

# Gestural De-Aggregation via Prosodic Structure

Dani Byrd, Rebeka Campos-Astorkiza, Michael Shepherd

Department of Linguistics, University of Southern California, Los Angeles, CA, USA

dbyrd@usc.edu, rebekaca@usc.edu, mshepher@usc.edu

**Abstract.** *It has been proposed that when two gestures involving the same articulator are organized with sufficient overlap, the resulting ‘gestural aggregation’ blends or averages the two movements to result in a smooth, often plateau-like kinematic trajectory (Munhall & Löfqvist 1992). We analyze the effects of a phrasal boundary on sequences of two consonants across a word boundary. Prosodic structure has been shown to affect the temporal overlap among gestures (Byrd et al. 2000). Given the overlap account of gestural aggregation, we expect the case of juncture geminates—two identical sequential consonants—to be influenced by prosodic structure in a similar fashion as other coproduced gestures are. In fact, our results show that, as expected, the overlapped gestures are temporally pulled apart or de-aggregated across a phrase boundary. [Supported by NIH DC 03172]*

## 1. Introduction

### 1.1. Articulatory overlap and juncture geminates

Coproduction in speech occurs when two gestures temporally overlap. When these gestures share all or some of their articulators, they have been described as *blended* in that the gestural parameters are (basically) averaged (Saltzman & Munhall 1989). When two abutting consonants are identical, they are called juncture geminates. The temporal behavior of juncture geminates is important to study as they are canonical *sequences* but are produced as *single* articulator constriction movements. When consonants abutting at word edges use different articulator sets, we can observe their coproduction rather clearly; we see two different constrictions being made in a temporally overlapped fashion (Browman & Goldstein 1990). However, when the two consonants share the same articulator, as in the case of juncture geminates, their individual characteristics become more difficult to observe. This type of coproduction has been called *gestural aggregation* (Munhall & Löfqvist 1992). The fact that two individual gestures are being coarticulated may or may not be obvious from the articulatory movements. For example, Munhall and Löfqvist (1992) examining the aggregation of two overlapping laryngeal gestures across a word boundary observe a single smooth movement at fast speaking rates.

For cases when only one continuous movement is observed, two analyses seem possible. First, overlap with blending of two separate gestures could simply result in a movement with only a single displacement extremum (Saltzman & Munhall 1989). However, summation of the two underlying gestures could also result in a single trajectory (Munhall & Löfqvist 1992). Previous studies suggest that juncture geminates are the result of extreme overlap and blending rather than a summation process. The degree of overlap between coproduced consonants varies with the speech rate so that overlap increases as the rate gets faster (Byrd & Tan 1996, Hardcastle 1985). Munhall and Löfqvist (1992) looked at laryngeal gestures in juncture geminates across different speech rates. In fact, at slower rates, Munhall and Löfqvist observed two distinct

laryngeal movements; at faster rates, a single smooth trajectory was present. These results suggest that the single movement in fast speech was the consequence of great overlap between the two gestures. Comparable results were reported in Löfqvist and Yoshioka (1981).

Juncture geminates can also be differentiated from single consonants in terms of their durational characteristics and their degree of articulator displacement. Kelso & Tuller (1987) note that the results of gestural summation would be a larger gesture with increased amplitude and steeper onset and offset slopes. Byrd (1995) used electropalatography (EPG) to investigate lingual juncture geminates in English. Lingual-palatal contact patterns indicate that these juncture geminates were produced with a single raising and lowering of the tongue. Byrd (1995) found the coproduced articulation for juncture geminates to be longer than the movement for a single gesture. Byrd's results also indicated no consistent increase in maximum contact for the geminated consonants relative to single consonants syllable onsets. Also, constriction formation and release contact slopes of the juncture geminates appeared comparable to those of singleton onset and coda slopes, respectively. These data suggest that a blending process that averages the spatial target values for two overlapping consonantal gestures is at work (Byrd 1995). Munhall and Löfqvist (1992) also found no consistent tendency for the combined single movement to be larger than an individual (non-coproduced) movement, although a simulated summation of the gestures predicts such a difference. Vaxelaire (1995) examined X-ray film data on constrictions for lingual juncture geminates in French and found slightly different results in that the extent of contact between the tongue and palate is greater for the French juncture geminates than for the similar single consonant. Geminates had longer articulatory durations than singletons, as predicted from previous studies.

