Intragestural dynamics of multiple prosodic boundaries

Dani Byrd*

Haskins Laboratories, 270 Crown Street, New Haven, CT 06511, U.S.A.

Elliot Saltzman

Haskins Laboratories, 270 Crown Street, New Haven, CT 06511,
and Department of Physical Therapy, Boston University, U.S.A.

Recent work has demonstrated that temporal lengthening of articulatory gestures adjacent to an intonational phrase boundary can be interpreted as resulting from a lower underlying gestural stiffness. Magnetometer data from three subjects were examined to determine whether multiple levels of prosodic boundaries can be distinguished in the spatio-temporal patterning of articulation and whether these patterns are consistent with the lowered gestural stiffness account. Eighty tokens were recorded of a /CVCV#CVCV/ sequence with five differing boundary conditions intended to elicit multiple prosodic boundaries. Magnitude and duration of each consonantal closing and opening gesture, and the temporal latency between consonants were determined. Three levels of interconsonant latency in the /CV#CV/ portion of the utterance can be statistically differentiated. These differences result from lengthening of postboundary closing movements and, to a lesser extent, the pre-boundary opening movements. Concomitant spatial changes are discussed. The kinematic data are used to model the intragestural dynamics of boundary-adjacent lengthening. This modeling suggests that although lowered gestural stiffness is the main source of lengthening, another parameter—a rise-time for gestural activation—is also necessary. Speaker-specific differences in dynamics are accounted for by corresponding differences in the manner in which rise-time varies with stiffness. © 1998 Academic Press Limited

1. Introduction

The study of the relation between linguistic prosody and speech production has demonstrated that there are a variety of prosodic influences on articulation, including effects of syllable structure and phrasal, emphatic, and lexical stress. In addition to syllable structure and stress, phrasing is another aspect of prosody that has been the focus of recent attempts to investigate the effects of prosodic structure on speech motor control.

*Correspondence to Dani Byrd.

Phrasing refers to the organization of words into structural constituents related to syntactic, semantic, and pragmatic information. Phonologists have reason to believe that a variety of phrasal domains exist which are organized into cumulatively larger domains. This organization is referred to as the Prosodic Hierarchy (see Shattuck-Hufnagel & Turk, 1996, for an overview). The standard view is that an utterance is represented obligatorily at all levels of the hierarchy. Put more rigorously, linguists assume that (1) there is a small, fixed set of prosodic domains, (2) an utterance is exhaustively parsed into a sequence of such domains, and (3) "within each language the domains are strictly ordered along a scale such that (except for the largest domain) each domain is properly contained within a domain of the next higher level in the scale" (Inkelas, 1990), that is, the domains are *strictly layered* (Selkirk, 1984).[1] In a given language, for example, these prosodic domains might include the Prosodic Word, the Intermediate (or Phonological) Phrase, the Intonational Phrase, and, possibly, the Utterance. The perceived strength of a juncture at domain edges will depend on the level in the prosodic hierarchy of the constituents abutting to form the juncture—the higher the level, the stronger the boundary.

Phonetic evidence in support of a prosodic hierarchy has been well documented with regard to F_0 (e.g., Beckman & Pierrehumbert, 1986; Nespor & Vogel, 1986), glottalization (Dilley, Shattuck-Hufnagel & Ostendorf, 1996), VOT (Jun, 1995), and acoustic final lengthening (e.g., Oller, 1973; Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992; Ladd & Campbell, 1991). Consider, for example, phrase-final lengthening—i.e., longer acoustic or articulatory duration in the final position of certain prosodic domains—which has been extensively discussed in the speech production literature. Typically, the last vowel before a large phrasal boundary is lengthened, but other units such as final consonants, final VC's, final syllables, and final words have been identified as subject to domain-final lengthening as well. Acoustic data reported in Wightman *et al.* (1992) indicate that several distinct types of phrasal boundaries can be distinguished by their degree of acoustic final lengthening. This result encourages the view that prosodic boundaries of a variety of strengths are active in determining temporal aspects of speaking.

As compared to intonational and acoustic studies, little research has addressed the effects of prosodic structure on the temporal characteristics of supraglottal articulatory patterns. Edwards, Beckman & Fletcher (1991), Beckman, Edwards & Fletcher (1992), and Beckman & Edwards (1992) report data on jaw movement that can be modeled in a dynamical systems framework as demonstrating a decreased stiffness parameter within a mass–spring gestural model. This parameter modulation is argued to yield the increased duration and decreased peak velocity of the final oral closing gesture of a VC sequence in phrase-final position. These studies of jaw movement motivate the consideration of how phrasal structure affects the movement of primary articulators, e.g., lips or tongue, in consonant articulation. Phrase-initial lengthening has also been reported (e.g., Byrd, Kaun, Narayanan & Saltzman, in press; Wilbur & Zelaznik, 1997, for ASL).

A number of studies have found articulations of a larger magnitude in word-initial as compared to non-initial positions (Vaissière, 1988; Krakow, 1989; Goldstein, 1992, citing Cooper, 1991; Browman & Goldstein, 1995; Keating, 1995; Byrd, 1996). Some studies have reported spatially larger articulations in phrase-initial positions. The acoustic study of Pierrehumbert & Talkin (1992) finds a larger glottal opening for [h] when it is initial after a phrase boundary than when it is in the middle of the phrase. Wilbur & Zelaznik

---

[1] But see Inkelas (1990), Selkirk (in press), and Ladd (1986), all discussed in Shattuck-Hufnagel & Turk (1996), for possible exceptions to strict layering.

(1997) report longer movement durations for ASL signs in phrase-final as compared to phrase-medial position. One of the few lines of study that has demonstrated that multiple phrasal domains can be differentiated in supralaryngeal articulation is that of Fougeron and Keating. [See also Dilley, Shattuck-Hufnagel & Ostendorf (1996) for similar evidence in looking at word-initial glottalization, and de Jong (1995) for similar results in his consideration of syllable stress.] Fougeron & Keating (1996, 1997) find an increase in the magnitude of lingual gestures (as measured by linguapalatal contact) for consonants initial in increasingly large domains, and (1997) larger articulations in domain-initial positions at each level than in medial and final positions at that level. In light of the "the-farther-the-longer" behavior typical of both speech and limb movements, whereby durations tend to be longer for movements with larger displacements (e.g., Kelso, Vatikiotis-Bateson, Saltzman & Kay, 1985; Ostry & Munhall, 1985; Saltzman, Löfqvist & Mitra, to appear), it seems reasonable to hypothesize that these phrase-initial gestures are also longer even though temporal effects were not the focus of these studies. These studies have encouraged us to look for evidence of multiple prosodic domains in the temporal characteristics of articulatory gestures.

These data and our own have led us to conceive of boundary-adjacent lengthening as a local slowing of the gestures in the immediate vicinity of sufficiently strong prosodic boundaries or, equivalently, at the juncture between prosodic domains of a sufficiently high position in the prosodic hierarchy. Ultimately, we wish to describe the control of such prosodically conditioned durational patterns in a dynamical systems framework.

The most general meaning of the term "dynamical systems framework" refers to a mathematical model that specifies how a system evolves in time. More specifically, the term can be used to refer to the mathematical relation between the current kinematic state (e.g., position and velocity) and parameter values of a system and the next state of the system. One particular dynamical systems approach called task dynamics (Saltzman, 1986; Saltzman & Munhall, 1989) has been used over the last decade to model speech behavior by many speech scientists and linguists. Saltzman & Munhall (1989) outline the advantages of such a theoretical approach—namely, that it offers a potentially unified account of speech kinematic patterns, their stability to perturbations, and their lawful warping due, for example, to speech rate and prosody. Various researchers have modeled effects of prosody on articulation using concepts laid out in the task dynamic model; see, for example, Edwards *et al.* (1991), de Jong (1991), Beckman *et al.* (1992), Beckman & Edwards (1992), Harrington, Fletcher & Roberts (1995), and Byrd *et al.* (in press). The modeling of the empirical results presented below which describes the control of boundary-conditioned lengthening in a dynamical systems framework continues in this vein by elaborating on these previous modeling efforts.

The articulatory study described here considers intragestural articulatory characteristics in the immediate, or local, neighborhood of a variety of phrasal boundaries. [For a discussion of phrasal effects on *inter*gestural organization see Byrd (in press) and Byrd *et al.* (in press).] The following general questions are addressed: What are the temporal signatures of prosodic domain edges in articulation? And, how can these signatures be understood in terms of the dynamics of gestural control? The hypotheses driving the work presented in the remainder of this paper are that (a) multiple (more than two) levels of prosodic boundaries can be distinguished by their temporal articulatory characteristics, specifically by the boundary-adjacent lengthening of gestures; and (b), as has been suggested in previous work, this lengthening can be modeled, at least in part, by lowered gestural stiffness. The results presented below bear on these hypotheses, and

bring to light new issues regarding the adequacy of current gestural models and the existence of speaker-specific dynamical strategies for prosodically modulating gestural durations.

2. Method

2.1. *Stimuli*

We investigate these experimental hypotheses by examining the articulatory behavior of lip closing and opening gestures for [m] adjacent to differing prosodic boundaries. These boundaries were elicited using five sentences in which the target sequence [mamə(#)mi(mi)] was embedded. Using meaningful sentences, this string was placed in different syntactic structures which were intended to elicit prosodic boundaries of varying strengths between the second and third syllables. These sentences are shown in Table I. The first sentence is a no-boundary or word-internal control condition. The second has the target syllable separated by a word boundary; the third has a boundary between enumerated items in a list; the fourth has a vocative boundary; and the last an utterance boundary.

   We will be discussing only the movements in the immediate vicinity of the boundary. That is, in the [mamə(#)mi(mi)] sequence, we will be evaluating the preboundary opening from the second [m] and the postboundary closing for the third [m]. Note that in all but one of the conditions the target syllables are the fifth and sixth or seventh of the sentence; in the vocative condition they are the third and fourth. The speakers were instructed to read the sentence in a casual style with no substantial pauses.

