Journal of Phonetics 41 (2013) 442–452

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Evaluation of prosodic juncture strength using functional data analysis

Benjamin Parrell [a,*], Sungbok Lee [a,b], Dani Byrd [a]

[a] University of Southern California, Department of Linguistics, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693, USA
[b] Viterbi School of Engineering, University of Southern California, CA 90089-1693, USA

ARTICLE INFO

Article history:
Received 27 September 2012
Received in revised form
6 August 2013
Accepted 16 August 2013
Available online 13 September 2013

ABSTRACT

Prosodic structure has large effects on the temporal realization of speech via the shaping of articulatory events. It is important for speech scientists to be able to systematically quantify these prosodic effects on articulation in a way that is capable both of differentiating between the degree of prosodic lengthening associated with varying linguistic contexts and that is generalizable across speakers. The current paper presents a novel method to automatically quantify boundary strength from articulatory speech data based on functional data analysis (FDA). In particular, a new derived variable—the Deformation Index—is proposed, which is the area under FDA time-deformation functions. First using synthetic speech produced with the TaDA task dynamics computational model, the Deformation Index is shown to be able to capture *a priori* known differences in boundary strengths instantiated in the π-gesture framework. Additionally, this method accurately distinguishes between types of boundaries in non-synthetic speech produced by four speakers.

1. Introduction

1.1. The articulation and 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 (acoustics: *e.g.*, Oller, 1973; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992; articulation: *e.g.*, Byrd, Kaun, Narayanan, & Saltzman, 2000; Byrd, Krivokapic, & Lee, 2006), and there is evidence, though mixed, that gestures increase in magnitude near these boundaries (*e.g.*, Beckman & Edwards, 1992; Beckman, Edwards, & Fletcher, 1992; Byrd & Saltzman, 1998; Byrd, Lee, Riggs, & Adams, 2005; Byrd et al., 2006; Cho, 2005, 2006; Cho and Keating, 2001; Fougeron, 2001; Fougeron & Keating, 1997; Tabain, 2003). While there is ample evidence of these spatial and temporal effects in speech, quantifying the overall effect of differing types or strengths of prosodic boundaries has proven difficult 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. Both inter- and intra-speaker variation exist for how the juncture between any two phrases will be realized. Secondly, past studies examining the effects of linguistic variables such as prosody on articulatory timing have relied on kinematic landmarks to define speech intervals of interest and compare their durations, ignoring the details of the continuous time course, or time evolution, between those landmarks. This is the standard practice in articulatory studies of prosody that examine timing; for example, articulatory closure or release intervals may be measured, or likewise time from gesture onset to peak velocity. While we have learned a great deal from such studies, they do not reveal the temporal detail of durational articulatory change along the continuous time axis. The present study attempts to ameliorate these difficulties. First, our utilization of the functional data analysis (FDA) time-registration technique allows the analysis of entire, continuous kinematic trajectories, capturing non-linear warping in time and space (Ramsay & Silverman, 2005). Second, we use articulatory speech synthesis to overtly control juncture strength via a π-gesture. This overt control allows for a proof-of-concept that articulatory trajectory deformations at junctures—which reflect both temporal and spatial changes—can insightfully reflect changes in juncture strength. We then confirm similar results using non-synthetic speech.

FDA time-registration preserves information on the detailed, continuous pattern of articulatory timing that unfolds during an interval. In previous research on speech, the FDA time registration method has been applied for speech in the analysis of lip movements (Ramsay, Munhall, Gracco, & Ostry, 1996), analysis of speech and voice signals (Lucero & Koenig, 2000; Lucero, Munhall, Gracco, & Ramsay, 1997), and the variability analysis of oral airflow data in children's speech (Koenig, Lucero, & Perlman, 2008). The main focus of these studies has been either to demonstrate the FDA time registration method and other related statistical methods or to estimate signal average and variability in an optimal way from repeated productions

* Corresponding author. Tel.: +1 213 740 2986; fax: +1 213 740 9306.
E-mail address: parrell@usc.edu (B. Parrell).

0095-4470/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2013.08.001

of the same utterance. FDA has been particularly successful in assessments of variability in speech kinematics, as it provides separation of spatial and temporal variability (Lucero, 2005). FDA has also been used to quantify the variability of different speech articulators during the production of intervocalic consonants (Lucero & Löfqvist, 2005). This study showed that FDA, through the ability to analyze spatial variability, can reveal the varying constraints on speech articulators due to production of different consonants.

