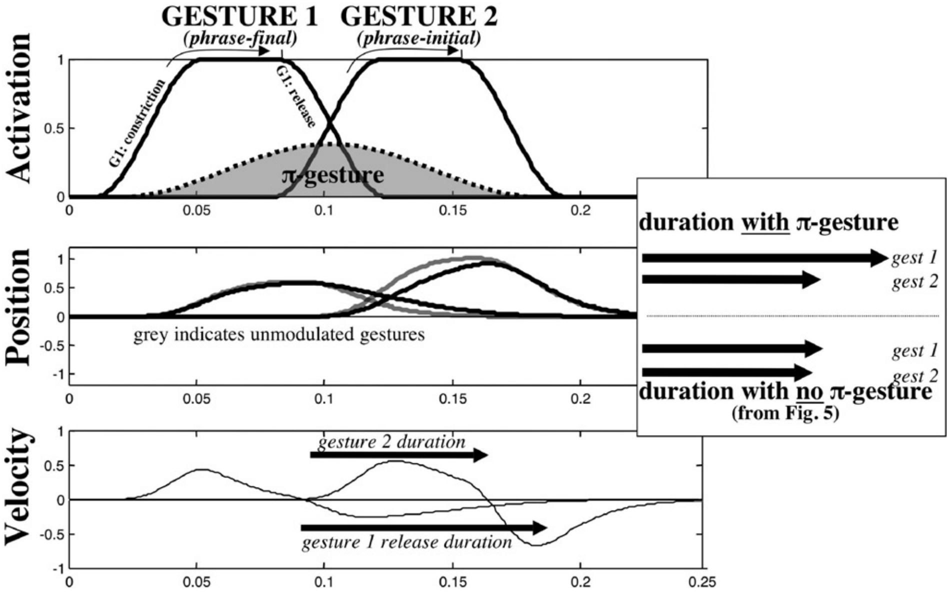


Fig. 6. Movement trajectories for two gestures spanning a phrase boundary. The π-gesture is centered between the two constriction gesture and has an arbitrary prosodic strength of 0.3. Heavy arrows indicate movement durations adjacent to the boundary. (For display purposes, all velocity trajectories have been multiplied by scaling factors of 0.025.)

neutral position, that is, the gesture's activation onset; the offset of raising is caused by the onset of the target shift back to the neutral position, that is, the gesture's activation offset or, equivalently, the activation onset of the associated neutral gesture. Had the activation interval of gesture 2 been extended long enough for this phrase-initial closing to reach its target (i.e., had no truncation been involved), then lengthening of this gesture due to its lowered stiffness would have been comparable to that of gesture 1's phrase-final neutral release. This release is, in contrast, substantially elongated since its endpoint is defined (purely kinematically) only as its time of return to neutral position, which proceeds unimpeded. Finally, the small delay effect on the observed kinematic onset (defined from the velocity trajectory) of the phrase-initial closing (see Table 1) also contributes to the lengthening of the gesture 1 release.

5.2.1. Variation in relative timing at phrase edges

A second change that has been reported at phrase boundaries is that the relative timing of gestures becomes less overlapped. This has been shown for gestures spanning a boundary in Byrd et al. (2000), for gestures initial in a phrasal domain by Hardcastle (1985), Jun (1993), and Keating, Cho, Fougeron, and Hsu (1999), for gestures early in a phrasal domain by McClean (1973), and for gestures final in a domain (for at least one subject) by Edwards et al. (1991). Stiffness scaling, however, does *not* account for changes in relative timing, as noted by Edwards et al. (1991). In fact, in our simulations, the amount of gestural overlap actually increases with the stiffness scaling associated with a π-gesture (see Table 1). This is due primarily to the fact that stiffness scaling does not influence the timecourse of gestural activations that govern the switching between successive gestural targets.

Table 1
Simulation results for two gestures of different tract variable—result column one includes no π-gesture, column two includes a π-gesture implemented via stiffness scaling, column three includes a π-gesture implemented via time scaling (i.e., clock slowing)

Two gestures, different tract variables	No π-gesture ($\alpha = 0$)	π-gesture using stiffness scaling ($\alpha = 0.4$)	π-gesture using time scaling ($\alpha = 0.3$)
Lowering onset time (s)	0.0902	0.0932	0.0976
Lowering offset time (s)	0.1632	0.1854	0.1792
Lowering duration (s)	0.0730	0.0920	0.0816
Raising onset time (s)	0.0942	0.0962	0.1062
Raising offset time (s)	0.1572	0.1622	0.1802
Raising duration (s)	0.0630	0.0660	0.0740
Kinematic overlap ("C2 inside C1"): $\frac{\text{(lowering offset time} - \text{raising onset time)}}{\text{lowering duration}}$	0.9452	0.9675	0.8946
Kinematic overlap ("C1 inside C2"): $\frac{\text{(lowering offset time} - \text{raising onset time)}}{\text{raising duration}}$	1.0952	1.3515	0.9865

For the phrase final gesture, duration was measured from time of lowering onset (first sample less than –1.0 velocity units) to lowering offset (first sample greater than –1.0 velocity units); phrase initial gestural duration was measured from time of raising onset (first sample greater than +1.0 velocity units) to raising offset (first sample less than +1.0 velocity units).