   The last four constructions were designed to elicit a selection of prosodic junctures. A great deal of linguistic research on the prosodic domains active in the phonology of English suggests that the word (W) boundary, the intermediate (or phonological) phrase boundary (PP), the intonational phrase (IP) boundary, and the utterance (U) boundary are boundaries that speakers employ; although theoreticians disagree about the exact number and characteristics of these active prosodic domains (Shattuck-Hufnagel & Turk, 1996). Work in prosodic phonology and phonetics (e.g., Fougeron & Keating, 1996, 1997; see also Shattuck-Hufnagel & Turk, 1996; Ross & Ostendorf, 1996) has illuminated certain difficulties for experimental design and interpretation. Among these are that (a) subjects can differ in both the types and number of prosodic phrasings they choose to use for a particular set of written stimuli, (b) the U and IP levels may not be distinct (Shattuck-Hufnagel & Turk, 1996), and (c) utterance-medial IPs and

TABLE I. Stimuli sentences for five experimental boundary conditions (Boldface was not present in the stimuli seen by subjects)

Boundary condition	Sentence
none (word-medial)	Poppa begged "mom**mamí**a" meanly upon coming.
word	Poppa-Pikt and Mom**ma-Mi**mi tapped Coby.
list	Poppa, Pikt, Mom**ma, Mi**mi, and Bibi tapped Coby.
vocative	Quick Mom**ma, Mi**mi tapped Coby.
utterance	Poppa picked Mom**ma. Mi**mi tapped Coby.

PPs may be distinguishable only by a "subjective sense of disjuncture [as indicated, for example, by lengthening]" and not by intonation (Beckman & Ayers, 1994). (It is, in fact, exactly this temporal behavior that we evaluate below as the dependent variables in our study.)

Our experimental goal was to elicit from each speaker *multiple* prosodic domains, rather than a particular domain for a given experimental condition (although we screen the data to ensure consistency within subject). We did not hope for or expect different speakers to necessarily use the same domains for the same experimental condition. Accordingly, statistical analyses (see below) are performed within-subject. For the sake of consistency across subjects in reporting the results we refer to the experimental conditions atheoretically—i.e., not by the name of any targeted theoretical domain but simply as the following experimental conditons: NONE, WORD, LIST, VOCATIVE, and UTTERANCE. In addition to data on the temporal behaviour at the juncture—the focus of the paper—information is also provided below on the acoustic and intonational patterning displayed for each condition for each subject, thereby allowing the reader to consider the data in his or her own theoretical terms, e.g., as demonstrating groupings of minor *vs.* major boundary or of phonological or intermediate *vs.* intonational boundary. Our further focus in Section 4 will be to demonstrate *how* boundary-adjacent lengthening at any phonological level may occur or, at least, be modeled in the speech production system.

2.2. *Subjects and procedure*

The EMMA magnetometer system (Perkell *et al.*, 1992) was used to track the movement of the upper and lower lips for 16 tokens of each of the five sentences produced by three speakers. Two of these speakers, a male (Speaker J) and a female (Speaker K), were young adult college undergraduates; the third, a female (Speaker F), was a middle-aged university professor in speech science. For all subjects the conditions were blocked and recorded in the following order, repeated twice: eight repetitions of UTTERANCE, eight repetitions of VOCATIVE, eight repetitions of LIST, eight repetitions of WORD, eight repetitions of NONE. For all subjects this experiment was part of a longer magnetometer recording session, and was the first experiment recorded. This procedure yielded a total of 80 sentences for each speaker.

2.3. *Data collection*

The EMMA magnetometer system was used to transduce the horizontal (x) and vertical (y) movements of small coils attached to the articulators in the midsagittal plane. The technical specifications of the EMMA magnetometer system are outlined in Perkell et al. (1992) (see also Gracco & Nye, 1993; Löfqvist, 1993). Single transducers were placed on the nose, upper and lower teeth (maxilla and jaw, respectively), upper and lower lips (at the vermilion border), and tongue tip. Three transducers for Speakers J and K, and one for Speaker F, were placed on the tongue body. The EMMA data were sampled at 625 Hz after low-pass filtering at 200 Hz before voltage-to-distance conversion. After voltage-to-distance conversion (with a filter cutoff of 17 Hz), correction for head movement (using the nose and maxillary reference transducers) and rotation to the occlusal plane (determined using a bite plate on which two receivers are fixed), the position signals were smoothed by a 25-point triangular filter.
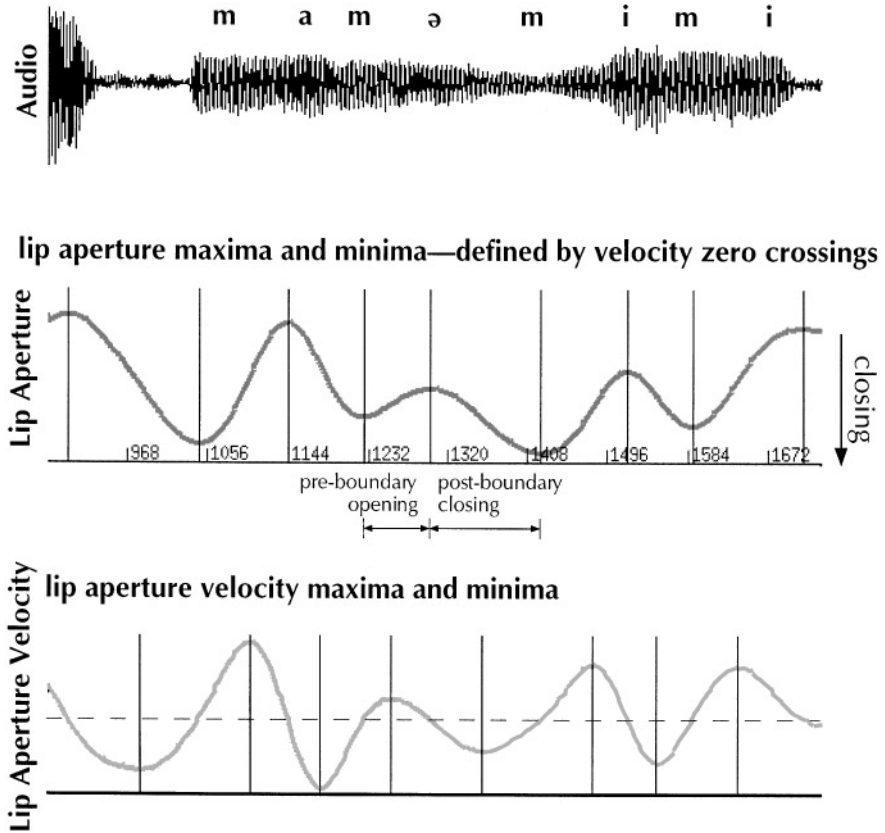
Figure 1. Token of the [mamə(#)mi(mi)] sequence. Top panel: audio waveform. Middle panel: Lip aperture. (Vertical strokes indicate onset and end of each lip closing or opening.) Bottom panel: Lip Aperture Velocity. (Vertical strokes indicate velocity peaks.)

2.4. *Data analysis*

For this study, horizontal (x) and vertical (y) position signals for two transducers are examined: the upper lip and the lower lip. We call the Euclidean distance between these receivers Lip Aperture.[2] Lip Aperture was calculated according to the following formula:

$$\text{Lip Aperture} = \sqrt{((\text{lip aperture } x)^2 + (\text{lip aperture } y)^2)}$$

where

$$\text{lip aperture } x = \text{upper lip } x - \text{lower lip } x$$

$$\text{lip aperture } y = \text{upper lip } y - \text{lower lip } y$$

(Note that because the receivers are placed at the vermilion border of the lips, no particular value for Lip Aperture corresponds to the achievement of labial closure; zero Lip Aperture will not in fact occur as this would mean that the receivers were effectively on

---

[2] Since the environment included no rounded segments, the parameterization for the Lip Aperture signal includes both the x and y dimensions so that any component of the closure movement not perpendicular to the occlusal plane will not be lost in the signal calculation (see, e.g., Löfqvist & Gracco, 1997, Figure 4).

top of one another, and the Lip Aperture signal would change during acoustic closure due to tissue compression (Löfqvist & Gracco, 1997). The Lip Aperture signal was differentiated and smoothed by a 25-point filter to create a Lip Aperture Velocity signal.

The *HADES* signal analysis program (Rubin, 1995) was used to identify algorithmically in the lip aperture signal the onset (and ending) times of the lip closing and opening movement for each of the [m]s in the [mamə(#)mi(mi)] sequence. These events were defined by identifying the zero-crossings in the lip aperture velocity signals. The time and lip aperture value at each of these points were recorded. Additionally, for each opening and closing movement, time and magnitude of the peak velocities were collected. These labeled time points are shown in Fig. 1.

Tokens for a movement were excluded if multiple velocity peaks occurred for either of the boundary-adjacent movements. Such exclusions were rare; only one token was excluded for Speaker K and one for Speaker F. Full pauses at the boundary were rare to non-existent in the dataset, as indicated by the fact that (near zero-velocity plateaus occurred in only one token (the token excluded for Speaker F by the criteria just outlined).

Based on movement onset, peak, end and peak velocity, a variety of dependent variables were calculated. These variables (see Fig. 2) include:

(a) the preboundary opening and postboundary closing movement duration;
(b) the transboundary interval (i.e., preboundary opening + postboundary closing duration);
(c) the preboundary opening and closing displacement (i.e., the difference between starting and final lip aperture);
(d) the interval from the onset of the closing movement to its peak velocity.

These measurements reflect the duration and magnitude of the individual opening and closing articulations and help illuminate certain aspects of the intragestural dynamics. In general, the statistical analyses test the effects of boundary condition (5 levels) on the dependent measures using separate analyses of variance for each subject. When a significant ($p \leqslant 0.05$) main effect of boundary condition occurs, a Fisher's protected
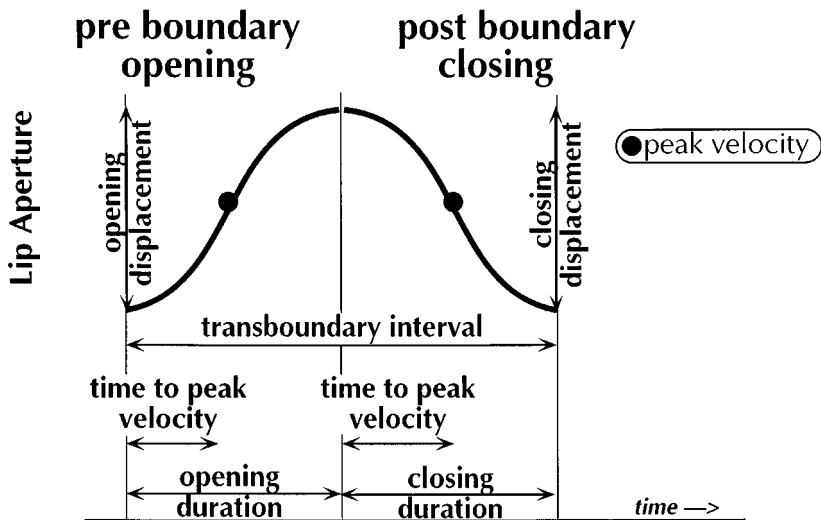

Figure 2. Schema indicating the measured experimental variables.

post hoc test is used to test for significant differences between each pair of boundary conditions. With a Bonferroni correction for 10 pairwise comparisons the conservative level of significance is $p < 0.005$. Differences reported below meet this conservative criterion for confidence level. Differences observed at the less conservative (i.e. non-Bonferroni corrected) level of 0.05 are also noted below, with the additional notation of ($p < 0.05$), as weaker evidence of a difference.