## **1.2. Prosodic effects on articulatory gestures**

Prosodic structure affects the spatial and temporal characteristics of individual gestures, as well as the relative coordination among different gestures (Byrd & Saltzman 2003, Byrd et al. 2000). With respect to the influence on the intragestural characteristics, articulatory studies have shown that gestures become longer near prosodic boundaries (e.g., Edwards et al. 1991, Beckman & Edwards 1992, Byrd & Saltzman 1998, Byrd et al. 2000, Cho 2005).

Articulatory studies have examined overlap patterns under the influence of prosodic structure and found that temporal overlap is less among gestures separated by or adjacent to a boundary (McClellan 1973, Byrd et al. 2000, Byrd & Saltzman 1998, Byrd & Choi 2006, *subm.*). Byrd et al. (2000) analyzed the timing patterns across a phrasal boundary of the tongue tip and upper lip gestures for two nasal consonants [n, m] in Tamil. The time between extrema was significantly longer in the boundary condition. As for the relative timing, the word-final gesture reached its extremum position significantly earlier in the word-initial gesture when a phrase boundary intervened. These results and those of Byrd and Choi (2006, *subm.*) and simulations in Byrd & Saltzman (2003) show that gestures are pulled apart across a phrasal boundary, resulting in less temporal coproduction between these gestures.

Critically, these findings have been modeled as a local slowing in the temporal pacing of constriction gesture activation trajectories within Byrd and Saltzman's  $\pi$ -gesture model of phrasal structure in speech production (Byrd & Saltzman 2003). If

phrase boundaries can be understood, as Byrd and Saltzman suggest, by transitory warping of the local speech rate, then juncture geminates, just like the sequences of non-identical consonants in the above studies, should show lesser overlap when they occur across a phrase boundary, perhaps even displaying signs of de-aggregation or ‘pulling apart.’

#### 1.4. Goals of the present study

This study evaluates the effects of phrasal boundaries on the intra- and intergestural characteristics of blended gestures produced with the same articulator. This allows us to investigate the process of gestural aggregation and the manner in which it is affected by prosodic structure. Juncture geminates are expected to be affected by prosodic boundaries in the same way as other gestures. The intragestural prosodic lengthening is expected to be reflected in longer constriction formation duration and somewhat in longer plateau duration. The intergestural changes in overlap when the juncture geminates occur across a phrase boundary are expected to be manifested in longer plateau duration as the two abutted gestures slide apart, possibly yielding longer total sequence duration. In fact, the intervening phrase boundary condition may even be associated with a two-peaked gestural trajectory.

## 2. Method

The experiment stimuli were constructed to test the effects of an utterance-level phrase boundary between two consonants  $C_1$  and  $C_2$  produced using the same articulator. The stimulus sentences are given in Table I. Contrast in prosodic structure was generated via the use of nominal compounds, specifically, proper name compounds, versus sequence of proper names spanning a phrase boundary. That is, the target consonants were either part of the same phrase (i.e. a compound name) or separated by an intonational phrase boundary. The target sequences were the juncture geminate [dd] for the same articulator and manner condition, [dz] for the same articulator but different manner condition, and [db] for the different articulator condition. The sentences were blocked by boundary condition and ten consecutive repetitions of each sentence were recorded, yielding a total of 60 tokens for each subject. Three native speakers of American English participated in the study and will be referred to as NP, KZ and JB.