Clearly, an appropriate simulation of phrase edges would ideally generate changes in both duration and relative timing, and we hypothesized that both types of prosodic effects might be captured by *time slowing at the gestural control level*—i.e., slowing the timecourse of gestural activation. This is because the transgestural process of slowing a hypothesized central clock would have both intragestural and intergestural timing consequences.[8] This implements the suggestion of Edwards et al. (1991) that "final lengthening is like a localized change in speaking tempo…[that] cannot be equated directly with the specification of stiffness." (p. 369)

5.3. Simulations of clock slowing

In the following simulations, we move from a π-gesture whose transgestural effect is on gestural stiffness parameters to a π-gesture whose transgestural effect is on the timecourse of gestural activations—i.e., the π-gesture acts to slow the speed of a hypothesized underlying central clock whose rate of timeflow determines the local utterance rate.[9] Thus, whereas in our

---

[8] Byrd and Saltzman (1998) explored the effects on gestural velocity profile asymmetries of prosodically-conditioned variations in the rise times of gestural activations (in conjunction with variation of gestural stiffness values). These simulations did not invoke clock-slowing, did not link slowing to the time-course of π-gestures, and only focused on the intragestural consequences of a single gesture's activation rise-time. In contrast, our current paper links clock slowing explicitly to the activation timecourse of the π-gesture and deals with the effects of such prosodically-modulated slowing on the intragestural (amplitude and duration) and intergestural (relative timing) kinematic properties of a pair of (simulated) gestures.

[9] These simulations were presented in preliminary form in Saltzman and Byrd (2000b).

stiffness-modulation simulations all gestural activations ($a(t)$) were functions of unscaled "standard" time (t), in our clock-slowing simulations all activations ($a(\tau)$) are functions of the scaled clock time (τ). In these simulations, the π-gesture modulates clock time as follows:

$$\dot{\tau} = \frac{\mathrm{d}}{\mathrm{d}t}\tau = (1 - \alpha a_{\pi}) \tag{4}$$

where t is unscaled time whose rate of change, $\dot{t}$, is by definition equal to 1.0; α is the prosodic or boundary strength of the π-gesture; and a_{π} is the π-gesture's activation.

From this equation it can be seen that when the π-gesture is inactive, $\dot{\tau} = 1$, clock time is the same as standard unscaled time. However, when a π-gesture is active, clock time is slowed in proportion to the π-gesture's activation level, resulting in the transgestural temporal "stretching" of all associated activation waveforms (constriction, neutral, *and* π-gestures). In Fig. 7, we see the difference between unscaled and scaled (i.e., slowed) time evolution (upper panel) as well as the effects of time slowing on the timecourse of a single gestural activation (lower panel).

Thus, a π-gesture whose effect is to slow the rate of time ticks of a central clock will induce transgestural and local slowing of the time course of an utterance. It is worth noting that modulating clock rate is equivalent to modulating the instantaneous frequency parameter of an underlying oscillatory timekeeper.[10]

Fig. 8 presents the simulation results for two constriction gestures spanning a phrase boundary where the phrase boundary is realized by a π-gesture that causes central clock slowing or time-stretching of the gestural activations.

The effect of this clock slowing is to stretch the gestural activations and, consequently, the movements. Note that this stretching is not uniform; the amount of slowing increases as a π-gesture's activation reaches and leaves its maximum. If the π-gesture is centered at the boundary, this means that stretching will be greatest closest to the boundary. Clock slowing causes gestures to be longer (increased duration) and slower (lower peak velocities). Such a pattern is reported for phrase initial closing gestures by Cho (2001, submitted). However, it is important to be mindful of the fact that changes in a gesture's *displacement* can also affect peak velocity; bigger gestures are often faster. This means that peak velocity values may or may not be observed to be lowered at boundaries depending on the complex interplay of temporal and spatial changes, and this seems to agree with the generally inconsistent findings regarding prosodically governed peak velocity changes. For example, Cho (2001, submitted) finds no peak velocity changes or displacement changes in a […bV][bV…] sequence for either of the pre and postboundary lip openings, but slower and larger closings immediately after the boundary (we refer only to Cho's unaccented data). However, the findings of Edwards et al. (1991) for lip opening, show faster and larger

---

[10] Viewed in this way, clock modulation can be interpreted as parameter-dynamic process. Note, however, that modulation of clock rate is not equivalent to modulation of delta-t in a discrete simulation. Delta-t modulation would simply time warp the trajectories of all system variables—activation, tract, and articulator. The result would be a temporal stretching of all these trajectories, with no changes to any of their magnitudes. One of us (Saltzman) had used such delta-t modulation in a non-prosodic context (Saltzman et al., 2000). Our current clock rate modulation model serves to stretch only the system's activation trajectories, without changing their spatial characteristics. Since the activation trajectories serve effectively as forcing functions that drive motions of the tract and articulator variables, temporal changes in activation trajectories result in both temporal (e.g., gestural durations and intergestural timing) and spatial (maximum amplitudes and peak velocities) changes in tract and articulator variable trajectories.

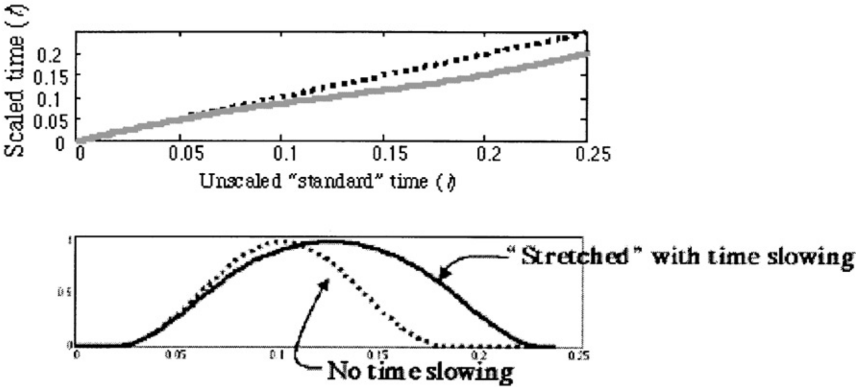


Fig. 7. Upper panel: Unscaled time evolution (dotted) and time evolution with time scaling (i.e., slowing) (solid). Bottom panel: The timecourse of a single gestural activation as it is stretched due to time slowing.