In the following section, we report the effects of the boundary condition on the kinematic characteristics of the preboundary opening and postboundary closing gestures. In Section 4, we describe how the lengthening of the postboundary closing gesture can be modeled with an extended task dynamics model (Saltzman & Munhall, 1989). We concentrate on modeling the closing gesture because it is most purely "consonantal", that is, lip aperture is directly controlled for bilabial closing but is only an indirect measure of tongue and jaw control for vocalic opening (and/or any active lip aperture release gesture).

3. Results

3.1. *Overview of acoustic and intonational behavior*

We view timing and intonation as often co-occurring but potentially independent cues to juncture. Indeed, evaluation of prosodic boundary strength or category often relies on temporal cues. The articulatory source of these cues is the focus of the present experiment. However, before presenting the articulatory results, we will provide an overview of the intonational and acoustic behavior of the data, so that the reader has additional information with which to evaluate the prosodic structure used in producing the experimental stimuli.

One of the authors (DB) listened to multiple repetitions of each sentence and determined that speakers were consistent in the intonational contours they used. Fig. 3 presents schematic intonation contours with the target phrase shaded. These contours were determined by the authors in consultation with two independent phonetically trained listeners. Certain differences between the speakers are apparent, and we might expect some parallel differences in articulatory behavior. One notable difference between speakers is that at the onset of the second phrase in the VOCATIVE condition Speaker F has only a small rise where Speakers K and J have a High commensurate with that initiating the first phrase of the sentence. In the WORD conditions, speakers differ in how early in the target phrase the High occurs. The NONE condition also differed in the vicinity of the target phrase among the three speakers.

In addition to intonational patterns, one other source of information about juncture that has been well researched is the acoustic lengthening of segments at a boundary, generally at a domain's right edge. This has been called final lengthening (for a review see Vaissière, 1988). This temporal patterning of acoustic segments results, of course, from the temporal patterning of the articulation. Because the acoustics are a complex result of temporally overlapping articulatory movements, there is no one-to-one mapping between acoustic segment duration and articulatory gesture duration. For example, in a V # C sequence, lengthening of the phrase-initial articulatory closure is likely to result in the acoustic lengthening of the phrase-final vowel since, all else equal, closure will be achieved later with respect to the vowel. That is, phrase-initial lengthening in articulation could well result in phrase-final lengthening in the acoustics. This paper presents
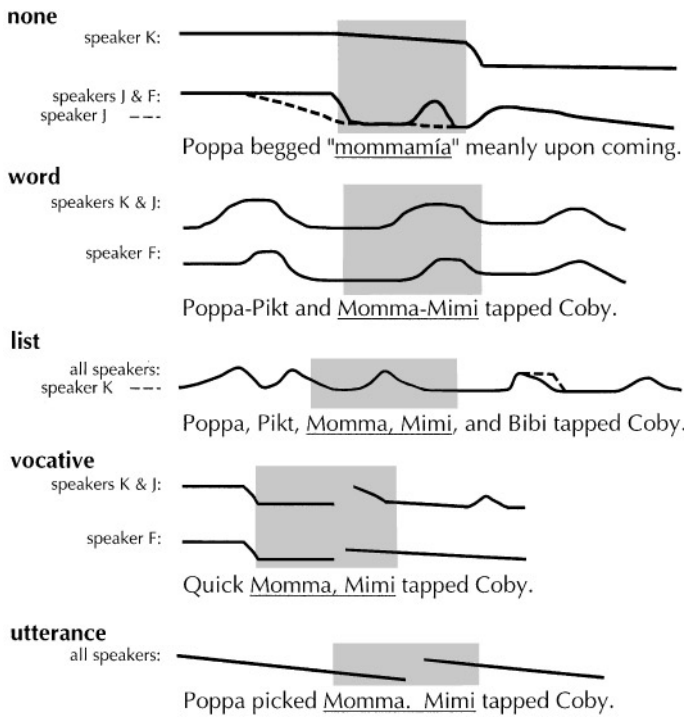
Figure 3. Schematic intonational contours (target phrase shaded) used by the three subjects. (Listening determined that speakers were consistent in the intonational contour they used for multiple repetitions of each sentence.)

information on one part of the articulatory patterning—namely, the lengthening of lip movements in the vicinity of boundaries. Of course, other articulators contribute to the decisions a researcher might make in acoustically segmenting this target phrase, e.g. velum, tongue, and glottal behavior. For this reason, we present a brief overview of the acoustic segment durations in our data.

Acoustic segmentation of vowels and nasals in the [mamə(#)mi(mi)] sequence was done using time-aligned waveforms and spectrograms and was based on the presence/absence of higher resonant structure in the waveform, amplitude discontinuity, and the presence/absence of nasal resonance in the spectrogram. Additionally, between the phrase-final and -initial vowels a voicing break, i.e. an interval showing no periodic glottal pulses, sometimes occurred. The duration of any such interval was also recorded. Voicing breaks generally occurred at the beginning of the initial [m] of the [mi(mi)] sequence (i.e., immediately following the [ə]), but sometimes voicing abated briefly somewhat later in the nasal. For each subject and each condition, Table II presents the duration of [ə], the number of occurrences and duration of voicing breaks, and the duration of the initial [m] of the [mi], including any voicing break.

Several points are noteworthy. First, the frequency (and duration) of voicing breaks appears, as we will see below, to agree well with the groupings each subject displays for the duration of the articulatory gestures at the boundary; i.e., conditions displaying longer articulatory durations are also likely to have more and longer voicing breaks. Second, an ANOVA testing for the effect of boundary condition on the duration of

TABLE II. Mean durations for three acoustic segment measurements for three speakers and five boundary conditions for the [mamə(#)mi(mi)] sequence

	ə# duration (ms)	voicing break	#m duration (ms) (including voicing break)
	Mean	Mean dur, n	Mean
Speaker J			
none	73	—	99
word	73	—	90
list	121	85 $n = 7$	154
vocative	111	190 $n = 16$	238
utterance	101	172 $n = 11$	199
Speaker K			
none	82	—	77
word	70	—	72
list	90	21 $n = 3$	100
vocative	87	69 $n = 3$	126
utterance	111	100 $n = 15$	173
Speaker F			
none	57	—	100
word	58	—	100
list	82	—	97
vocative	68	—	92
utterance	79	—	101

the [ə] finds a significant effect for all subjects [Sp. J $F(4,75) = 42.831$, $p < 0.0001$; Sp. K $F(4,75) = 23.947$, $p < 0.0001$; Sp. F $F(4,74) = 46.255$, $p < 0.0001$]. For speaker J, all categories differ ($p < 0.005$) except UTTERANCE and VOCATIVE, LIST and VOCATIVE, and WORD and NONE. For Speaker K, the UTTERANCE condition has a significantly longer vowel than all others and the WORD condition is significantly shorter than VOCATIVE and LIST ($p < 0.005$). For Speaker F, all conditions differ ($p < 0.005$) except UTTERANCE and LIST and WORD and NONE. In general, the change in vowel length (excluding the duration of any voicing break present following the vowel) from the shortest to longest condition was about $+45$ ms for the two slower speakers (J and K) and about $+25$ ms for the faster speaker (F). Finally, since we expect there to be some systematic relation between the acoustic final lengthening and the transboundary articulatory interval described above (i.e., preboundary opening + postboundary closing duration), we consider the correlation of these two measures. The correlation of the ə + voicing break duration with the duration of the transboundary articulatory interval is significant and positive for all subjects ($p < 0.0001$) with r ranging from 0.82 to 0.91.

3.2. *Transboundary articulatory interval*

The transboundary articulatory interval is from minimum lip aperture for the preboundary [m] to minimum lip aperture for the postboundary [m]. This interval is systematically related to, although not the same as, the interval used to measure acoustic final lengthening in the literature on juncture or boundary effects. This is because the transboundary articulatory interval encompasses the period from oral release for the

preboundary vowel through the moment at which the lips (or lip receivers in this case) are most closely approximated for the following labial consonant.

There is a significant main effect of boundary condition on the transboundary articulatory interval [Sp. J $F(4,75) = 73.899$, $p < 0.0001$; Sp. K $F(4,74) = 52.2$, $p < 0.0001$; Sp. F $F(4,74) = 43.521$, $p < 0.0001$]. Post hoc tests indicate that for Speaker J, all pairs except NONE *vs.* WORD and VOCATIVE *vs.* UTTERANCE differed ($p < 0.005$). For Speaker K, all pairs except NONE *vs.* WORD and LIST *vs.* VOCATIVE differed ($p < 0.005$);and for Speaker F, all pairs except NONE *vs.* WORD, VOCATIVE *vs.* WORD, VOCATIVE *vs.* NONE, and LIST *vs.* UTTERANCE differed ($p < 0.005$). These groupings are indicated in Fig. 4 in the following way. Conditions which did not differ significantly in the post hoc test are displayed using bars with the same fill pattern. Bars with different patterns indicate conditions that are significantly different.