Table I. Stimuli used in experiment.

intervening boundary?	sentences
	<i>juncture geminates [dd] (gestural aggregation):</i>
Yes	JIMMY loved Dodd. Deb bet DAD didn't know that.
No	JIMMY loved Dodd-Deb. Bet TAD didn't know that.
	<i>same-articulator, non-geminate [dz]:</i>
Yes	KIMMY loved Dodd. Zeb bet DAD didn't know that.
No	KIMMY loved Dodd-Zeb. Bet TAD didn't know that.
	<i>different-articulators, singleton [d]:</i>
Yes	TIMMY loved Dodd. Bub bet TAD didn't know that.
No	TIMMY loved Dodd-Bud. Bet DAD didn't know that.

The Electromagnetic Midsagittal Articulometer (EMMA) system (Perkell et al. 1992) was used to track the horizontal (x) and vertical (y) movements of transducers adhered to the tongue tip and lips. The transducer trajectory data were sampled at 625 Hz and the acoustic data at 20kHz. The data were corrected for head movement using reference transducers on the nose and maxilla and were rotated to the occlusal plane.

In this paper, the focus is on the kinematic behavior of the tongue tip (TT) gesture produced in each of the target sequences, i.e. [dd], [dz], [db]. Note that for [dd] and [dz], a single smooth movement of the tongue tip was produced by all speakers in all conditions with one exception. (Speaker JB displayed a somewhat different behavior for the [dd] and [dz] in the boundary condition. This will be later discussed). Using the trajectory analysis software MVIEW (under development by Mark Tiede), five points were defined for the TT trajectory: gestural onset, plateau onset, maximum constriction, plateau offset and gestural offset. Time and spatial value for these landmarks were derived from the tongue tip x and y velocity trajectories. The gestural onset and offset correspond to the beginning of the constriction formation and the end of the release. They were calculated as threshold-crossing points in the tangential velocity trajectory, where the threshold was defined as a percentage of the range between the maximum and minimum local speed. The threshold was set to 0.2. The plateau onset and offset correspond to the beginning and end of the constriction plateau. These were also calculated as threshold-crossing points, with a value of 0.3. Maximum constriction is defined as a zero-velocity crossing point. Additionally, peak velocity timepoints during both the closing and opening movements were recorded. Figure 1 illustrates the seven measured timepoints.

It is worth noting that speaker JB, unlike the other two speakers, produced the sequences [d#z] and [d#d] in the boundary condition (only) with two distinct tongue tip gestures. Figure 2 illustrates these two-gesture productions of [d#z] and [d#d] across a phrasal boundary. For these tokens, the landmarks from both gestures were combined