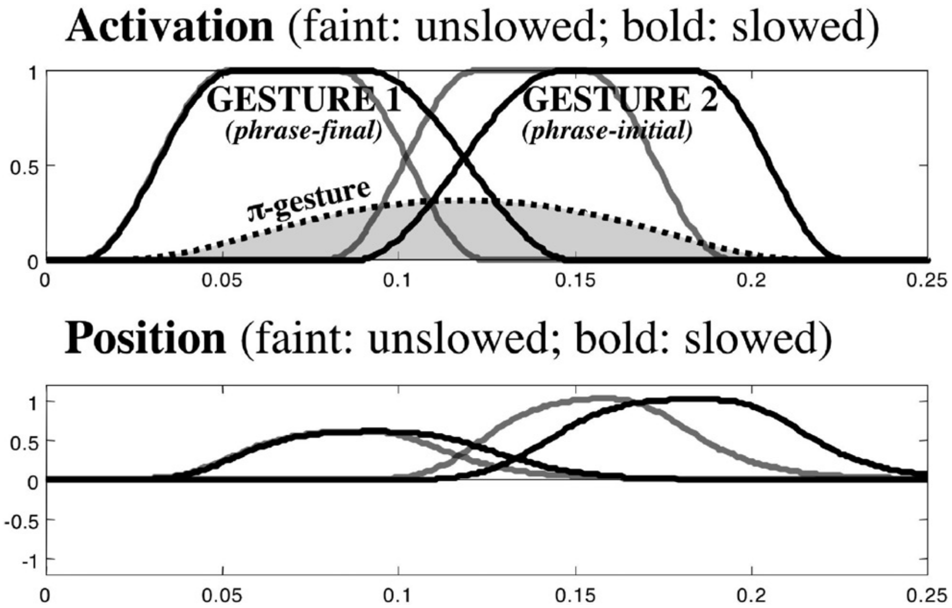


Fig. 8. Simulation results for two constriction gestures spanning a phrase boundary where the phrase boundary is realized by a π-gesture (strength=0.3) that causes central clock slowing or time-stretching of the gestural activations. (For display purposes, all velocity trajectories have been multiplied by scaling factors of 0.025.)

openings phrase finally for most subjects (again in the case of unaccented syllables). Our own experimental data have patterned variously in this regard, tending at boundaries to show slower peak velocities for comparable displacements but faster peak velocities when displacements increase, as they often do. Interestingly, Cho's data are for full vowels while Byrd and Saltzman (1998) and Edwards et al. (1991) uses reduced vowels. This suggests that in Cho's study the preboundary opening may have been earlier or more removed from the π-gesture, and consequently less affected by it.

Significantly, in our simulations boundary-induced clock slowing not only increases gestural durations but also results in a decrease in gestural overlap (see Table 1). Such overlap changes are reported in Byrd et al. (2000). Decreases in overlap likely contribute to the previous findings (Edwards et al., 1991), noted above, of larger displacements for preboundary openings involving reduced vowels, as the jaw and lips would have more time to lower before being required to turn around for the postboundary lip closings. Of course, the separate (single articulator, single tract variable) gestures in our simulations have no common articulators to be affected by overlap changes.

To summarize thus far, implementing the π-gesture as clock slowing enables us to unify the two empirically reported phrase edge phenomena of lengthening and lessened overlap under a single dynamical "umbrella."

5.3.1. *Spatial magnitude effects at phrase edges*

Gestures have been reported to be spatially larger in phrase initial position. This phenomenon has been dubbed "initial strengthening" in work by Fougeron and Keating (1997), Cho and Keating (2001), Keating et al. (1999). More generally these authors have used the term initial strengthening to refer to increasing the perceptual saliency of prosodic boundaries. In fact, various phenomena have been proposed to fall under the rubric of initial strengthening. These include increased lingua-palatal contact for lingual consonants (Fougeron & Keating, 1997; Keating et al., 1999), longer VOTs (Jun, 1993, 1995; Cho & Jun, 2000), lower RMS [h]s (Pierrehumbert & Talkin, 1992), and more lip rounding in rounded vowels (van Lieshout, Starkweather, Hulstijn, & Peters, 1995) in phrase initial positions. These findings and their apparent common sensitivity to phrase boundaries, encourage us to consider whether modulation in the temporal domain may have significant consequences in the spatial domain. Fougeron and Keating (1997) and Cho and Keating (2001) suggest that, at least in certain languages, increased lingua-palatal contact in phrase initial position might result from a lack of undershoot (in the sense of Lindblom, 1963) in that position due to increased segmental duration. While we do not suggest that *all* "initial strengthening" phenomena are necessarily the result of temporal changes related to phrasal structure, it seems worthwhile to consider the temporal domain when attempting to understand spatial variability. Certain instances, though perhaps not all, of initial strengthening might have their foundation in clock slowing. In fact, we have already seen that π-gestures can lessen gestural overlap, possibly accounting for the longer VOTs in phrase initial position. Regarding the linguapalatal contact data for French presented in Fougeron (2001), Fougeron notes that there is more contact when the phrase initial segment is preceded by a large final lengthening, and she suggests that the lengthening and strengthening mechanisms may be related, or, in fact, one and the same. Also, the degree of phrase-initial lengthening of closure duration in French and Korean also seems to be correlated with amount of linguapalatal contact during stops (Fougeron, 2001; Cho & Keating, 2001), though this correlation is weaker in English (Fougeron & Keating, 1997).