This articulatory interval distinguishes three levels of phrasing for two speakers (J and K), and two levels for the third (F). Note that Speaker F's behavior in the first three
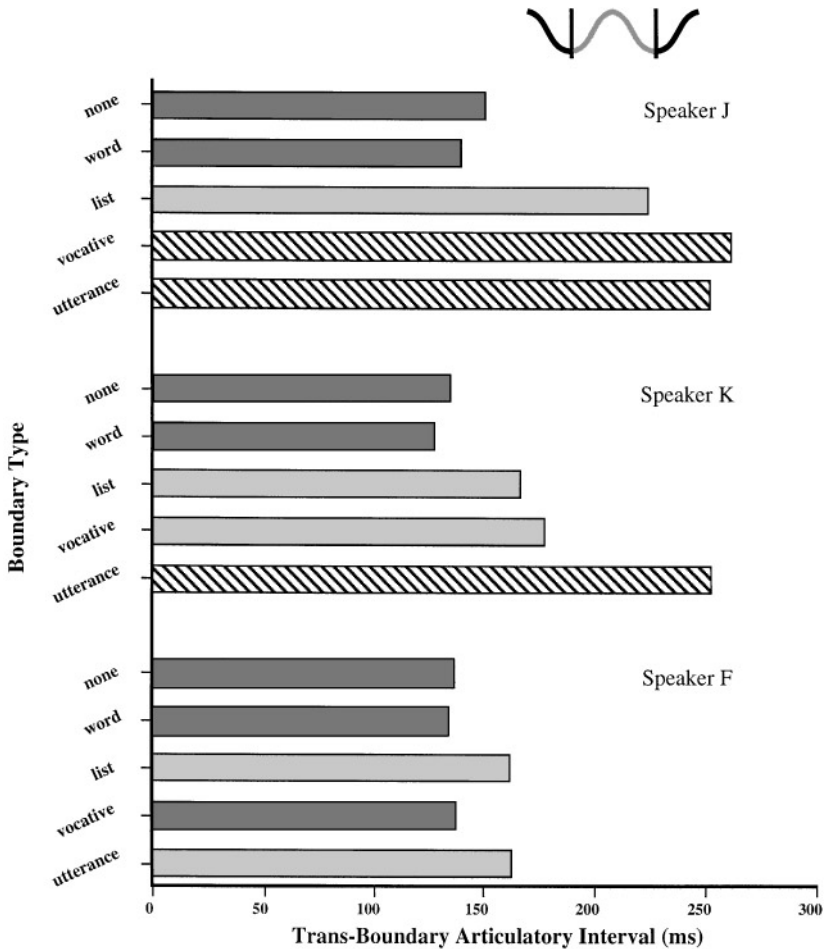

Figure 4. The transboundary articulatory interval for three speakers and five experimental boundary conditions. Bars with like shadings group together in post-hoc tests ($p < 0.005$).

conditions in Fig. 4 parallels that of the other speakers; however, this speaker did two things differently. First, recall that this speaker used an intonational pattern for the VOCATIVE condition that involved only a small rise at the beginning of the phrase. We see that this condition does not group with her UTTERANCE and LIST conditions for which she used substantial rises. Second, she does not display the large lengthening that the other speakers had in the UTTERANCE condition. This is a point to which we will return below in the discussion of variability. Note here, however, that it is likely that this speaker would have implemented this boundary with a pause had she been allowed, but was forced to implement it with a more intermediate level boundary in light of the instructions not to pause in the reading. (She informed the experimenters that she read that sentence "as though there was a semi-colon there.")

3.3. *Preboundary and postboundary durations*

The two component durations of the transboundary interval—preboundary opening and postboundary closing—are considered next. Means and standard deviations for preboundary opening and postboundary closing durations are given in Table III.

The postboundary closing gesture displays the same lengthening patterns (i.e., the same groupings in the post hoc comparisons) as that described above for the transboundary interval [main effect: Sp. J $F(4,75) = 55.352$, $p < 0.0001$; Sp. K $F(4,74) = 56.069$, $p < 0.0001$; Sp. F $F(4,74) = 22.983$, $p < 0.0001$]. The patterns for the preboundary opening are similar, showing a significant main effect of boundary condition for all subjects [Sp. J $F(4,75) = 53.886$, $p < 0.0001$; Sp. K $F(4,74) = 23.255$, $p < 0.0001$; Sp.

TABLE III. Means and standard deviations for duration (ms) and displacement (cm) for preboundary opening and postboundary closing movements for three speakers and five boundary conditions

	Preboundary opening		Postboundary closing	
	Duration (ms)	Displacement (cm)	Duration (ms)	Displacement (cm)
Speaker J				
none	70.9 (7.22)	0.394 (0.15)	79.4 (8.74)	0.541 (0.17)
word	62.8 (16.44)	0.275 (0.13)	76.9 (7.74)	0.413 (0.11)
list	105.0 (10.94)	0.850 (0.15)	118.6 (9.03)	1.051 (0.13)
vocative	109.9 (14.18)	0.753 (0.25)	151.3 (25.32)	1.021 (0.21)
utterance	110.3 (12.02)	0.882 (0.2)	140.9 (29.02)	1.108 (0.22)
Speaker K				
none	58.2 (6.68)	0.195 (0.06)	76.3 (5.25)	0.296 (0.05)
word	52.9 (6.49)	0.128 (0.04)	74.7 (5.08)	0.229 (0.05)
list	65.9 (13.74)	0.251 (0.12)	100.4 (22.45)	0.416 (0.15)
vocative	62.6 (7.72)	0.239 (0.09)	110.7 (24.55)	0.483 (0.13)
utterance	82.8 (10.26)	0.383 (0.09)	169.8 (31.40)	0.635 (0.11)
Speaker F				
none	59.9 (4.60)	0.243 (0.06)	76.2 (2.97)	0.368 (0.05)
word	56.2 (12.00)	0.204 (0.07)	77.8 (6.42)	0.363 (0.04)
list	73.6 (3.46)	0.342 (0.08)	88.1 (4.27)	0.513 (0.09)
vocative	61.7 (5.60)	0.262 (0.06)	75.6 (3.94)	0.384 (0.06)
utterance	70.4 (5.24)	0.366 (0.10)	92.1 (10.79)	0.597 (0.09)

F $F(4,74) = 17.964$, $p < 0.0001$]. For Speaker J, NONE and WORD group together and are differentiated from the three other conditions ($p < 0.005$). Among the latter conditions, LIST and VOCATIVE, UTTERANCE and VOCATIVE, and LIST and UTTERANCE are not statistically differentiated. That is, the pattern for Speaker J's preboundary opening differs from that of his postboundary closing in that LIST fails to be differentiated from UTTERANCE and VOCATIVE. Speaker K's pattern mirrors that of her postboundary closing movement—i.e., NONE and WORD group together, LIST and VOCATIVE group together, and UTTERANCE is differentiated from all other conditions ($p < 0.005$)—however, the comparison between VOCATIVE and NONE does not reach significance for the preboundary opening. Finally, the comparisons between VOCATIVE and WORD and between LIST and NONE are significant at the weaker ($p < 0.05$) level. Similarily, speaker F's preboundary opening lengthening pattern also parallels her postboundary closing pattern—i.e., NONE and WORD group together, LIST and UTTERANCE group together, and VOCATIVE is differentiated from NONE ($p < 0.005$). The comparison between VOCATIVE and WORD does reach the weaker ($p < 0.05$) significance level in the case of the preboundary opening, but does not for postboundary closing duration. The preboundary opening and postboundary closing movements are shown aligned at the onset of the postboundary closing (or, equivalently, the end of the preboundary opening) in Fig. 5. The local lengthening caused by the prosodic boundaries above the word level appears to operate roughly symmetrically on the boundary-adjacent gestures.

3.4. *Variability*

One additional factor of interest was the durational variability of these movements and any relation it might have to the strength of the adjacent prosodic boundary. To evaluate this question, consider the standard deviations of the postboundary closing movement durations shown in Fig. 6 (see also Table III).

Generally, greater durational variability is observed for stronger prosodic boundaries. For Speakers J and K the standard deviations become larger for each of the groupings of increasing mean duration described in Section 2.2. For Speaker F's WORD condition, the standard deviation is slightly larger than those of the NONE and VOCATIVE conditions with which it patterns in mean duration. Finally, while Speaker F's mean closing duration values do not uniquely differentiate the UTTERANCE boundary, variability in those closing durations is substantially greater for this condition than for the other conditions.

Interestingly, in tasks involving the perception and production of isochronous intervals, a generalized version of Weber's Law (see e.g., Engen, 1971; Ivry & Hazeltine, 1995) has been found to hold such that the variability of temporal intervals typically increases as interval duration increases (Repp, 1997). Repp further finds that variability behaves similarly in IOIs (inter-onset intervals) in musical performance for rhythmical sequences of different note durations. He concludes that "timing control is least accurate at structural boundaries associated with long note values and/or expressive lengthening" (Repp, 1997). Our data on phrase-initial lengthening appear also to demonstrate this type of general behavior.

3.5. *Displacement*

The displacements for the preboundary opening and postboundary closing movements are shown in Fig. 7. As in Fig. 4, bars with different patterns indicate conditions that are significantly different for that subject.
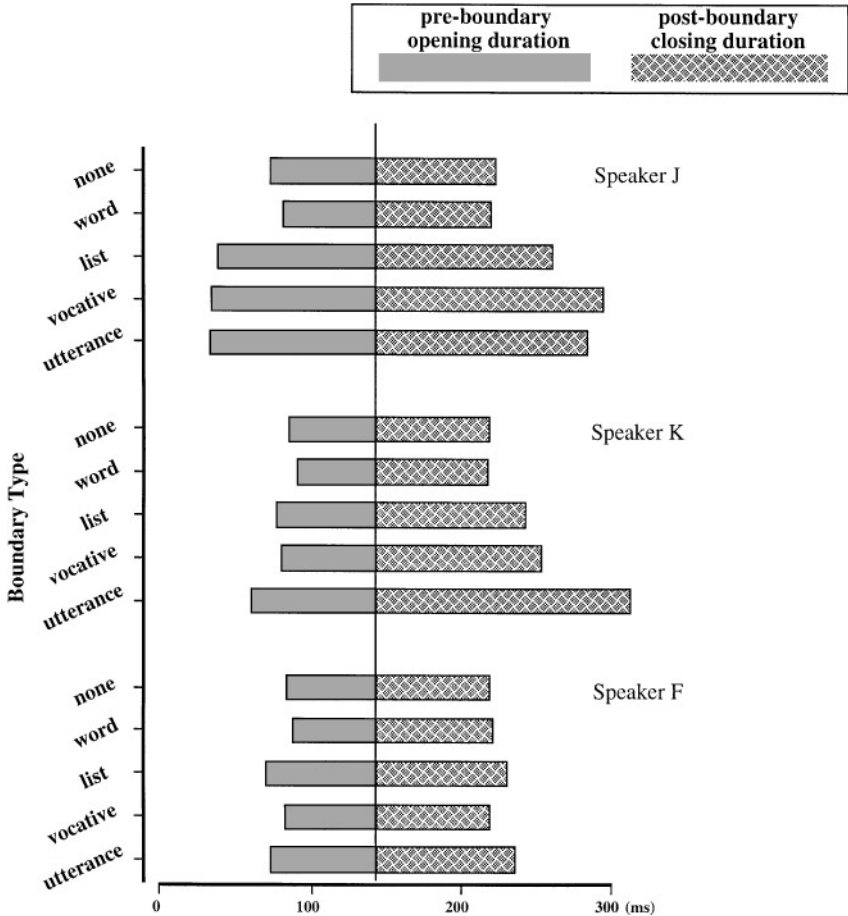
Figure 5. Preboundary opening and postboundary closing durations for three
speakers and five experimental boundary conditions. Vertical line marks onset of
post-boundary closing (and end of preboundary opening).

