Evaluation of Juncture Strength using Articulatory Synthesis of Prosodic Gestures and Functional Data Analysis

Benjamin Parrell, Sungbok Lee , Dani Byrd

Department of Linguistics, University of Southern California
parrel@usc.edu, sungbokl@usc.edu, dbyrd@college.usc.edu

Abstract

Prosodic boundary gestures (pi-gestures) (Byrd & Saltzman, J. Phon., 2003) have been introduced to model the local slowing or lengthening of articulatory gestures in the vicinity of phrase boundaries. In this paper, pi-gestures are simulated within the TaDA task dynamics computational model and examined using functional data analysis (FDA) to evaluate articulatory lengthening in terms of underlying pi-gesture activation duration and strength from a realistic control model. A new derived variable of "deformation index" (area under FDA time-deformation functions) is shown to capture differences in π-gesture effect due to boundary strength.

Index Terms: pi-gesture, phrase boundaries, prosodic lengthening, functional data analysis

1. Introduction

1.1. Articulation & modeling of prosodic boundaries

Prosodic structure has been shown to affect both the temporal and spatial properties of the articulation of speech gestures. Speech shows a local slowing in the vicinity of a prosodic boundary (e.g., Byrd et al. 2000, Byrd et al. 2006), and there is evidence, though mixed, that gestures increase in magnitude near a prosodic boundary (e.g., Fougeron & Keating 1997; Byrd & Saltzman 1998, Fougeron 2001, Tabain 2003, Cho 2005, 2006), though these results can be mixed (Byrd et al. 2005, 2006)

Within the Articulatory Phonology (e.g., Browman and Goldstein 1992) model for representing the phonological structure of speech, π -gestures have been proposed to account for prosodic juncture effects (Byrd et al. 2000, Byrd & Saltzman 2003). Under this paradigm, phrase boundaries are modeled as prosodic gestures (π-gestures) with a temporal activation interval, similar to constriction gestures. The π-gestures act to locally slow down the clock that controls the temporal unfolding of articulatory gestures during the interval when they are active. The activation interval of π-gestures has been modeled using ramped functions, such that there is a stronger effect near the center of the gesture than at the edges, thereby capturing that articulatory effects have been observed to diminish as the distance of the constriction gesture from the boundary (roughly, phrase edge) increases. Modeling of π -gestures has shown to capture temporal and spatial effects of prosodic boundaries on speech (Byrd & Saltzman 2003). Crucially for our work here, differences in π -gesture activation duration and strength are hypothesized as possible mechanisms for capturing the juncture strength differences between varying prosodic boundaries. Longer and/or stronger pi-gestures will yield greater prosodic slowing, in accord with a stronger prosodic boundary. However, the impact of manipulating pi-gesture activation intervals and strength of activation is not well understood, nor are the interactions between these two gestural parameters. Evaluating these effects is complicated for a number of reasons.

First, in natural or even laboratory speech, it is difficult to overtly control the strength of a prosodic boundary that is produced. Secondly, past studies examining the effects of linguistic variables such as prosody on articulation have relied on kinematic landmarks to define speech intervals of interest, ignoring the continous time course, or time evolution, between those landmarks. The present study attempts to ameliorate these difficulties by using articulatory speech synthesis to overtly control juncture strength via a π-gesture. This allows for a proof-of-concept that spatiotemporal deformations of articulatory trajectories at junctures can insightfully reflect changes in boundary strength. Functional Data Analysis (FDA) (Ramsay & Silverman 2005) allows the analysis of entire, continuous kinematic trajectories (Lee et al. 2006), capturing non-linear warping in time and space.