Fig. 9 illustrates the results of a simulation in which two constriction gestures within the same tract-variable are defined in "phrase" initial position; i.e., the π-gesture initiates the sequence rather than intervenes between the two constriction gestures. The constriction gestures undergo

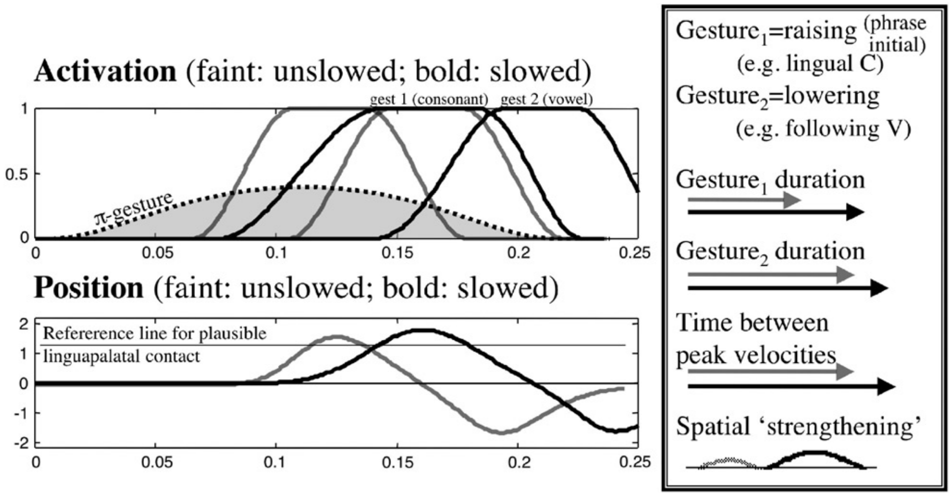


Fig. 9. Simulation results in which two constriction gestures (a closing followed by a widening) within the *same* tract-variable are defined in ‘‘phrase’’ initial position, i.e., the π-gesture initiates the sequence rather than intervening between the two constriction gestures. The constriction gestures undergo clock-slowing under the influence of the π-gesture (strength=0.4).

clock-slowing under the influence of the π-gesture. The constriction gestures in these simulations are meant to emulate a lingual consonant and vowel gesture, as this was the type of sequence examined by Keating, Fougeron, and colleagues. Accordingly, gesture 1 has a positive target position (closed) and gesture 2 has a negative target position (open). [Since both gestures are now defined within the *same* tract-variable (emulating the tongue dorsum constriction), their intrinsic parameters (i.e., target, stiffness, and damping) are therefore *blended* such that the resultant, ongoing tract-variable parameters become activation-weighted averages of these intrinsic parameter values (see Saltzman & Munhall (1989) for further details on parameter blending).] This might, by way of concrete example, roughly be the gestures for the underlined consonants in the utterance ‘‘*ByeDad.Gottago.*’’ (though recall these simulations are for single-articulator gestures, which [g] and [a] are not).

As shown in Fig. 9 and Table 2, the simulated constriction gestures in phrase initial position are longer in duration and less overlapped (as indicated, for example, by the lengthened time between peak velocities) in the presence of a π-gesture. As a further consequence of this overlap change, the π-gesture causes the (consonantal) closing gesture to become larger in spatial magnitude and to increase its duration and degree of closure or contact (as indicated by the magnitude of movement above the arbitrary horizontal reference in Fig. 9).

Thus it seems that transgestural perturbations of clock rate due to a π-gesture that locally slows time flow in an utterance can result in appropriate kinematic changes not only in the temporal domain but also in the spatial domain. In further exploration of initial strengthening phenomena it will, we believe, prove useful to examine concurrently both gestural magnitude and the timecourse of the actual associated kinematic movement trajectories from movement

Table 2
Simulation results for two gestures sharing the same tract variable, both with and without the presence of a π-gesture

Two gestures, same tract variables	No π-gesture ($\alpha = 0$)	π-gesture using time scaling ($\alpha = 0.4$)
Raising onset time (s)	0.0732	0.0932
Raising offset time (s)	0.1252	0.1612
Raising duration (s)	0.0520	0.0680
Lowering onset time (s)	0.1252	0.1612
Lowering offset time (s)	0.1934	0.2413
Lowering duration (s)	0.0682	0.0801
Time between peak velocities	0.0400	0.0480

For the first phrase-initial gesture (raising), duration was measured from time of raising onset (first sample greater than +1.0 velocity units) to raising offset (first sample less than +1.0 velocity units); for the second phrase-initial gesture (lowering), duration was measured from time of lowering onset (first sample less than –1.0 velocity units) to lowering offset (first sample greater than –1.0 velocity units).

onset to target attainment. This approach should allow researchers to distinguish phrasal clock-slowing effects from other constriction-related changes, such as increased gestural magnitude in initial position or under accent, or weakening/reduction in particular prosodic environments.

5.4. Exploration of model parameters

In this final section on the clock-slowing simulation of π-gestures, we present a brief exploration of the model parameters in order to convey a sampled range of the qualitative kinematic variation generated by the model that can serve as a foundational reference for further exploration of clock-slowing. In Fig. 10, examples A and B serve as reference for the examples in the rest of the figure. Example A shows the constriction activation trajectories and gestural state (position and velocity) trajectories when there is no π-gesture (as in Fig. 5). Example B shows the same trajectories when a clock-slowing π-gesture is added (a symmetrically shaped π-gesture with an arbitrary strength of 0.3 centered between two constriction gestures). In the remaining parts of the figure, the left column shows an early (C) and late (D) temporal alignment of the π-gesture with the constriction gestures. The middle column shows a weaker (E) and stronger (F) π-gesture centered between the two constriction gestures. The right column shows a centered, symmetric π-gesture with shortened rise and fall times and a plateau at maximum activation (G), followed by a positively (H) and negatively (I) skewed π-gesture with no plateaus centered between the two constriction gestures. For all examples, activations and gestural kinematics are shown.