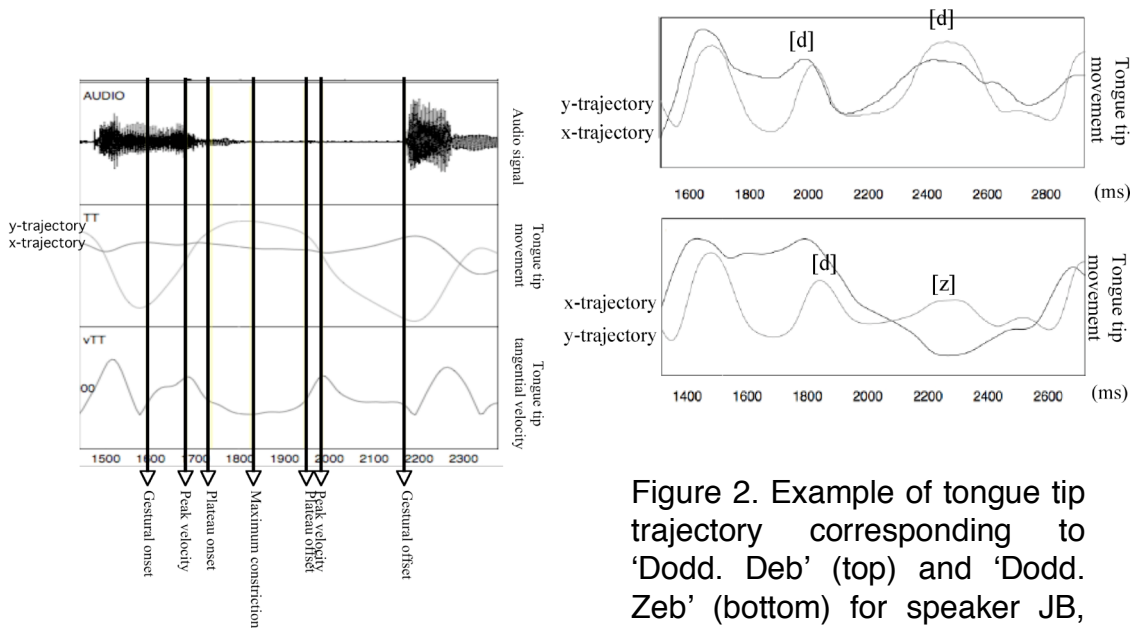


Figure 1. Timepoint markings.

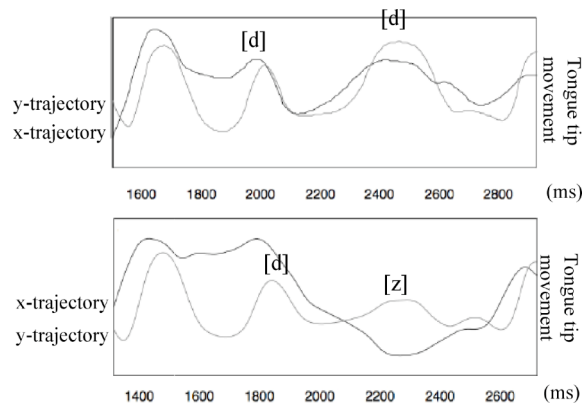


Figure 2. Example of tongue tip trajectory corresponding to 'Dodd. Deb' (top) and 'Dodd. Zeb' (bottom) for speaker JB, illustrating the two-gesture production of sequences [d#d] and [d#z] across a boundary.

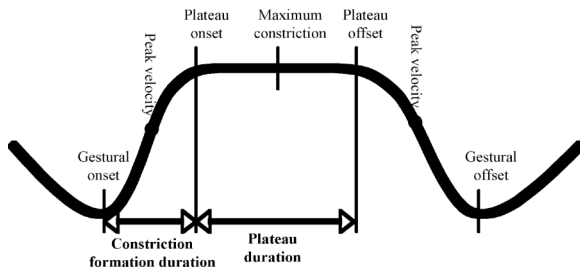


Figure 3. Articulatory landmarks & derived dependent variables.

for the analysis. Specifically, the gestural onset, first peak velocity and plateau onset were taken from the first TT gesture, and the plateau offset, second peak velocity and gestural offset were selected from the second gesture. The maximum constriction point was taken from the gesture with the highest plateau. For [d#z], this point was from the first gesture, and for [d#d] it was located in the second gesture.

Based on the tongue tip movement landmarks, two articulatory intervals were derived from the measured timepoints and used as dependent variables:

1. Constriction formation duration: time from gestural onset to plateau onset;
2. Plateau duration: time from plateau onset to plateau offset.

The articulatory landmarks and derivable variables are represented in Figure 3. Constriction formation duration is taken as an indicator of the closing gesture profile, and plateau duration serves as an index of gestural overlap for blended gestures. Sequence type ([dd], [dz], [db]) is not predicted to have an effect on constriction formation duration but may affect plateau duration. The effect of the presence or absence of a phrase boundary intervening between the two consonants is hypothesized to be significant for both duration variables but may be different in degree for the plateau duration depending on the sequence type.

Individual two-factor ANOVAS for each subject were conducted in order to evaluate the effects of boundary and sequence ([dd], [dz], [db]) on constriction formation duration and plateau duration. PLSD post-hoc tests were carried out for both dependent variables to evaluate pairwise comparisons among the different sequence types. The significance level for all the statistical tests was set at  $p < .05$ .