1.2. Functional Data Analysis: An Overview

Functional data analysis (FDA) (Ramsay & Silverman 2005) preserves information on the detailed, continuous pattern of articulatory timing that unfolds during an interval. We (Lee, et al. 2006; Byrd et al. 2008) presented an FDA approach that allows the analysis of entire, continuous kinematic trajectories obtained in a movement tracking experiment examining the influence of a phrasal boundary on articulatory patterning. FDA time deformation functions reveal detailed patterns of delaying (i.e., slowing of internal clock-rate) of articulator movement in the presence of a phrase boundary as the speech stream approaches and recedes from the phrase edge. The gradual increase and decrease of clock-slowing around a phrase edge is a theoretically predicted pattern within the pi-gesture model (Byrd & Saltzman 2003), which would be more difficult to visualize and validate with a traditional interval-based approach. The present study extends this work to determine if FDA can be used to distinguish the effects of different strengths of prosodic boundaries on speech articulation and to connect these explicitly to controlled manipulation of pi-gesture activation strength and duration. We do this via integration of the FDA time deformation functions to quantify prosodic effects in articulation.

1.3. Current study goals

While Byrd & Saltzman (2003) demonstrated that the temporal effect of a π-gesture increases with its activation strength, the spatial and temporal effects of the interactions of π-gesture activation strength and duration were not explored in detail. Neither has any previous study attempted to use FDA to recover differences between prosodic boundaries of varying strength. This study has two specific goals. First, it seeks to determine the ability of FDA analysis to capture the boundary strength differences for continuous articulatory trajectories. In order to do this, we have chosen to leverage synthetic articulatory speech data. By using articulatory synthesis and explicitly controlling the duration and strength of the π-gestures used

to instantiate the prosodic boundary, we know *a priori* which differences should in principle be recoverable. Additionally, it crucially provides us with an unambiguous control signal, created with no π-gesture and therefore no local slowing; this is important for use in the FDA. The second goal of the study is to examine consequences of variation in π-gesture activation strength and duration. Specifically, this study will manipulate both activation strength and duration of the π-gesture, in a fully-crossed manner, to determine whether one or the other may have a larger effect on boundary-adjacent articulatory lengthening. For both of these goals, we will be pursuing a new FDA measure, the DEFORMATION INDEX, derived from the integration of FDA time-deformation functions. We anticipate that this measure will offer the field a new tool to quantitatively assess boundary strength in articulatory data.

2. Method

The current study uses the Task Dynamic Application (TaDA) developed at Haskins Laboratories to produce both acoustic and articulatory output (Saltzman et al. 2008; Nam et al. 2005). Within this computational model of speech production, articulatory constriction gestures are the basic compositional units of speech. These gestures are goal-directed actions with specified dynamical parameter values for stiffness (within a critically-damped mass-spring model), constriction degree, and constriction location. Each action or gesture acts on one (vocal) tract variable (such as Lip Aperture, Tongue Tip constriction degree, etc.), which in turn are made up of synergies of articulators (for example, Lip Aperture calls on the upper and lower lip and jaw articulators). The temporal patterning of these actions is modeled via intergestural coupling relations that relies on a constellation of planning oscillators associated with each gesture (Goldstein et al., 2006, 2008). These relations can be represented in a coupling graph, which both reflects the phonological structure of the utterance and determines the coordination of the gestures involved in producing that utterance. From the coupling graph, a gestural score is created with the activation times and durations of the various gestures. The model synthesizer then uses that score to create an articulatory pattern in time and its corresponding acoustic signal.

The current version of TaDA incorporates π-gestures into the gestural score. These gestures act to locally slow the temporal unfolding or pacing of constriction gesture activation (Byrd & Saltzman 2003). These prosodic gestures can be placed directly into the gestural score and can be manipulated in terms of their temporal location, activation duration, and activation strength.

2.1. Generation of test speech materials

Using TaDA, a series of four-syllable utterances was created with π-gestures of varying strength and duration. There were two segmental patterns used, differing in the presence or absence of a coda consonant at the prosodic boundary: [CV.CV#CV.CV] and [CV.CVC#CV.CV]. All vowels were [a], and two separate sets of utterances were created with different values for C, one with the bilabial stop [p] and one with the alveolar [t] (table 1). For each condition (2 coda types x 2 consonants), the π -gesture's midpoint was coordinated synchronously with the midpoint of the constriction gesture for the consonant after the prosodic boundary, ([CV.CV#**C**V.CV]). The duration of the constriction gesture for each consonant was 120 or 130 ms; for each vowel, 240 or 250 ms (the expected variation in these ranges is due to the dynamic generation of the gestural score from the coupling graphs).