Fig. 10 illustrates several interesting properties of the model's clock slowing π-gestures. Most obvious is the fact that stronger π-gestures (F) induce more gestural slowing, more peak velocity reduction, and longer intergestural intervals than weaker π-gestures (E). Likewise, when a constriction is coproduced squarely within the domain of a π-gesture it is more affected than when only the trailing edges are coproduced; i.e., the left gesture in C and the right gesture in D are most strongly affected. Additionally, the overall phase (time) shift generated by a π-gesture, evaluated

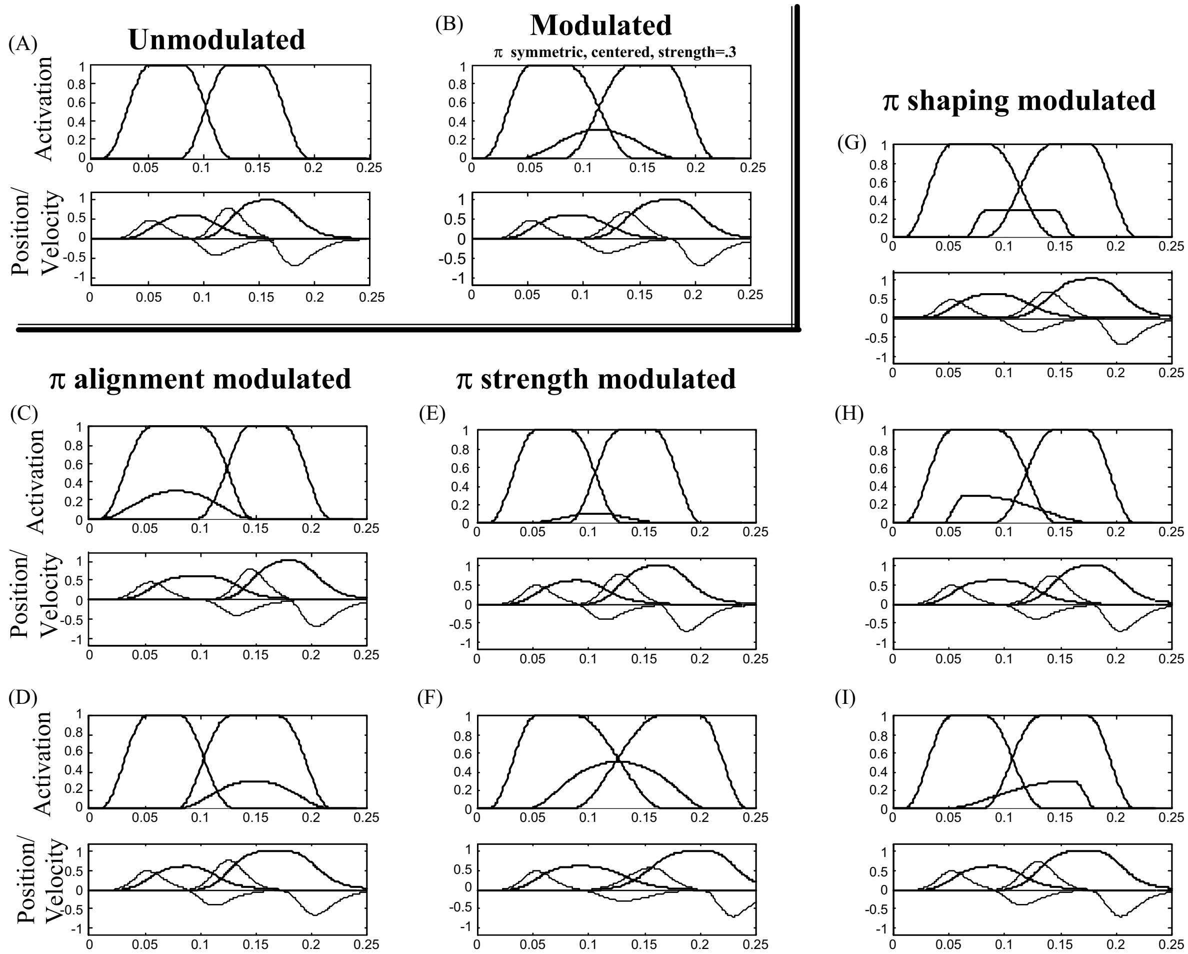


Fig. 10. (A) An unmodulated sequence of two constriction gestures (for reference); (B) a symmetric π-gesture with an arbitrary strength of 0.3 centered between two constriction gestures; (C) an early temporal alignment of a symmetric π-gesture (strength=0.3); (D) a late temporal alignment of a symmetric π-gesture (strength=0.3); (E) a weaker (strength=0.1) symmetric π-gesture centered between the constriction gestures; (F) a stronger (strength=0.5) symmetric π-gesture centered between the constriction gestures; (G) a symmetrical, centered π-gesture (strength=0.3) with a plateau at maximum activation; (H) a positively skewed π-gesture (strength=0.3) centered between the two constriction gestures; and (I) a negatively (bottom) skewed π-gesture (strength=0.3) centered between the two constriction gestures. For each condition, activations are shown in the top panel with the π-gesture activation shaded, and gestural kinematics are shown in the bottom panel, with position and velocity (thinner line) overlaid.

as the time between the beginning of gesture 1 and the end of gesture 2, reflects the area under the π-gesture's activation curve. Since the areas under these curves are comparable in B, C, D, G, H, and I, the resulting overall time shifts are the same; these shifts are greater than in A and E, and less than in F. Finally, the relative amounts of temporal change "allocated" across the constriction gestures in the π-gesture's temporal domain vary with location of the center of gravity of the π-activation wave with respect to the constriction gestures. This can be a function of either the π-gesture shape (e.g., skew) or its relative timing with respect to the constriction gestures; e.g., similar effects are observed in C and H, and in D and I. Of course, the magnitude of this local effect is proportional to the π-gesture's ongoing strength. So activation peak strength (i.e., boundary strength), center of gravity location, and area under the curve are the major determinants of π-gesture effects.

5.5. *Summary of π-gesture simulations*

Local slowing of a central clock appears to be a plausible way to capture prosodically-driven shaping of articulatory behavior. Unlike stiffness modulation, which only affects gestural durations, clock rate modulation generates several experimentally observed prosodic effects: gestural lengthening, reduced intergestural overlap, and increases in spatial magnitude.