Our approach, both in past work (Byrd, Lee, & Campos-Astorkiza, 2008; Lee, Byrd, & Krivokapic, 2006; Parrell, Lee, & Byrd, 2010a, 2010b) and in the current study, differs from other applications of FDA to speech data in that we use FDA not to quantify the variability within multiple repetitions of the same utterance (either spatially or temporally), but to capture differences between multiple *patterns* of production, here particularly between classes of prosodic boundaries. We have previously presented an FDA time-registration 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 (Lee et al., 2006). FDA time deformation functions, after alignment of test and reference (control) signals, 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 π-gesture model (Byrd & Saltzman, 2003), which would be more difficult to visualize and validate with a traditional interval-based approach. Byrd, Lee, and Campos-Astorkiza (2008) went on to use the FDA approach to show that interspeaker differences in boundary strength may be a source of important qualitative differences in the articulatory patterning of the boundary-adjacent gestures. The present study extends this work to determine if FDA time-registration can be used to distinguish the effects of *different strengths* of prosodic boundaries on speech articulation. We do this by pursuing a suggestion proffered in Lee et al. (2006), namely that integration of the FDA time deformation functions can be used to quantify prosodic effects in articulation. If so, the FDA time-registration method will be a potent tool for capturing prosodic boundary strength variation. We will be pursuing a new FDA-based 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. We test this new measure first on synthesized speech as a proof-of-concept, then on non-synthetic, natural speech.

2. Study 1: synthesized speech

2.1. Methods

2.1.1. Instantiating prosodic boundaries in synthesized speech

Our first test of the method uses synthetic articulatory speech data. By using articulatory synthesis and explicitly controlling the strength of the prosodic boundaries, we know *a priori* which differences should in principle be recoverable. Additionally, it crucially provides us with an unambiguous control signal, created with no boundary present; such a control signal serves as the baseline to which different classes of boundaries are compared in calculating the Deformation Index (for details, see Section 2.1.4). To instantiate the effects of prosodic boundaries, we employ π-gestures in the synthesis (Byrd & Saltzman, 2003) as discussed below. Importantly, we do not aim to test the validity of such a theoretical model *per se*; rather, we simply use this model as an established computational method that has been shown to model some major aspects of the spatial and temporal effects of prosodic boundaries in speech.

Within the Articulatory Phonology (Browman & Goldstein, 1992 and elsewhere) model for representing the phonological structure of speech, π-gestures have been proposed to account for prosodic juncture effects (Byrd & Saltzman, 2003; Byrd et al., 2000). Under this paradigm, phrase boundaries are modeled as prosodic gestures (π-gestures) with a temporal activation interval, similar to constriction gestures. But rather than activating a constriction variable, a π-gesture acts 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 distance of the constriction gesture from the boundary (roughly, phrase edge) increases. Modeling of π-gestures has been 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/or strength are hypothesized as possible mechanisms for capturing the juncture strength differences between varying prosodic boundaries. Longer and/or stronger π-gestures yield greater prosodic slowing in computational modeling of speech gestures, hypothesized to reflect stronger linguistic prosodic boundaries.

2.1.2. The task-dynamics model

For articulatory synthesis we use the Task Dynamic Application (TaDA) developed at Haskins Laboratories to produce articulatory and consequent acoustic output; details are available in Nam, Goldstein, Saltzman, and Byrd (2005) and Saltzman et al. (2008). This program implements the framework of Articulatory Phonology (Browman & Goldstein, 1992) and Task Dynamics (Saltzman & Munhall, 1989). Within this well-established 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, Tongue Dorsum 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 rely on a constellation of planning oscillators associated with each gesture (Browman & Goldstein, 2000; Goldstein, Byrd, & Saltzman, 2006; Goldstein, Nam, Saltzman, & Chitoran, 2009). 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. For any given input to the model, parameter information (*e.g.*, tract variable, constriction degree, constriction location, stiffness) for each gesture is accessed from a dictionary, and a coupling graph establishing the dynamical coordination between these gestures is constructed. 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 gestural score to create an articulatory pattern in time and finally its corresponding acoustic output signal.

The current version of TaDA incorporates π-gestures into the gestural score (Nam et al., 2005). These gestures act to locally slow the temporal unfolding or pacing of constriction gesture activation as described in Byrd and 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.