### 3. Results

#### 3.1. Constriction formation interval

A main effect of boundary on the constriction formation duration is obtained for all the subjects: Speaker JB [ $F(1, 50)=44.4, p < .0001$ ], speaker KZ [ $F(1, 48)=43.2, p < .0001$ ], speaker NP [ $F(1, 48)=207.9, p < .0001$ ]. The constriction formation interval is longer in

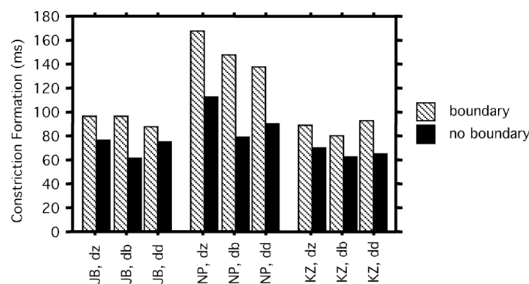


Figure 4. Boundary & CC effects on constriction formation duration.

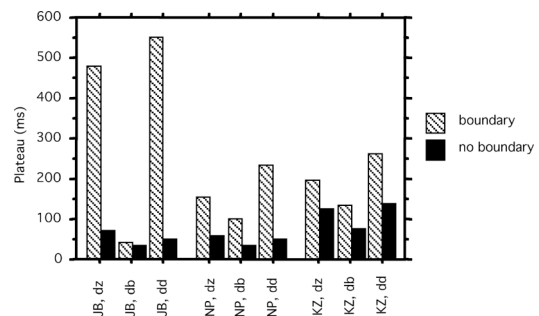


Figure 5. Boundary & CC effects on plateau duration.

the presence of a phrase boundary. As for the sequence type, this independent variable has a significant effect on constriction formation only for speaker NP [ $F(2, 48)=19.5, p<.0001$ ]. According to a PLSD post-hoc test, NP's constriction formation for [dz] is different from [dd] and [db] ( $p<.0001$ ), so that [dz] presents a longer duration; [dd] and [db] are not significantly different from each other. Speakers JB and KZ do not show a significant effect of sequence type. The interaction between boundary and sequence is significant only for speaker JB [ $F(2, 50)=3.8, p<.0001$ ]. It is a magnitude interaction such that the boundary has a stronger effect on [d] constriction duration in [db] than in [dd] or [dz]. Figure 4 shows the constriction formation duration for the different boundary and sequence conditions split by speaker. Table II includes the corresponding means and standard deviations.

Table II. Constriction and plateau durations (ms) [*b* indicates boundary condition and *nb* indicates no boundary condition]

		Constriction formation duration (ms)						Plateau duration (ms)					
		[db]		[dd]		[dz]		[db]		[dd]		[dz]	
		<i>b</i>	<i>nb</i>	<i>b</i>	<i>nb</i>	<i>b</i>	<i>nb</i>	<i>b</i>	<i>nb</i>	<i>b</i>	<i>nb</i>	<i>b</i>	<i>nb</i>
Speaker JB	<i>mean</i>	96	61.6	87.1	75.4	96.8	76	40.8	32	549.3	51.4	480.8	71.2
	<i>SD</i>	9.9	8.5	14.1	4.3	20.1	9.4	2.5	5.3	41.8	4.3	119.7	12.2
Speaker NP	<i>mean</i>	148	78.4	138	89.6	167	112.8	101	33.6	235	49.6	156	60
	<i>SD</i>	12.1	6.3	21.3	10.5	20.7	13.3	57.2	5.1	54.4	8.3	69.3	12.1
Speaker KZ	<i>mean</i>	80.5	63	92.2	64.7	89.3	69.8	133	75	264.2	139.3	193.8	124.9
	<i>SD</i>	10.2	15.9	12.6	6.4	14.7	8.6	48.4	14.1	67.7	15.3	34.5	17.7