Table. 1. *Phrases used in generation of synthetic speech*

	Labial [p]	Alveolar [t]
Pre-boundary coda	pa.pap#pa.pa	ta.tat#ta.ta
No pre-boundary coda	pa.pa#pa.pa	ta.ta#ta.ta
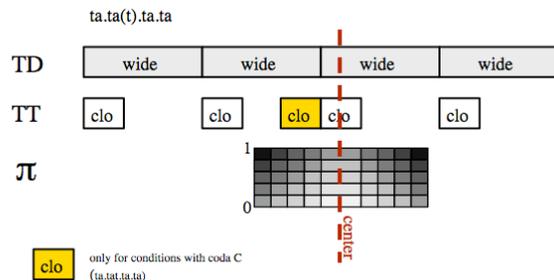
Figure 1: Schematic gestural score

The π-gesture activation strength and duration were manipulated as shown in Figure 1. The strength of the π-gesture ranged from .2 to 1 (where one is maximal activation in arbitrary units) in five steps of .2. The π -gesture activation duration also increased in five steps, with the first step equal to the duration of the synchronous closure gesture, and each subsequent step increasing in duration by 20ms on both sides of center (i.e., a 40ms total increase). A control utterance, with no π-gesture, was also generated; it was otherwise identical to the utterances with pi-gestures. All gestures were generated with cosine-ramped activations and deactivations, following Byrd & Saltzman (2003). This resulted in a total of 100 synthesized test utterances (5 activation strength steps x 5 activation duration steps x 2 utterance types x 2 consonants) and 4 control utterances (2 utterance types x 2 consonants).

2.2. Functional data analysis of the model trajectory

We used FDA time alignment to examine the TaDA-generated articulatory trajectory of the consonant articulation at and around the boundary. This means that we examined either the lip aperture (for [p]) or tongue tip constriction degree (for [t]) trajectory. We used the FDA landmark time registration method described in Lee et al. (2006) (see Figure 2a). We will be comparing prosodic effects shown in the articulatory trajectory in a test signal with the comparable control (or "reference") signal in which no boundary effects are present. After smoothing of the original curves (trajectories), a linear time normalization is applied to each individual signal by resampling so that each signal has 500 equally sampled data points (Figure 2b). This length normalization step removes any linear time lengthening effects and is required by the implementation of the FDA time alignment procedure. Twenty B-splines of the order 6 and λ value of 1E-12 are used in the regularized FDA smoothing method (Lee et al. 2006).

Next, for a landmark time registration of two curves (i.e., the reference and one test curve), twelve B-splines of the order 4 (i.e., piece-wise cubic-splines) and λ value of 1E-12 are used in the regularized FDA time alignment of the two curves. As landmarks, the four minima locations (see Figure 2c) are chosen and used as internal break points during the time alignment procedure.

After time alignment, a time deformation function Ftest (t) is computed as follows: Ftest(t) = htest(t) - (h)ref(t), which represents delay (Ftest(t) > 0) or advance (Ftest(t) < 0) of the internal clock time of a test signal with respect to the reference. As the endpoints for this analysis are anchored or 'pinned' at the edges of the interval of interest, timing effects at

the two end points of the interval are not discernable.
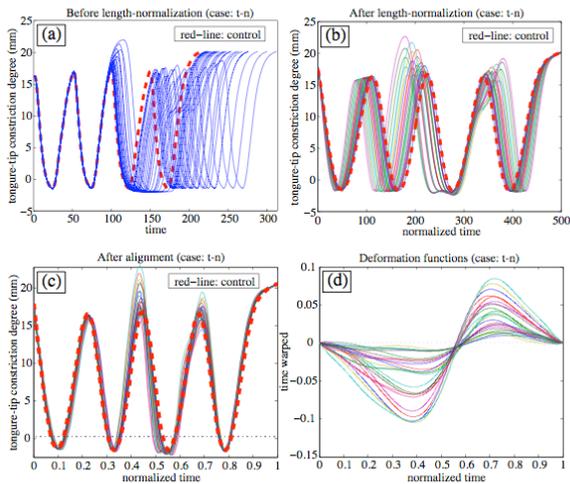
Figure 2: *For [tata#tata]. (a) Outputs of TaDA articulatory synthesis based on systematically varied boundar conditions. (b) Lineraly time normalized outputs by uniform sampling of 500 points. (c) Outputs after FDA time normlization procedur; all extrema positions are well aligned. (d) Time deformation functions capturing the differences in detailed time evolution pattern as a function of boundary conditions.*