6. Future directions

The clock mechanism we have proposed can be invoked to capture speech rate differences (i.e., "global" timing) associated with sociolinguistic, dialectal, or individual characteristics. Without further constraint, however, local modulation of clock rate could be engendered by any number of segmental and/or prosodic factors. At this point in the development of the model, we seek to make and explore conservative assumptions regarding such modulation. In this spirit, we have restricted our application of this mechanism to the domain of boundary-related phenomena. Although a clock-modulation mechanism has been invoked by others to deal with the consequences of segmental context on the "temporal microstructure" of syllables, e.g., vowel length before voiced and voiceless consonants (Port & Cummins, 1992), we prefer to look first to an account of such phenomena at the level of intergestural timing (i.e., relative phasing of gestures) and/or intrinsic gestural dynamics. Local modulation of utterance timeflow is, in our view, limited to characterizing prosodic durational effects. Thus, accentual lengthening may also prove to be a type of local temporal modulation best handled by clock slowing, possibly accompanied by spatial expansion.

Future simulation and empirical work on prosodic gestures within this framework should investigate the timecourse and coordination of the π-gestures themselves. Over what interval does phrase-edge slowing obtain? That is, how remote from a phrase boundary is slowing observed to initiate and end? During its domain, what π-gesture activation shaping best captures the slowing down and speeding up of articulatory patterning? The coordination of the π-gestures with segmental and intonational events must be understood. Are the articulatory slowing effects that have been observed at phrase edges limited to that position, or can the phonological structure (e.g., stress) or semantic structure (e.g., informational content) of the phrase attract the slowing effect away from the very edge of the phrase earlier or later? Variation in the specific duration and coordination of π-gestures might account for certain types of language- and speaker-dependent differences that have been observed in prosodic realization.

Additionally, how should pausing be handled within this computational framework? Of course, the articulatory behavior during pauses of various sorts should be examined experimentally in order to evaluate the extent to which π-gestures might generate these intrapausal behaviors. In the cases where a hiatus in the events of speech production is substantial, might two π-gestures be present—one delimiting the end of the first informational unit and one initiating the next? Such a situation might be explored in terms of the effective "splitting" of one π-gesture into two in the face of limits on the malleability of the clock controlling the flow of speech events. Finally, it is possible that a cessation or initiation of speaking requires the stopping or starting of the clock. Such stops and starts are typically coincident with the edges of informational units and, therefore, would typically be accompanied by a π-gesture under our account.

Further, it remains an open question as to whether activation levels or boundary strengths of π-gestures are gradient or categorical in nature. Do the articulatory patterns at phrase boundaries suggest the existence of small set of categorically distinct boundary types or do they support an analysis in which the strength of disjuncture between phrases can be understood as gradient in degree?[11] If boundaries are defined gradiently, one would want to explore whether phonological processes can be understood to take place at all junctures under/above a particular strength, rather than at a single or arbitrary group of phonological phrase category types?

Finally, modeling of the hypothesized central clock must become more sophisticated and biologically realistic, for example, involving an ensemble or population of neural oscillators. These investigations and concomitant computational modeling of their results will provide a profile of the manner in which multi-gesture articulatory patterning is shaped by prosodic context.

7. Conclusion

Understanding the organization of units of speech production as a function of the informational composition of utterances is critical to developing a unified account of how abstract linguistic structure is communicated in spoken language. Articulatory patterning at phrase edges is one example of how the surface expression of phonological units can vary in a linguistically principled way. We have examined the relation between phrasal structure and the control and coordination of articulation within a dynamical systems model of speech production. We suggest that boundary-related durational patterning can be conceived of as resulting from prosodic gestures or π-gestures, which share much with the familiar dynamical description of constriction gestures. Pi-gestures, however, function purely to stretch or shrink the local temporal fabric of an utterance. This local modulation of the "clock-rate" that controls the temporal unfolding of an utterance is such that the clock slows increasingly as the boundary is approached and speeds up again as the boundary recedes. This modulation of local speaking rate gives rise to kinematic effects that have been observed at phrase edges, and endows phrases with a degree of temporal elasticity. Viewing phrase boundaries as warping the temporal fabric of an utterance represents a promising confluence of the fields of prosody and of speech dynamics.

---

[11] For an interesting discussion of boundary strength, see Swerts (1997), who further comments that prosodic labelers "may be confused about the exact spot at which a boundary occurs, but they may agree that within a certain "region" there is [a] change of information unit" [p. 520]—a phenomenon not incompatible with a prosodic gesture approach in which the junctural element has a span in time.

Acknowledgements

The authors gratefully acknowledge the support of NIH grants DC-03172 and DC-03663, and thank attentive audiences at the UCLA Linguistics Department, the UCSB Linguistics Department, the Eighth Conference on Laboratory Phonology, the CREST Workshop on Speech Motor Control and Modeling in Sapporo, and the NATO Advanced Study Institute on Dynamics of Speech Production and Perception. The authors thank Louis Goldstein for the materials for Fig. 1. Finally, the authors thank Mary Beckman, Fred Cummins, an anonymous reviewer, and the Editor for extremely insightful and thoughtful reviews that significantly improved the presentation of this material.

References

Abry, C., & Lallouache, T. (1995). Modeling lip constriction anticipatory behavior for rounding in French with MEM (Movement Expansion Model). In K. Elenius, & P. Branderud (Eds.), *Proceedings of the XIIth international congress of phonetic sciences* (pp. 152–155). Stockholm: KTH and Stockholm University.

Bailly, G., Laboissière, R., & Schwartz, J. L. (1991). Formant trajectories as audible gestures: an alternative for speech synthesis. *Journal of Phonetics*, *19*, 9–23.

Barbosa, P., & Bailly, G. (1994). Characterisation of rhythmic patterns for text-to-speech synthesis. *Speech Communication*, *15*, 127–137.

Beckman, M. (1996). The parsing of prosody. *Language and Cognitive Processes, 11*(1/2), 17–67.

Beckman, M. E., & Edwards, J. (1992). Intonational categories and the articulatory control of duration. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 359–375). Tokyo: Ohmsha, Ltd.

Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. J. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II: gesture, segment, prosody* (pp. 68–86). Cambridge: Cambridge University Press.