### 3.2. Plateau duration

All speakers show a main effect of boundary on the plateau duration: Speaker JB [ $F(1, 50)=443.7, p<.0001$ ], speaker KZ [ $F(1, 48)=66.1, p<.0001$ ], speaker NP [ $F(1, 48)=108.8, p<.0001$ ]. The plateau is longer in the boundary condition. The type of sequence also shows a main significant effect for all speakers, such that the plateau is longer for [dd] & [dz] than for [db]: Speaker JB [ $F(2, 50)=139.8, p<.0001$ ], speaker KZ [ $F(2, 48)=30, p<.0001$ ], speaker NP [ $F(2, 48)=15.1, p<.0001$ ]. The results from PLSD post-hoc tests further show that for all the speakers, the three sequences present plateau durations different from each other at a significance level of  $p<.05$ . As for the interaction between sequence and boundary, it was also significant for all the speakers: Speaker JB [ $F(2, 50)=75.9, p<.0001$ ], speaker KZ [ $F(2, 48)=4.03, p<.024$ ], speaker NP [ $F(2, 48)=10.17, p<.0002$ ]. This magnitude interaction is such that the plateau duration for [dd] and [dd] is more lengthened by the presence of a boundary than that for singleton [d] in the [db] sequence. Figure 5 shows each speaker's plateau duration for the different boundary and sequence conditions. Table II reports the means and standard deviations.

Speaker JB shows a high mean and standard deviation for the [d#z] and [d#d] boundary condition. As noted above, JB produced this set of stimuli with two distinct tongue

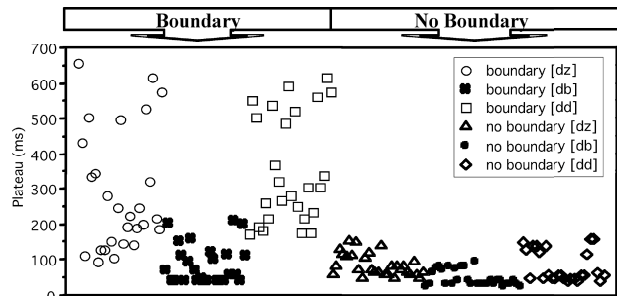


Figure 6. Scattergram showing the distribution of plateau durations (all speakers pooled).

tip gestures, unlike the two other speakers. This fact accounts for the exceptionally long plateaus for this particular sequence. Furthermore, JB's [d#z] and [d#d] productions have a great amount of variability compared to the other sequences. In fact, this same situation is found across all the subjects; Speakers KZ and NP also display a high variability in plateau duration for the boundary condition. The scattergram in Figure 6 shows the distribution of plateau durations for all the speakers pooled. Byrd et al. (2000) also found greater variability in intergestural timing under the boundary condition. They interpret this as an indication that intergestural timing is less constrained when the consonants are in separate phrasal domains.

#### **4. Discussion and conclusion**

The data presented above confirm the predictions of the overlap account of gestural aggregation. According to this analysis, aggregated gestures should display similar patterns for onset slopes to single productions but show longer plateau duration than non-coproduced gestures. Our results are in accordance with these predictions. We find no effect of sequence on constriction formation duration indicating that this variable is comparable for singleton [d] productions and [d] in abutted [dd] and [dz]. Only speaker NP presents a longer constriction formation for the tongue tip gesture in the sequence [dz] than that in [db] and [dd]. However, even for this speaker, the juncture geminate and the singleton [d] have similar onset slopes. Thus, in line with previous studies, we find that gestural aggregation is not a summation process but rather the result of temporal overlap.

The results reported here shed light on effects of phrase structure on gestural aggregation. Our data suggest that phrase boundary effects on juncture geminates are comparable to those on singleton gestures. Aggregation of gestures spanning two phrasal domains result in longer constriction formations and plateau durations than in cases when the juncture geminates are phrase medial. Longer constriction formations and plateaus in juncture geminates can be straightforwardly understood as the result of intragestural lengthening and lesser intergestural overlap.