Thus, the time deformation function reflects how the trajectories that were synthesized with a prosodic boundary (implemented with a π-gesture of some particular strength and duration of activation) are delayed or advanced relative to the control trajectory in which no boundary occurred. Because the π-gesture was synthesized to be synchronous with the center of the onset consonant gesture, we expect prosodically lengthened articulatory trajectories to be advanced before that synchronized point and delayed after it (relative to control).

Recall that our intent here is to pursue the suggestion in Lee et al. (2006) that the area under the time-deformation function could provide a valuable derived measure for capturing the effects of the prosodic juncture. Therefore, using a trapezoid rule, the Deformation Index (the area under the curve of each time-deformation function) is calculated as a measure of the strength of the π-gesture/prosodic boundary. Because of the length normalization (see Figure 2b), the time slowing effects are spread over the entire time region and, as such, the time deformation function changes its sign at the center of the π-gesture (i.e., from negative to positive, see Figure 2d). Therefore, in order to compute the area under the deformation curve as the measure of lengthening effect, we take the absolute value of the curve.

3. Results

Results of examining the Deformation Index—the area under the time deformation function—from the four conditions are shown in Figure 3. Two major patterns are clearly visible. First, the five π-gesture activation strengths have clearly distinct lengthening effects when compared at the same π-gesture activation duration. Indeed, for the most part the π-gesture strengths are distinct regardless of activation duration (Figure 4). One can also see in Figure 3 that the activation strength of a π-gesture has a much stronger influence on its ultimate articulatory effects than its activation duration. For example, at duration level 1, the varying strengths of the π-gestures result in lengthening differences of approximately 35 arbitrary units, whereas at strength level 1, the varying durations of the π-gestures differ in their effect only by 2 units. While the π-gesture's duration clearly does have an effect on the amount of

lengthening in the output, it is much smaller than the influence of its activation strength. However, the two parameters reinforce one another; the effects of activation duration are much more noticeable at high activation strengths (Figure 5).
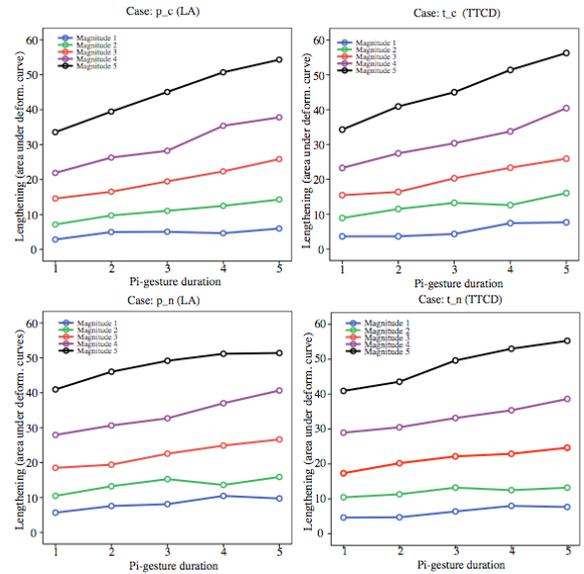
Figure 3: *Left—coda condition; right—no coda; top—[p]; bottom— [t]. The 5 levels of π-gesture activation are distinct at any given duration, and show little to no overlap across all durations. Also the magnitude of the π-gesture has a much larger effect than its duration.*
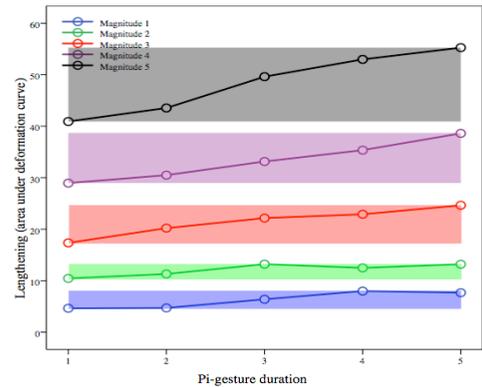
Figure 4: *([tata#tata]) The lengthening effect of π-gestures of different strengths are distinct at any given π-gesture duration (see shading); even at the durational extremes they remain largely separate.*
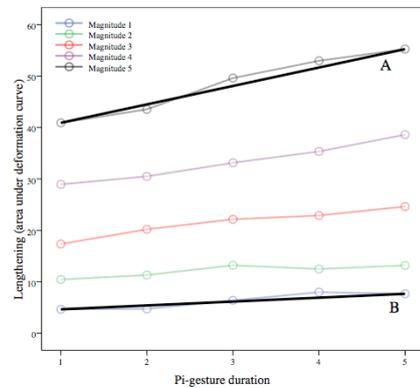
Figure 5: *([tata#tata]) The lengthening effect of π-gesture duration is greater at high π-gesture activation strengths. Compare the slope of line A (activation strength 1) with the slope of line B (activation strength 0.2).*