Berkovits, R. (1993a). Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics*, *21*, 479–489.

Berkovits, R. (1993b). Progressive utterance-final lengthening in syllables with final fricatives. *Language and Speech*, *36*(1), 89–98.

Berkovits, R. (1994). Durational effects in final lengthening, gapping and contrastive stress. *Language and Speech*, *37*(3), 237–250.

Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: effects on language comprehension. *Language and Speech*, *38*, 365–391.

Browman, C. P. (1994). Lip aperture and consonant releases. In P. A. Keating (Ed.), *Phonological structure and phonetic form: papers in laboratory phonology III* (pp. 331–353). Cambridge: Cambridge University Press.

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

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

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston, & M. E. Beckman (Eds.), *Papers in Laboratory Phonology: I. Between the Grammar and the Physics of Speech* (pp. 341–376). Cambridge: Cambridge University Press.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, *49*, 155–180.

Browman, C.P., & Goldstein, L. (1995). Dynamics and articulatory phonology. In R. Port, & T. van Gelder (Eds.), *Mind as motion: dynamics, behavior, and cognition*. Cambridge, MA: MIT Press.

Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée*, *5*, 25–34.

Bullock, D., & Grossberg, S. (1988). Neural dynamics of planned arm movements: Emergent invariants and speed–accuracy properties during trajectory formation. *Psychological Review*, *95*, 49–90.

Byrd, D. (1996a). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*(2), 209–244.

Byrd, D. (1996b). A phase window framework for articulatory timing. *Phonology*, *13*(2), 139–169.

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, *57*(1), 3–16.

Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. In M. B. Broe, & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V* (pp. 70–87). Cambridge: Cambridge University Press.

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple phrasal boundaries. *Journal of Phonetics*, *26*, 173–199.

Cambier-Langeveld, T. (1997). The domain of final lengthening in production of Dutch. In Coerts, & de Hoop (Eds.), *Linguistics in the Netherlands*. (pp. 13–24). Amsterdam: John Benjamins.

Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: articulatory evidence from Georgian. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology VII* (pp. 419–448). Berlin: Walter de Gruyter.

Cho, T. Manifestation of prosodic structure in articulatory variation: evidence from lip movement kinematics in English. *Laboratory Phonology VIII* (submitted).

Cho, T. (2001). *Effects of prosody on articulation in English*. PhD dissertation. University of California, Los Angeles. (Distributed as UCLA Dissertations in Linguistics, number 22).

Cho, T., & Jun, S.-A. (2000). Domain initial strengthening as enhancement of laryngeal features: aerodynamic evidence from Korean. *Chicago Linguistics Society*, *36*, 31–44.

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29*(2), 155–190.

Coker, C. H. (1976). A model of articulatory dynamics and control. *Proceedings of the IEEE*, *64*, 452–460.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: a literature review. *Language and Speech*, *40*, 141–201.

Docherty, G. (1992). *The timing of voicing in British English obstruents*. Berlin: Foris.

Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, *89*(1), 369–382.

Ferriera, F. (1993). The creation of prosody during sentence production. *Psychological Review*, *150*, 233–253.

Flash, T., & Sejnowski, T. (2001). Computational approaches to motor control. *Current Opinion in Neurobiology*, *11*, 655–662.

Fletcher, J. (1991). Rhythm and final lengthening in French. *Journal of Phonetics*, *19*(2), 193–212.

Fougeron, C. (2001). Domain-initial articulatory variations in French. *Journal of Phonetics*, *29*(2), 109–135.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *101*, 3728–3740.

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

Fujimura, O. (2000). The C/D model and prosodic control of articulatory behavior. *Phonetica*, *57*, 128–138.

Gaitenby, J. H. (1965). The elastic word. *Haskins Laboratories Status Report on Speech Research*, SR-2.

Gibson, J. J. (1975). Events are perceivable but time is not. In J. T. Fraser, & N. Lawrence (Eds.), *The study of time II* (pp. 295–301). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gibson, J.J. (1986). *The ecological approach to visual perception.* Hillsdale, NJ: Lawrence Erlbaum (Originally published: Boston: Houghton-Mifflin, 1979).

Goldstein, L., & Fowler, C. Articulatory phonology: a phonology for public language use (submitted).

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

Harrington, J., Fletcher, J., & Roberts (1995). Coarticulation and the accented/unaccented distinction: evidence from jaw movement data. *Journal of Phonetics*, *23*, 305–322.

Hawkins, S. (1992). An introduction to task dynamics. In G. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology I: gesture, segment, prosody* (pp. 9–25). Cambridge: Cambridge University Press.

Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody*. PhD dissertation, Ohio State University.

Jun, S.-A. (1995). Asymmetrical prosodic effects on the laryngeal gesture in Korean. In B. Connell, & A. Arvaniti (Eds.), *Phonology and phonetic evidence: papers in laboratory phonology IV* (pp. 235–253). Cambridge: Cambridge University Press.

Keating, P. A., Cho, T., Fougeron, C., & Hsu, C.-S. (1999). Domain-initial strengthening in four languages. *UCLA Working Papers in Linguistics*, *97*, 139–156 (Also to appear in *Papers in laboratory phonology VI*. Cambridge: Cambridge University Press).

Klatt, D. H. (1976). Structure of a phonological rule component for a synthesis-by-rule program. *IEEE Transactions on Acoustics Speech, and Signal Processing*, *24*, 391–398.

Koiso, H., Shimojima, A., & Katagiri, Y. (1998). Collaborative signaling of information structures by dynamic speech rate. *Language and Speech*, *41*, 323–350.

Kröger, B. J., Schröder, G., & Opgen-Rhein, C. (1995). A gesture-based dynamic model describing articulatory movement data. *Journal of the Acoustical Society of America*, *98*, 1878–1889.