In fact, our results show that the plateau durations increased both for single and aggregated gestures. However, as the interaction between sequence and boundary indicates, the plateau duration for singleton [d] is lengthened to a lesser extent than the coproduced articulations of [dz] and [dd]. This asymmetry in the boundary effect seems to be the result of differences in intergestural patterning. Tongue tip aggregation is the result of overlap between two separate gestures, whereas singleton [d] is formed by only one gesture. A phrase boundary, thus, modifies the intragestural characteristics of both productions, but in the case of juncture geminates, it also decreases the overlap between the two underlying gestures resulting in a substantial increase of the plateau duration. Moreover, the effect of boundary on de-aggregation can even result in two-peak productions for juncture geminates, as found for speaker JB. This suggests that prosodically-driven modulation of gestural overlap (Byrd & Saltzman 2003) can result not only in quantitative but also qualitative differences in output articulatory kinematics. The presence of a phrase boundary can even lead to total de-aggregation of two gestures involved in the production of a juncture geminate.

To conclude, this contribution is two-fold. First, we provide further evidence supporting the overlap account of gestural aggregation. Second and most importantly,

we analyze prosodic effects on juncture geminates, showing that gestural de-aggregation can occur across a phrase boundary due to a sufficient decrease in overlap.

## References

- 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 Linguistics Structure*. Tokyo 359-375.
- Browman, C. & Goldstein, L. 1990. Tiers in articulatory phonology, with some implications for casual speech. In Kingston, Beckman (eds.) *Papers in Laboratory Phonology I: between the grammar and physics of speech*. Cambridge: Cambridge UP, 341-376.
- Byrd, D. (1995) Articulatory characteristics of single and blended lingual gestures. In K. Elenius & P. Branderud (eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 438-441.
- Byrd, D., and Choi, S. (2006, subm.) At the juncture of prosody, phonology, and phonetics—The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. *Papers in Laboratory Phonology 10*. Mouton
- 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. Acquisition and the Lexicon*. Cambridge University Press, 70-87.
- Byrd, D. & Saltzman, E. (1998) Intragestural dynamics of multiple phrasal boundaries. *Journal of Phonetics* 26, 173-199.
- Byrd, D. & Saltzman, E. (2003) The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31, 149-180.
- Byrd, D. & Tan, C. C. (1996) Saying consonant clusters quickly. *J. of Phon.*, 24(2), 263-282.
- Cho, T. (2005) Manifestation of prosodic structure in articulation: evidence from lip movement kinematics in English. In L. Goldstein (ed.) *Laboratory Phonology 8: Varieties of Phonological Competence*. New York: Walter De Gruyter Inc.
- Edwards, J., Beckman, M. E., & Fletcher, J. (1991) The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America* 89, 369-382.
- Hardcastle, W. J. (1985) Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication*, 4, 247-263.
- Kelso, J.A.S. & Tuller, B. (1987) Intrinsic timing in speech production: Theory, methodology, and preliminary observations. In Keller & Gopnik (eds.), *Sensory and Motor Processes in Language*. Hillsdale, NJ: Erlbaum, 203-222.
- McClellan, M. (1973) Forward coarticulation of velar movement at marked junctural boundaries. *Journal of Speech and Hearing Research* 16, 286-296.
- Löfqvist, A. & Yoshioka, H. (1981) Laryngeal activity in Icelandic obstruent Production. *Nordic Journal of Linguistics* 4, 1-18.
- Munhall, K. G. & Löfqvist, A. (1992) Gestural aggregation in speech: laryngeal gestures, *Journal of Phonetics* 20, 93-110.
- Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I. & Jackson, M. 1992. Electromagnetic midsagittal articu-lometer (EMMA) systems for transducing speech articulatory movements. *JASA* 92: 3078-3096.
- Saltzman, E. L. & Munhall, K. G. (1989) A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1, 333-382.
- Vaxelaire, B. (1995) Single vs. double (abutted) consonants across speech rate: X-ray and acoustic data from French. In K. Elenius & P. Branderud (eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 384-387.