There was a significant main effect of boundary on both the opening and closing
displacements [opening: Sp. J $F(4,75) = 37.16$, $p < 0.0001$; Sp. K $F(4,74) = 19.969$,
$p < 0.0001$; Sp. F $F(4,74) = 12.926$, $p < 0.0001$; closing: Sp. J $F(4,75) = 56.128$,
$p < 0.0001$; Sp. K $F(4,74) = 36.298$, $p < 0.0001$; Sp. F $F(4,74) = 36.087$, $p < 0.0001$]. For
the most part, the displacement values behaved according to the farther-the-longer
pattern; i.e., those conditions that had longer durations had larger displacements. The
preboundary opening displacements pattern in pairwise comparisons of boundary con-
ditions ($p < 0.005$) just as the preboundary opening durations for all pairwise compari-
sons except one; for Speaker K WORD is differentiated from VOCATIVE in displacement, but
for duration the significance level of the difference ($p = 0.0053$) does not quite reach the
Bonferroni-corrected criterial level. The postboundary closing displacements pattern just
as the postboundary closing durations in all pairwise comparisons of boundary condi-
tions ($p < 0.005$) except two; for Speaker F LIST is differentiated from UTTERANCE in
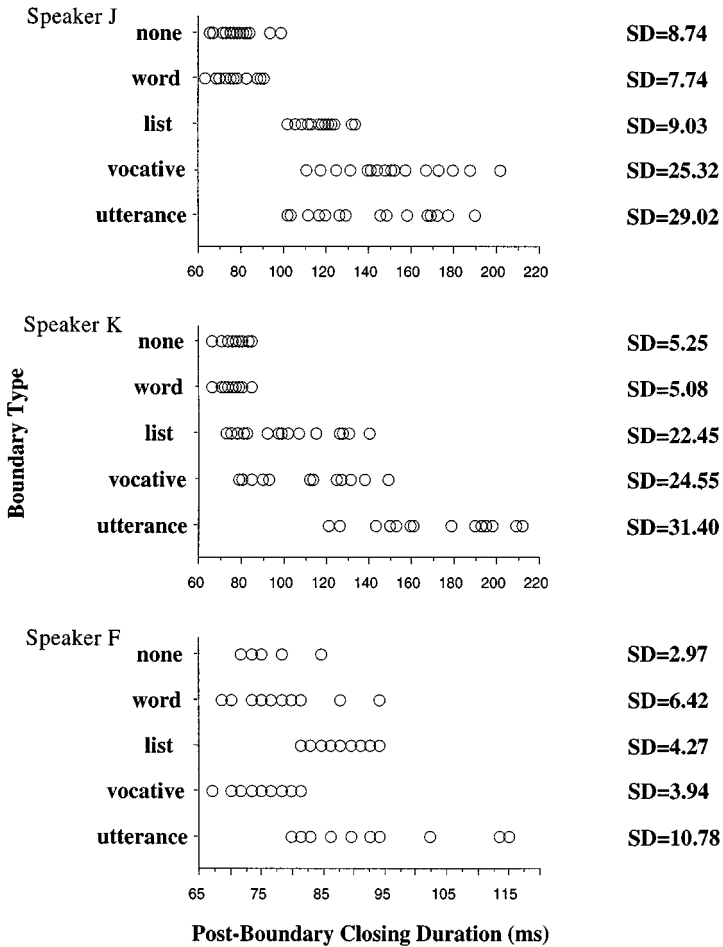displacement ($p < 0.005$) but not in duration; and for Speaker J LIST is not differentiated

Figure 6. Token distribution and standard deviations for postboundary closing durations for three speakers and five experimental boundary conditions.

from VOCATIVE in displacement but is in duration. To convey an overall idea of the consistency in patterning between duration and displacement, of 60 total pairwise comparisons (10 comparisons × 3 subjects × 2 gestures), all but three yield the same result for displacement as for duration. The correlation between displacement and duration is high and positive for all subjects, ranging from $r = 0.83$ to $r = 0.91$ ($p < 0.0001$) for preboundary opening and from $r = 0.77$ to $r = 0.89$ ($p < 0.0001$) for postboundary opening.

3.6. *Time-to-peak-velocity*

Because one of our goals is to understand the control of linguistically conditioned durational patterns in a dynamical systems framework, we wish to evaluate the claim (Edwards *et al.*, 1991; Beckman *et al.*, 1992; Beckman & Edwards 1992; Byrd *et al.*, in press) that a lowered gestural stiffness within a mass–spring gestural model (e.g., Saltzman & Munhall, 1989) is the source of lengthening at phrase boundaries.
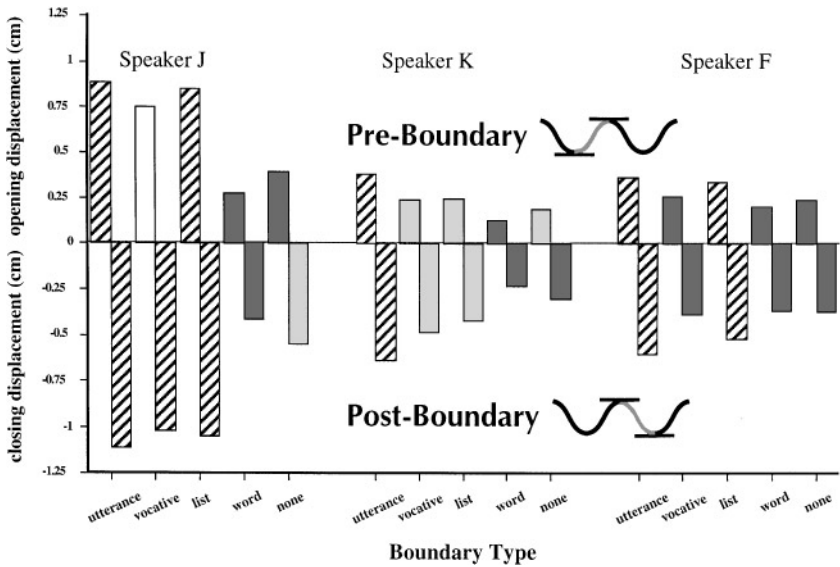
Figure 7. Preboundary opening and postboundary closing displacements for three speakers and five experimental boundary conditions. Bars with like shadings group together in post hoc tests ($p < 0.005$, except for unfilled bar for Speaker J which differs from UTTERANCE but not from LIST).

To review, speech gestures have been modeled as movements of a critically damped mass–spring system such as that shown in the equation in Fig. 8.

Within such a model, the dynamical parameter that most directly controls movement duration is the spring stiffness or natural frequency of the gesture. (See Byrd *et al.*, in press, for a review of other parameters affecting gestural duration). Gestures with a lower stiffness (Fig. 8, left panel) will reach their target position later than ones with higher stiffness (Fig. 8, right panel). Also, within this model gestures with lower stiffness will have later occurring peak velocities than ones with higher stiffness (see Fig. 8). We have attempted to assess gestural stiffness to determine if the prosodically conditioned lengthenings observed above for the postboundary closure are consistent with decreases in stiffness. To do this, we examined the time from onset to peak velocity for the post-boundary closing movements.

There is a significant effect of boundary condition on postboundary closing time-to-peak-velocities [Sp. J $F(4,75) = 30.872$, $p < 0.0001$; Sp. K $F(4,74) = 25.455$, $p < 0.0001$; Sp. F $F(4,74) = 26.218$, $p < 0.0001$]. Crucially, for all subjects, all pairwise comparisons of boundary conditions for postboundary closing-time-to-peak-velocity yield the same results as for postboundary closing durations. (Although for Speaker K, three of the post hoc comparisons for time-to-peak-velocity are significant at a weaker ($p \leq 0.03$) level; cf. $p < 0.005$ for the other comparisons). These results are shown in Fig. 9. Correlations between postboundary closing duration and time-to-peak-velocity are also high and positive for all subjects ranging from $r = 0.84$ to $r = 0.94$ ($p < 0.0001$). That is, time-to-peak-velocity behaves in the same way as duration. This indicates that local lowering of gestural stiffness at the edges of high prosodic domains may indeed be the source of boundary-adjacent lengthening.

Lower Stiffness Gesture (ω_0=3)

Higher Stiffness Gesture (ω_0=10)
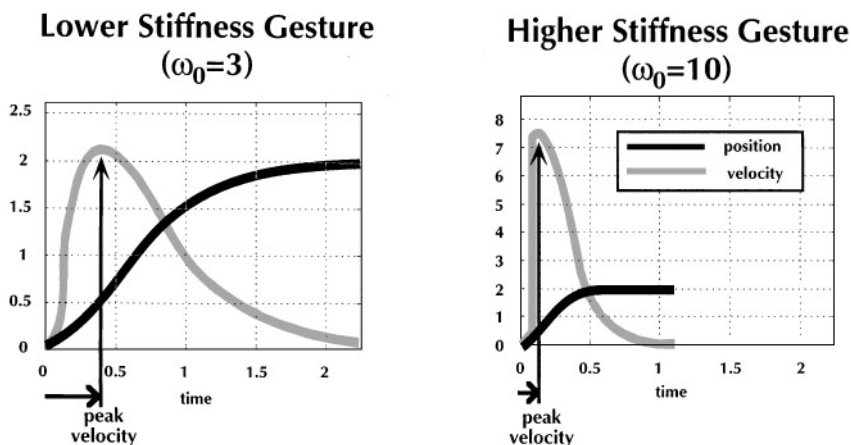
$$\ddot{x}+2\xi\omega_0\dot{x}+\omega_0^2(x-x_{\text{targ}})=0,\ \xi=1,\ x_{\text{targ}}=2$$

where: ξ=damping ratio=$1/2(b/\sqrt{mk})$,
b=damping coefficient, k=stiffness, m=mass
ω_0=natural frequency; ω_0^2=k/m

Figure 8. Two movements generated from the critically damped mass–spring equation shown. Dark trajectories are position (x) trajectories; lighter trajectories are velocity ($\dot{x}$) trajectories. Target position value is 2 in both left and right panels. Left panel shows lower stiffness gesture ($\omega_0 = 3$). This movement has longer times to reach target and to reach the peak velocity than the movement in the right panel which shows a higher stiffness gesture ($\omega_0 = 10$).