Laboissière, R., Schwartz, J.-L., & Bailly, G. (1991). Motor control for speech skills: a connectionist approach. In D. S. Tonretzky, J. L. Elman, T. J. Sejnowski, & G. E. Hinton (Eds.), *Connectionist models. Proceedings of the 1990 summer school* (pp. 319–327). San Mateo, CA: Morgan Kaufmann.

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

McClean, M. (1973). Forward coarticulation of velar movement at marked junctural boundaries. *Journal of Speech & Hearing Research*, *16*(2), 286–296.

Mussa-Ivaldi, F. A. (1995). Geometrical principles in motor control. In M. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 434–438). Cambridge: MIT Press.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dodrecht: Foris Publications.

O'Dell, M. L., & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the XIVth international congress of phonetic sciences*, *Vol. 2* (pp. 1075–1078). New York: American Institute of Physics.

Ostry, D. J., Gribble, P., & Gracco, V. L. (1996). Coarticulation of jaw movements in speech production: is context sensitivity in speech kinematics centrally planned? *The Journal of Neuroscience*, *16*(4), 1570–1579.

Perrier, P., Ostry, D. J., & Laboissière, R. (1996). The equilibrium point hypothesis and its application to speech motor control. *Journal of Speech and Hearing Research*, *39*, 365–377.

Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press [Linguistic Inquiry Monograph Series, 15].

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology I: gesture, segment, prosody* (pp. 90–117). Cambridge: Cambridge University Press.

Plaut, D. C. (1995). Lesioned attractor networks as models of neuropsychological deficits. In M. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 540–543). Cambridge, MA: MIT Press.

Price, P. J., Ostendorf, S., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, *90*(6), 2956–2970.

Port, R. F., & Cummins, F. (1992). The English voicing contrast as velocity perturbation. In *Proceedings of the 1992 International Conference on Spoken Language Processing (ICSLP '92)*, *Vol. 2* (pp. 1311–1314). Banff, Alberta, Canada. Edmonton: Priority Printing.

Repp, B. (1998). A microcosm of musical expression: I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America*, *104*, 1085–1100.

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

Saltzman, E. L. (1995). Dynamics and coordinate systems in skilled sensorimotor activity. In R. Port, & T. van Gelder (Eds.), *Mind as motion: Dynamic, behavior, and cognition* (pp. 149–173). Cambridge, MA: MIT Press.

Saltzman, E. (1999). Nonlinear dynamics of temporal patterning in speech. In P. L. Divenyi, & R. J. Porter (Eds.), *Proceedings of symposium on the dynamics of the production and perception of speech, a satellite symposium of the XIVth international congress of phonetic sciences*. Martinez, CA: East Bay Institute for Research and Education, Inc.

Saltzman, E., & Byrd, D. (1999). Dynamical simulations of a phase window model of relative timing. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the XIVth international congress of phonetic sciences* (pp. 2275–2278). New York: American Institute of Physics.

Saltzman, E., & Byrd, D. (2000a). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, *19*(4), 499–526.

Saltzman, E., & Byrd, D. (2000b). Demonstrating effects of parameter dynamics on gestural timing. May 2000 meeting of the Acoustical Society of America. *Journal of the Acoustical Society of America*, *107*(5,2), 2904.

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

Saltzman, E., Löfqvist, A., Kay, B., Kinsella-Shaw, J., & Rubin, P. (1998). Dynamics of intergestural timing: a perturbation study of lip–larynx coordination. *Experimental Brain Research*, *123*, 412–424.

Saltzman, E., Löfqvist, A., & Mitra, S. (2000). "Clocks" and "glue"—Global timing and intergestural cohesion. In M. B. Broe, & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V* (pp. 88–101). Cambridge: Cambridge University Press.

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

Sancier, M., & Fowler, C. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*, 421–436.

Sanderman, A. A., & Collier, R. (1997). Prosodic phrasing and comprehension. *Language and Speech*, *40*, 391–409.

Selkirk, E. O. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

Schafer, A. J. (1997). *Prosodic parsing*: *The role of prosody in sentence comprehension*. University of Massachusetts Doctoral Dissertation.

Schafer, A. J., & Speer, S. R. (1997). The influence of prosodic phrasal boundaries on the resolution of semantic lexical ambiguity. 134th meeting of the Acoustical Society of America, San Diego, CA. *Journal of the Acoustical Society of America*, *102*, 3138.

Schepman, A., & Rodway, P. (2000). Prosody and parsing in coordination structures. *The Quarterly Journal of Experimental Psychology*, *53*(2), 377–396.

Shadmehr, R. (1995). Equilibrium point hypothesis. In M. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 370–372). Cambridge, MA: MIT Press.

Shattuck-Hufnagel, S., & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 193–247.

Shattuck-Hufnagel, S., & Turk, A. (1998). The domain of final-lengthening in English. *Proceedings of 16th international congress on acoustics and 135th meeting Acoustical Society of America*, *Vol. 2*, pp. 1235–1236.

Steedman, M. (1991). Structure and intonation. *Language*, *67*, 262–296.

Steedman, M. (2000). *The syntactic process*. Cambridge, MA: MIT Press.

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *The Journal of the Acoustical Society of America*, *101*(1), 514–521.

van Lieshout, P. H., Starkweather, C. W., Hulstijn, W., & Peters, H. F. M. (1995). Effects of linguistic correlates of stuttering on *emg* activity in nonstuttering speakers. *Journal of Speech and Hearing Research*, *38*, 30–372.

van Santen, J. P. H. (1997). Segmental duration and speech timing. In Y. Sagisaka, N. Campbell, & N. Higuchi (Eds.), *Computing prosody* (pp. 225–249). New York: Springer.

Vatikiotis-Bateson, E., Hirayama, M., Honda, K., & Kawato, M. (1992). The articulatory dynamics of running speech: Gestures from phonemes. In *Proceedings of the 1992 International conference on spoken language processing (ICSLP '92)*, Vol. 2 (pp. 887–890). Banff, Alberta, Canada. Edmonton: Priority Printing.