Differences due to the four segmental conditions (stop consonant and coda differences) were limited. There were no differences between conditions with [p] and [t]. This is predicted pattern since the π-gesture is active over all concurrent gestures without regard to those gestures' particular active articulators. There were, however, slight differences due to the presence or absence of a preboundary coda consonant. We can see that the Deformation Index (the non-linear slowing effect) is slightly higher in [CV#CV] than in [CVC#CV] (figure 3). This difference is present only at shorter π-gesture activation duration, and the two conditions generally have equal Deformation Indexes at activation durations 4 and 5.

In addition to its temporal lengthening effects, the π-gesture also affects the spatial magnitude of the articulatory gestures, in agreement with data from previous studies (Byrd & Saltzman 2003). We can see in Figure 2(c) that magnitude differences occur, though only in the immediate area around the π-gesture. The π-gesture creates both a wider constriction degree during the preceding and following vocalic interval and a more tightly constricted consonant closure posture.

4. Discussion

Three main conclusions can be drawn from the results above. First, we note that Functional Data Analysis, using a new derived variable Deformation Index (area under time-deformation functions), can recover differences in π-gesture lengthening effect due to activation strength and duration of a π-gesture in synthetic articulatory trajectories from a realistic control model. Future research will test whether the same FDA technique is able to distinguish different prosodic boundary types in non-synthetic articulatory data. We know from studies examining piecewise durations between kinematic landmarks that such differences exist, but we have not had a mechanism before this to quantify such changes for entire continuous trajectories that are varying in the boundary-adjacent interval. Until now, there has been no way to accurately and automatically measure boundary strength in speech production, nor to distinguish with one measure between boundaries of different phonological categories. These results indicate that FDA holds promise to do just that.[1]

Second, the magnitude effects seen in the modeling data in the current study show that it is possible that both prosodic temporal lengthening and spatial strengthening effects are the result of one π-gesture instantiating a prosodic boundary, as hypothesized in Byrd & Saltzman (2003). While the new measure of Deformation Index is solely a measure of temporal warping and thus does not capture these magnitude effects, the time standardization and alignment techniques of FDA offer a way to visualize magnitude differences.