4. Modeling phrase-initial lengthening

4.1. Is stiffness the "whole story"?

The type of mass–spring gestural model described in Section 3.6 has been used by linguists and speech scientists in discussing the dynamic bases for boundary-adjacent lengthening. In this model, a gesture's intrinsic parameters—target, damping, and stiffness—remain constant throughout the gesture's period of activation. That is, a gesture's activation is described by a step function and acts to multiplicatively gate in the gesture's parameter values. Such a system is described by the following *tract-variable* (e.g., lip aperture; Browman & Goldstein, 1990; Saltzman & Munhall, 1989) equation of motion:

$$\ddot{x} = -a(t)b\dot{x} - a(t)k[x - a(t)x_0] \tag{1}$$

In this equation, the activation $a(t)$ varies as a function of time according to

$$a(t) = \begin{cases} 0 & \text{if } (t < t_{\text{on}}), \\ 1 & \text{if } (t_{\text{on}} \le t < t_{\text{off}}), \\ 0 & \text{if } (t \ge t_{\text{off}}), \end{cases} \tag{2}$$

where t_{on} and t_{off} denote, respectively, the gesture's onset and offset times. At gesture onset the activation function instantaneously changes from 0 to 1, and tract-variable
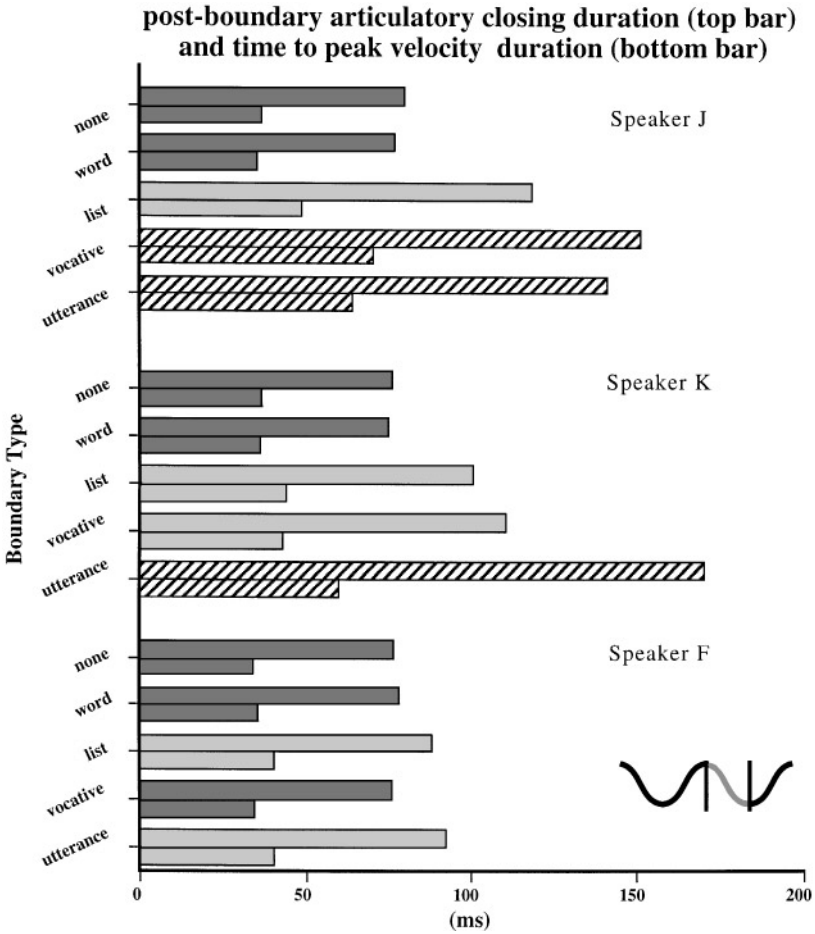
D. Byrd and E. Saltzman

Figure 9. Postboundary closing durations and absolute time-to-peak-velocities for three speakers and five experimental boundary conditions. Bars with like shadings group together in post hoc tests ($p \leq 0.03$).

system parameters go instantaneously from their default[3] values to the active gesture's intrinsic parameter values; at gesture offset activation switches to zero and the system parameters return instantaneously to their default values. Thus, during periods of gestural activation ($t_{on} \leq t < t_{off}$) Equation (1) assumes the following simpler form:

$$\ddot{x} = -b\dot{x} - k[x - x_0] \tag{3}$$

As we described earlier, this type of system displays longer time-to-peak-velocity when stiffness is decreased. Furthermore, in this step-function model, the proportional time-to-peak-velocity (i.e., the skew of the velocity profile or time-to-peak-velocity normalized by

[3] In the simulations discussed in this paper, we assume for purposes of simplicity that there is a one-to-one relation between tract-variables and model articulators. That is, only one articulator is used to generate motion of the simulated tract-variable, and thus the dynamics governing motions of the tract-variable and the model articulator are identical. In particular, the default parameter values in the tract-variable equation of motion become identical in this case with the default parameter values in the so-called *neutral attractor* defined at the articulator level (Saltzman & Munhall, 1989).

dividing by gesture duration) does not change as stiffness is varied. It has been realized for quite some time, however, that the step-function activation is an oversimplification (e.g., Bullock & Grossberg, 1988; Coker, 1976; Kröger, Schröder & Opgen-Rhein, 1995; Ostry, Gribble & Gracco, 1996; Perrier, Ostry & Laboissiere, 1996). In fact, stiffness changes within the step-function model cannot completely account for this experiment's phrase-initial lengthening data for two reasons.

4.2. *Two incorrect predictions of the step-function model*

4.2.1. *Proportional time-to-peak-velocity values*

First, the step-function model predicts proportional time-to-peak-velocities that are much too small. The peak velocities for the postboundary closing gestures in our data generally occur a little before the 50%-point of total gesture duration. However, the step-function model predicts peak velocities that occur a little before the 25%-point of the gesture (see Section 4.3 for further details).

4.2.2. *Speaker-specific kinematics*

Second, recall that the step-function model predicts invariant proportional time-to-peak-velocities. However, in some instances the postboundary closing gestures in our data display *different* proportional time-to-peak-velocities with boundary conditioned stiffness changes. Specifically, we find speaker-specific differences in the effect of boundary on the shape of the velocity profiles, i.e., on proportional time-to-peak-velocities.

For speaker F, peak velocities occur at approximately 45% of the way through the postboundary closing gesture for all boundary conditions. Similarly, for Speaker J, peak velocities occur at approximately 45% of the way through the phrase-initial closing gesture for four of the five boundary conditions and slightly earlier (41%) for the LIST boundary condition. However, for Speaker K, there is a significant effect of boundary condition ($F(4,74) = 16.589$, $p < 0.0001$) on proportional time-to-peak-velocities. The two smallest boundaries (i.e. NONE and WORD) peak at 47%, the LIST boundary at 44%, the VOCATIVE boundary at 40%, and the UTTERANCE boundary at 36%. (All differ significantly in post hoc tests except WORD and NONE. Of these comparisons, five pairs are different at the $p < 0.005$ level and the four pairs at the weaker $p < 0.05$ level.) The step-function model simply cannot account for Speaker K's data.

4.3. *The extended model*

In order to model both the gestural lengthening and the time-to-peak-velocity (*both* absolute and proportional) results, we used a gestural activation function with a ramped rise and fall, as opposed to the instantaneous rise and fall of the step-function model. In particular, we have simulated the phrase-initial gestures using half-cosines to model the rise- and fall-intervals of the associated gestural activation functions (see Fig. 10).

Additionally, since activation rises and falls gradually, the tract-variable's behavior during periods of gestural activation ($t_{\text{on}, 1} \leq t < t_{\text{off}, 2}$) is not as simple as that described by Equation (3). That is, during intervals of rising and falling gestural activation, the tract-variable's behavior is governed by a time-varying blend of default and active influences (Saltzman & Munhall, 1989). We used the following equations to model this
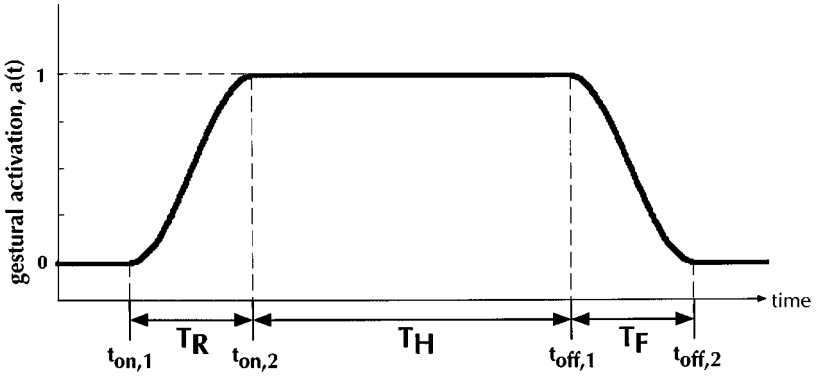
Figure 10. Activation function with half-cosine rise and fall intervals: T_R is rise-time; T_H is hold time; and T_F is fall-time of gestural activation wave. (Zero indicates no active gestural control; one indicates maximal gestural strength.)

more complex behavior:

$$\ddot{x} = \ddot{x}_g + \ddot{x}_d \tag{4a}$$

$$\ddot{x}_g = -a(t)b_g\dot{x} - a(t)k_g[x - a(t)x_{0,g}] \tag{4b}$$

$$\ddot{x}_d = (1 - a(t))[-b_d\dot{x} - k_d(x - x_{0,d})] \tag{4c}$$

In these equations, $\ddot{x}_g$ and $\ddot{x}_d$ denote respectively, the tract-variable components of acceleration due to active gestural and default tract-variable influences; system mass is assumed to be equal to 1.0; b_g, k_g, and $x_{0,g}$ denote, respectively, the active gesture's intrinsic damping, stiffness, and target parameters, b_d, k_d, and $x_{0,d}$ denote, respectively, the tract-variable's default damping, stiffness, and target parameters; and $a(t)$ varies over time as follows:

$$a(t) = \begin{cases} 0 & \text{if } (t < t_{on,1}) \\ 0.5\{1 - \cos(\omega_r[t - t_{on,1}])\}, & \text{if } (t_{on,1} \le t < t_{on,2}) \\ 1 & \text{if } (t_{on,2} \le t < t_{off,1}) \\ 0.5\{1 + \cos(\omega_f[t - t_{off,1}])\}, & \text{if } (t_{off,1} \le t < t_{off,2}) \\ 0 & \text{if } (t \ge t_{off,2}) \end{cases} \tag{5}$$

In this equation, $t_{on,1}$, $t_{on,2}$, $t_{off,1}$, and $t_{off,2}$ are defined as in Fig. 10; and ω_r and ω_f denote, respectively, the angular frequencies (rad/s) of the rising and falling intervals of the gesture's activation function. Thus, Equation (5) allows gestural simulations in which half-cosines model the rise-interval ($T_R = \pi/\omega_r$) and fall-interval ($T_F = \pi/\omega_f$) of the associated gestural activation functions (see Fig. 11).

Using Equations (4) and (5), we performed two sets of simulations, one for the pattern shown by Speaker K (who showed decreasing proportional times to peak velocity, $\%tv_p$, as gestural duration increased), and one for the pattern shown by Speaker J (arbitrarily chosen to represent the pattern shown by Speakers F and J in which $\%tv_p$ remained approximately constant with increasing gestural duration). For each set, we began by simulating a discrete phrase-initial prototype gesture for the shortest duration condition, using the mean duration of each speaker's NONE and WORD conditions
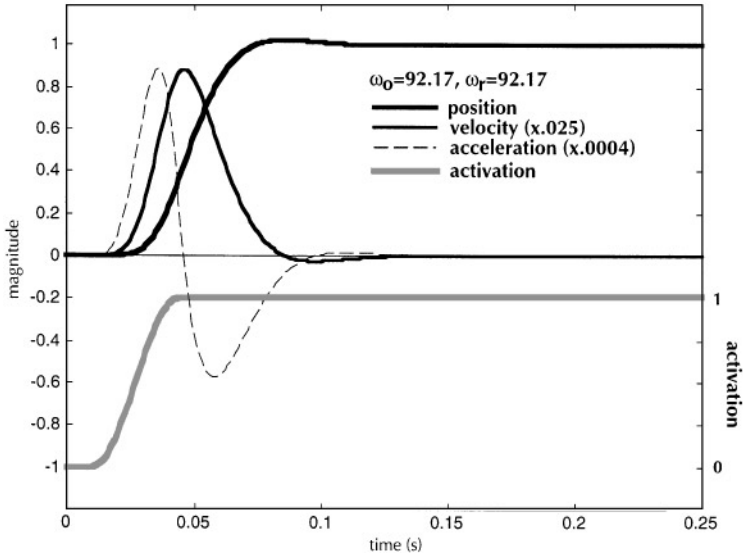
Figure 11. Top panel: Simulated gestural position, velocity (scaled $\times 0.025$), and acceleration (scaled $\times 0.0004$) with an intrinsic gestural frequency (ω_0) and a gestural activation rise frequency (ω_r) of 92.17 rad/s. Bottom panel: gestural activation wave with half-cosine rise. (Note: this simulation depicts speaker K's prototype gesture, also shown in Fig. 12, top panel).

(which did not differ significantly within speakers), and using a slightly underdamped motion form of the type commonly seen in discrete limb movements (e.g., Zelaznik, Schmidt & Gielen, 1986). For the simulated gestures, duration was defined from gesture onset (the first non-zero velocity sample) to offset (the negative-going velocity zero-crossing), time-to-peak-velocity was measured from gesture onset to peak velocity (the first negative-going acceleration zero-crossing), and $\%tv_p$ was defined as (time-to-peak-velocity)/(gesture duration).

In each prototype gesture, activation rise frequency (ω_r) was set equal to intrinsic gestural frequency ($\omega_{0,g}$) and intrinsic gestural frequency was chosen so that the simulated duration matched the mean duration of each speaker's NONE and WORD conditions ($\omega_{0,g} = 92.17$ for Speaker K; $\omega_{0,g} = 89.04$ for Speaker J). Additionally, $x_{0,g}$ was set arbitarily to 1.0; damping ratio, ξ_g, was set to 0.75 resulting in $b_g = 2\xi_g\omega_{0,g}$, and the system's default parameter values were fixed as follows: $x_{0,d} = 0.0$; $\omega_{0,d} = \sqrt{k_d} = \omega_{0,g}$; and $b_d = 2\omega_{0,d}$, yielding critical damping in the default case when no gesture is active. Despite the fact that this fitting procedure focused only on gesture duration, the simulated $\%tv_p$'s were either identical to (Speaker K; Fig.12, top panel) or within 2% of (Speaker J; Fig. 13, top panel) the values measured in the actual data. Note, however, that if activation is defined by a step function but all other parameters are the same as for each speaker's prototype gestures, $\%tv_p$ equals 23% in both cases which is far below any of the values observed in our data.

We show next that the speaker-specific patterns of gestural duration and proportional time-to-peak-velocity observed in the data can be modeled using simple changes to gestural frequency ($\omega_{0,g}$) and its activation wave's rise frequency (ω_r), varying b_g to keep ξ_g fixed at 0.75, and keeping all other parameters fixed at their prototype values.
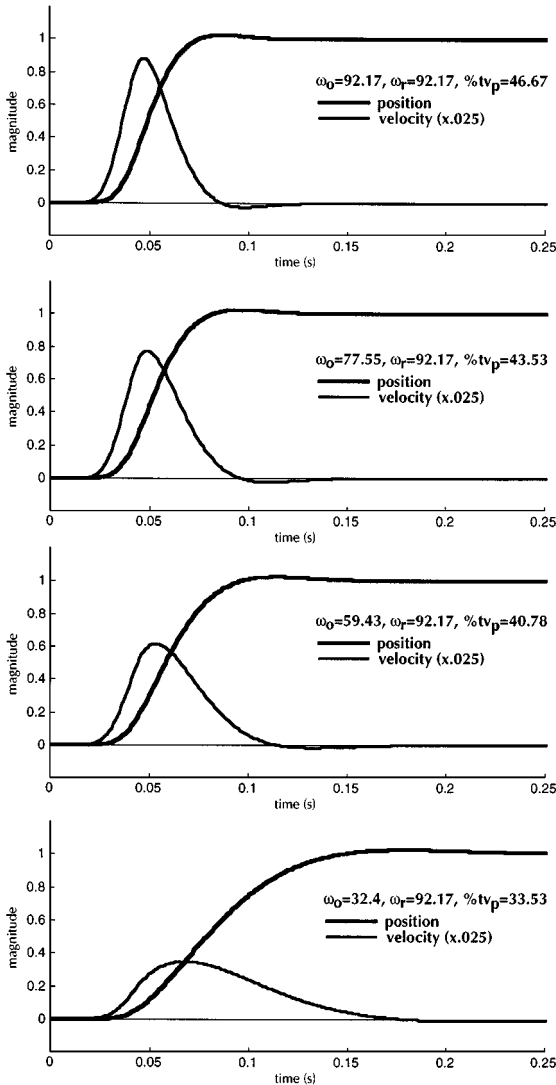
Figure 12. Speaker K's prototype pattern, showing the effects of reducing intrinsic gestural frequency (ω_0) while maintaining a constant value for activation rise frequency (ω_r). (Proportional time-to-peak-velocity ($\%tv_p$) decreases with gestural frequency).

4.3.1. *Speaker-specific dynamics*

With the ramped rise-time, the pattern shown by Speaker K (who exhibited differences in proportional-time-to-peak-velocity across the experimental conditions) can be modeled by decreasing gestural frequency but maintaining a constant activation rise-time interval (see Fig. 12). For these simulations, we decreased intrinsic gestural frequency, keeping activation rise-frequency constant, until simulated gesture duration equaled the duration obtained in the slowest, UTTERANCE condition (Fig. 12, bottom panel). Specifically, absolute time-to-peak-velocity lengthened and $\%tv_p$ dropped comparably in the
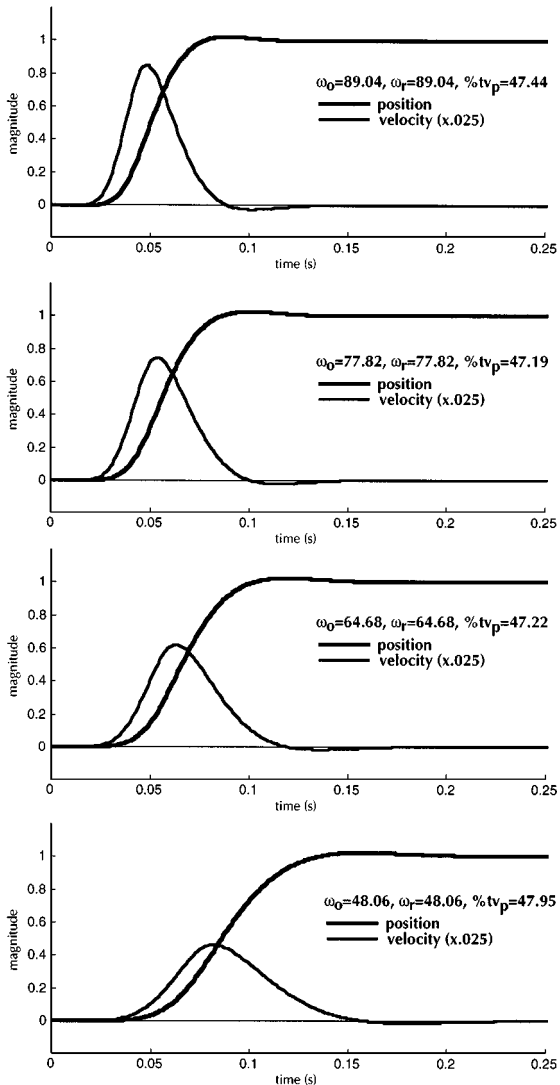
Figure 13. Speaker J's pattern, showing effects of reducing both intrinsic gestural frequency (ω_0) and activation rise frequency (ω_r), with the constraint that $\omega_r = \omega_0$. (Proportional time-to-peak-velocity ($\%tv_p$) remains approximately constant as frequency decreases.)

simulation and actual data. For illustrative purposes, we generated two intermediate duration simulations by using intrinsic gestural stiffnesses that were simply equispaced between the fastest and slowest simulation values (Fig. 12, middle panels).

An alternative form of speaker-specific dynamics within the extended model can generate the pattern of Speaker J (who displayed relatively constant $\%tv_p$ across the experimental conditions). For these simulations (Fig. 13), we decreased intrinsic gestural frequency, $\omega_{0, g}$, while constraining activation rise-frequency, ω_r, to equal intrinsic gestural frequency, until the simulated gesture duration equaled the mean duration of the

slowest, conditions VOCATIVE and UTTERANCE which did not differ significantly (Fig. 13, bottom panel). Notably, as in the actual data, $\%tv_p$ remained constant in the simulations across yoked decreases in stiffness and increases in activation rise-time. As for Speaker K's simulation set, we generated two intermediate duration simulations for each pattern type using intrinsic gestural stiffnesses that were simply equispaced between the fastest and slowest simulation values (Fig. 13, middle panels).