Lastly, we find that the activation strength of the π-gesture

---

[1] It is important to recall, however, that the Deformation Index is sensitive to the presence of additional gestures in the area affected by the π-gesture. There were differences at low π-gesture durations between phrases with a preboundary coda constriction and those without one. This may be due to the homorganic nature of the involved gestures. When it is relatively short in duration, the π-gesture's domain includes the closure gesture for the postboundary onset but not the preboundary coda. When there is a preboundary coda, the onset has a much smaller effect on the ultimate articulatory trajectory since the articulator already starts close to or at its target; if there is little or no movement the π-gesture's effect will not be visible. However, when the π-gesture is long, it does include the closure gesture for the coda. Since the coda closure starts from the articulator position during the vowel, which is similar to the starting position for the onset closure in [CV#CV] sequences, the resulting effects of the π-gesture for the two conditions is similar.

has a larger effect than its duration on articulatory lengthening in these synthetic utterances. This finding is a step toward understanding how parametric variation in local clock-slowing can be connected with articulatory trajectories. In the future, such approaches will allow more accurately modeling of prosodic scope effects in natural articulatory experiments.

5. Acknowledgements

6. References

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

Byrd, D., Kaun, A., Narayanan, S., and Saltzman, E. (2000). "Phrasal signatures in articulation," in *Papers in Laboratory Phonology V. Acquisition and the Lexicon*, edited by M. B. Broe and J. B. Pierrehumbert (Cambridge University Press, Cambridge), pp. 70–87.

Byrd, D., Krivokapic, J., and Lee, S. (2006). "How far, how long: On the temporal scope of phrase boundary effects," *JASA* 120, 1589-1599.

Byrd, D., Lee, S., and Campos-Astorika, R. (2008). "Phrase boundary effects on the temporal kinematics of sequential tongue tip consonants," *JASA* 123, 4456-4465.

Byrd, D., Lee, S., Riggs, D., and Adams, J. (2005). "Interacting effects of syllable and phrase position on consonant articulation," *JASA* 118, 3860–3873.

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

Byrd, D. and Saltzman, E. (2003). "The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening," J. of Phon. 31, 149-180.

Cho, T. (2005). "Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in English," *JASA* 117, 3867-3878.

Cho, T. (2006). "Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English," in *Laboratory phonology 8: Varieties of phonological competence*, ed. L. Goldstein, D. H. Whalen, C. Best (Mouton de Gruyter, Berlin/New York), pp. 519-548.

Fougeron, C. (2001). "Articulatory properties of initial segments in several prosodic constituents in French," *J. of Phonetics* 29,109-135.

Fougeron, C. and Keating, P. A. (1997). "Articulatory strengthening at edges of prosodic domains," *JASA* 106, 3728-3740.

Goldstein, L., Byrd, D., and Saltzman, E. (2006). "The role of vocal tract gestural action in units understanding the evolution of phonology," in *Action to Language via the Mirror Neuron System*, Ed. by M. Arbib (Cambridge University Press, Cambridge), pp. 215-249.

Goldstein, L., Nam, H., Saltzman, E., and Chitoran, I. (2008). "Coupled oscillator planning model of speech timing and syllable structure," in *Proceedings of the 8th Phonetics Conference of China and the International Symposium on Phonetic Frontiers*.

Koenig, L.L., Lucero, J.C., and Perlman, E. (2008). "Speech production variability in fricatives of children and adults: Results of functional data analysis," *JASA* 124, 3158.

Lee, S., Byrd, D., and Krivokapic, J. (2006). "Functional data analysis of prosodic effects on articulatory timing," *JASA* 119, 1666-1671.

Nam, H., Goldstein, L., Saltzman, E., and Byrd, D. (2005). "TADA: An enhanced, portable Task Dynamics model in MATLAB (A)," *JASA* 115, 2430.

Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*, 2nd ed. (Springer-Verlag, New York).

Saltzman, E. L. and Munhall, K. G. (1989). "A dynamical approach to gestural patterning in speech production," *Ecol. Psych.* 1, 333-382.

Saltzman, E., Nam, H., Krivokapic, J., and Goldstein, L. (2008). "A task-dynamic toolkit for modeling the effects of prosodic structure on articulation," in *Proceedings of the Speech Prosody 2008 Conference,* edited by P. A. Barbosa, S. Madureira, and C. Reis (Associação Luso-Brasileira de Ciências da Fala), pp. 175-84.

Tabain, M. (2003). "Effects of prosodic boundary on /aC/ sequences: Articulatory results," *JASA* 113, 2834-2849.