Finally, we note that a third speaker-specific pattern of change in velocity profile symmetry has been reported in the speech production literature (Shaiman, Adams & Kimelman, 1997). For these subjects, proportional time-to-peak-velocity increased as gesture duration increased. We note that this type of pattern is produced in simulations for which gesture duration is increased using increases in activation rise-time alone while holding gestural stiffness constant. For example, if $\omega_{0,g}$ is held constant as Speaker J's prototype gesture's value ($\omega_{0,g} = 89.04$) and only ω_r is decreased successively from its prototype value ($\omega_{r,} = 89.04$) to produce increases in gestural duration, the $\%tv_p$ values increase successively as shown in the following (ω_r, duration [s], $\%tv_p$) triplets: (89.04, 0.078, 47.44), (74.43, 0.083, 50.6), (56.12, 0.094, 56.38), (27.57, 0.146, 64.38).

4.3.2. Truncation and interparameter constraints

It is possible that intergestural truncation by preceding and following gestures contributes to shaping the gestural velocity profiles analyzed in our speakers' data. While we do not deny that truncation may influence velocity profile shapes, we note that several studies of *discrete* targeting movements performed by the limbs or speech articulators (e.g., Ostry, Cooke & Munhall, 1987; Nagasaki, 1989; Müller & Stelmach, 1992) have reported duration-conditioned changes in velocity profile symmetry that are similar to those we described above. In essence, discrete movements are single gestures that are not part of an overall gestural sequence and therefore, by definition, cannot be truncated by (non-existent) preceding and following gestures. Thus, we argue that although truncation may be a contributing factor in shaping gestural velocity profiles during ongoing speech production, it is not a necessary factor. Variations in gestural activation rise time, yoked in simple ways to variations in gestural stiffness, can account parsimoniously for the observed data patterns (see also Bullock & Grossberg, 1988, for a related approach in which changes to a so-called GO-signal are analogous to changes in our model's activation trajectory).

Thus, we interpret the results of the simulations reported above to imply that the standard mass–spring model of articulatory gestures assumed in much of the phonetics literature must be elaborated to include an activation rise-time parameter. This parameter, *at least for modeling our data*, does not seem to be manipulated in a way independent of the existing stiffness parameter. With very simple constraints between gestural stiffness and rise time, we were able to capture two types of data pattern which are similar in the manner in which gesture duration and absolute time-to-peak-velocity lengthened with lowered stiffness, but which are distinguished by whether proportional time-to-peak-velocity either decreased or did not change.

We are encouraged by these results to look for dependencies among motor control parameters and the way such parameters are influenced by linguistic structure of various sorts. For example, although the effects of prosody on gestural amplitude are not captured in the current modeling work, the strategy of attempting to identify the parameter changes that are the source of observed kinematic variation is a promising one.

5. Summary and conclusions

To summarize, multiple levels of prosodic boundaries can be distinguished by the temporal and spatial articulatory characteristics of the boundary-adjacent gestures. In a dynamical systems framework, all subjects appeared to use a lowered gestural stiffness as the main source of lengthening. However, a ramped activation rise-time also appears to be a necessary parameter. Prosodically conditioned variations of this rise-time parameter, when they occurred, were constrained to be simple functions of gestural stiffness itself.

These results and others (e.g., those of Beckman and colleagues and Byrd *et al.*, (in press) have led us to view boundary-adjacent lengthening as a local slowing in the immediate vicinity of prosodic boundaries of sufficient strength or of a sufficiently high position in the prosodic hierarchy. Determining whether these boundaries bear a one-to-one relationship with those prosodic boundaries demonstrated to be active in defining domains for phonological processes represents a promising area of research for the future.

References

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

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

Beckman, M. E. & Ayers, G. M. (1994) Guidelines for ToBI Labelling Available:http://www.ling.ohio_state Phonetics/E_ToBI/singer_tobi.html

Beckman, M. E. & Pierrehumbert, J. (1986) Intonational structure in Japanese and English, *Phonology Yearbook*, **3,** 255–309

Browman, C. P. & Goldstein L. (1990) Tiers in articulatory phonology, with some implications for casual speech. In *Papers in laboratory phonology I: between the grammar and the physics of speech* (J. Kingston & M. E. Beckman, editors), pp. 341–338. Cambridge, U.K. Cambridge University Press

Browman, C. P. & Goldstein, L. (1995) Gestural syllable position effects in American English. In *Producing speech: contemporary issues. For Katherine Safford Harris* (F. Bell-Berti & L. Raphael, editors), pp. 19–34. Woodbury, NY: AIP Press

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

Byrd, D. (in press) Intergestural coordination adjacent to multiple prosodic boundaries. In *Proceedings of the 135th Meeting of the ASA*, June, 1998, Seattle, Washington.

Byrd, D., Kaun, A., Narayanan, S. & Saltzman, E (in press) Phrasal signatures in articulation, *Papers in laboratory phonology V*

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

Cooper, A. (1991) *An articulatory account of aspiration in English.* Unpublished doctoral dissertation, Yale University

de Jong, K. (1991) An articulatory study of consonant-induce vowel duration changes in English, *Phonetica*, **48,** 1–17

de Jong, K. (1995) The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation, *Journal of the Acoustical Society of America*, **97,** 491–504

Dilley, L., Shattuck-Hufnagel, S. & Ostendorf, M. (1996) Glottalization of word-initial vowels as a function of prosodic structure, *Journal of Phonetics*, **24,** 423–444

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

Engen, T. (1971) Psychophysics. In *Experimental Psychology* (J. W. Kling & L. A. Riggs, editors) New York: Holt, Rinehart & Winston

Fougeron, C. & Keating, P. (1997) Articulatory strengthening in prosodic domain-initial position, *UCLA Working Papers in Phonetics*, **92,** 61–87 and to appear in *Journal of Acoustical Society of America* **101,** 3728–3740

Fougeron, C. & Keating, P. (1996) Variations in velic and lingual articulation depending on prosodic position: results for 2 French speakers, *UCLA Working Papers in Phonetics*, **92,** 88–96

Goldstein, L. (1992) Comments on chapters 3 and 4. In *Papers in laboratory phonology II*: *gesture, segment, prosody* (G. J. Docherty & D. R. Ladd, editors) pp. 120–124. Cambridge: Cambridge University Press

Gracco, V. L. & Nye, P. W. (1993) Magnetometry in speech articulation research: some misadventures on the road to enlightenment, *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*, **31,** 91–104

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

Inkelas, S. (1990) *Prosodic Constituency in the Lexicon*. New York: Garland Publishing

Ivry, R. B. & Hazeltine, R. E. (1995) Perception and production of temporal intervals across a range of durations: evidence for a common timing mechanism, *Journal of Experimental Psychology*: *Human Perception and Performance*, **21**(1), 3–18

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

Keating, P. A. (1995) Segmental phonology and non-segmental phonetics. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (K. Elenius & P. Branderud, editors) Vol. 3, pp. 26–32

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

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

Kröger, B., Schröder, G. & Opgen-Rhein, C. (1995) A gesture-based dynamic model describing articulatory movement data, *Journal of the Acoustical Society of America*, **98**(4)**,** 1878–1889

Ladd, D. R. (1986) Intonational phrasing: the case for recursive prosodic structure, *Phonology Yearbook*, **3,** 311–340

Ladd, D. R. & Campbell, W. N. (1991) Theories of prosodic structure: evidence from syllable duration, *ICPhs XII* **2,** 290–293

Löfqvist, A. (1993) Electromagnetic transduction techniques in the study of speech motor control. *PHONUM*: *Reports from the Department of Phonetics, University of Umeå*, Vol. 2, pp. 87–106

Löfqvist, A. & Gracco, V. (1997) Lip and jaw kinematics in bilabial stop consonant production, *Journal of Speech, Language, and Hearing Research*, **40,** 877–893

Müller, F. & Stelmach, G. E. (1992) Prehension movements in Parkinson's disease. In *Tutorials in motor behavior II* (G. E. Stelmach & J. Requin, editors), pp. 307–319. New York: Elsevier

Nagasaki, H. (1989) Asymmetric velocity and acceleration profiles of human arm movements, *Experimental Brain Research*, **74,** 319–326

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

Oller, D. K. (1973) The effect of the position in utterance on speech segment duration in English, *Journal of the Acoustical Society of America*, **54,** 1235–1247

Ostry, D. J. Cooke, J. D. & Munhall, K. G. (1987) Velocity curves of human arm and speech movements, *Experimental Brain Research*, **68,** 37–46

Ostry, D. J. Gribble, P. and 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

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

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I. & Jackson, M. (1992) Electromagnetic midsagital articulometer (EMMA) systems for transducing speech articulatory movements, *Journal of the Acoustical Society of America*, **92,** 3078–3096

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. & Talkin, D. (1992) Lenition of /h/ and glottal stop. In *Papers in laboratory phonology II*: *gesture, segment, prosody* (G. J. Docherty & D. R. Ladd, editors) Cambridge, U.K.: Cambridge University Press

Repp, B. (1997) Variability of timing in expressive piano performance increases with interval duration, *Psychonomic Bulletin and Review*, **4**

Ross, K. & Ostendorf, M. (1996) Prediction of abstract prosodic labels for speech synthesis, *Computer Speech and Language*, **10**, 155–185

Rubin, P. E. (1995) HADES: a case study of the development of a signal analysis system. *Applied speech technology*, pp. 501–520. Boca Raton, FL, CRC Press

Saltzman, E. L. (1986) Task dynamic coordination of the speech articulators: a preliminary model. Generation and modulation of action patterns, *Experimental Brain Research*, Series 15, 129–144

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

Saltzman, E. Löfqvist, A. & Mitra, S. (in press) "Clocks" and "glue"—global timing and intergestural cohesion. *Papers in laboratory phonology V*. This work was presented at the Conference on Laboratory Phonology, Evanston, Illinois, July, 1996

Selkirk, E. O. (in press) The prosodic structure of function words. In *International conference on boostrapping from speech to grammar in early acquisition*, (J. Martin & K. Demuth, editors) Brown University. Providence, RI: Lawrence Erlbaum.

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

Shaiman, S., Adams, S. G. & Kimelman, M. D. Z. (1997) Velocity profiles of lip protrusion across changes in speaking rate, *Journal of Speech and Hearing Research*, **40**, 144–158

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

Vaissière, J. (1988) Prediction of velum movement from phonological specifications, *Phonetica*, **45**, 122–139

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. (1992) Segmental durations in the vicinity of prosodic phrase boundaries, *Journal of the Acoustical Society of America*, **91**, 1707–1717

Wilbur, R. & Zelaznik, H. (1997) Kinematic correlates of stress and phrase position in ASL. Talk presented at the 1997 meeting of the Linguistic Society of America

Zelaznik, H. N., Schmidt, R. A. & Gielen, S. C. A. M. (1986) Kinematic properties of rapid aimed hand movements, *Journal of Motor Behavior*, **18**, 353–372