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

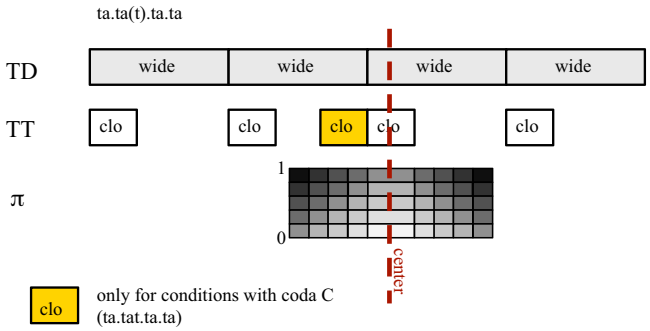

Fig. 1. Schematic gestural score showing articulatory and π-gestures. π-Gestures were centered on the closure gesture for the onset consonant of the third syllable and were varied in both duration (in five steps from 120 to 280 ms) and strength (in five steps from 0.2 to 1, where 1 is the maximum possible strength).

2.1.3. Generation of test speech materials

Using TaDA, a series of four-syllable utterances was created with π-gestures of varying strength and duration potentially located between the second and third syllables. 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 consonants—one with the bilabial stop [p] and one with the alveolar [t] (Table 1). For each condition (2 coda types × 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]). This specific alignment of the π-gesture in the synthesis was, to some extent, arbitrary; the alignment used in this study was chosen as a plausible first approximation to test our quantitative approach. While different alignments of the π-gesture with oral gestures in the neighborhood of the prosodic boundary would of course yield kinematic variation, we do not expect that the precise alignment of the π-gesture will affect the Deformation Index results as long it falls within the domain used in the FDA analysis. The duration of the constriction gesture for each consonant was 120 or 130 ms; for each vowel, 240 or 250 ms.[1]

π-Gesture activation strength and duration were manipulated as shown in Fig. 1. The strength of the π-gesture ranged from 0.2 to 1 (where one is maximal activation in arbitrary units) in five steps of 0.2. 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 20 ms on both sides of center (*i.e.*, a 40 ms total increase). A control utterance, with no π-gesture, was also generated; it was otherwise identical to the utterances with π-gestures. All gestures were generated with cosine-ramped activations and deactivations, following Byrd and Saltzman (2003). This resulted in a total of 100 synthesized test utterances (5 activation strength steps × 5 activation duration steps × 2 utterance types × 2 consonants) and 4 control utterances (2 utterance types × 2 consonants).

2.1.4. Functional data analysis of the model-generated trajectory data

We used FDA time alignment to examine the articulatory trajectory of the consonant articulation at and around the boundary. This always included the TaDA-generated articulatory trajectory produced for the activation of the onset [p] or [t] consonant in syllable three, and includes the coda [p] or [t] trajectory when a coda consonant [C#C] was present. This means that we examined either the Lip Aperture or tongue tip constriction degree trajectory. Here we outline the FDA-based time-alignment procedure (for further detail the reader may refer to Lee et al. (2006)). Throughout, the term "test" refers to an utterance with a boundary and the term "reference" refers to the control, no-boundary utterance. We will be comparing prosodic effects shown in the articulatory trajectory in a test signal with the comparable control signal in which no boundary effects are present.

First, the original curves (*i.e.* trajectories) are smoothed by applying the regularized FDA smoothing method to the curves (Fig. 2a). Twenty B-splines of the order 6 and λ value of 1E-12 are used in the regularized FDA smoothing method, as empirically determined in Lee et al. (2006). It is noted that the FDA parameter values were empirically tuned in a trial and error fashion and the minute λ value for trajectory smoothing was chosen as optimal as the articulatory trajectories are already smooth enough in the FDA data representation methodology. Then a linear time normalization is applied to each individual curve by an equal-distance resampling so that each curve has 500 equally sampled data points (Fig. 2b). This length normalization step removes any differences in overall duration, such as might arise from variation in speech rate. An additional practical purpose of duration normalization is to distribute the same linguistically salient articulatory events (*e.g.*, peak amplitudes or velocities) at somewhat similar locations in time in the test and control signals so that the gap between those events falls into the range that FDA non-linear time alignment algorithm (described below) can manage. Next, the time-normalized test signals are non-linearly warped to match the reference signal. This is accomplished by locally expanding or compressing the internal clock time of the reference signal to minimize the distance between—or align—the test and reference signals (signals after non-linear alignment shown in Fig. 2c). This results in a common time path (or "time-warping function") $h(t)$, which represents the local timing differences (slowing or advancing) between the internal time of the test signal with respect to the clock time of the reference signal. For the

[1] This slight variation in duration is to be expected from the TaDA model. Activation durations and the coordination between different gestures in the model are both defined in terms of the relative phase of planning oscillators associated with each gesture, not in terms of absolute time. Each time the model is run to generate a new utterance, these oscillators must settle into stable relationships. A small amount of noise in this dynamical process causes slightly different durations from trial to trial.

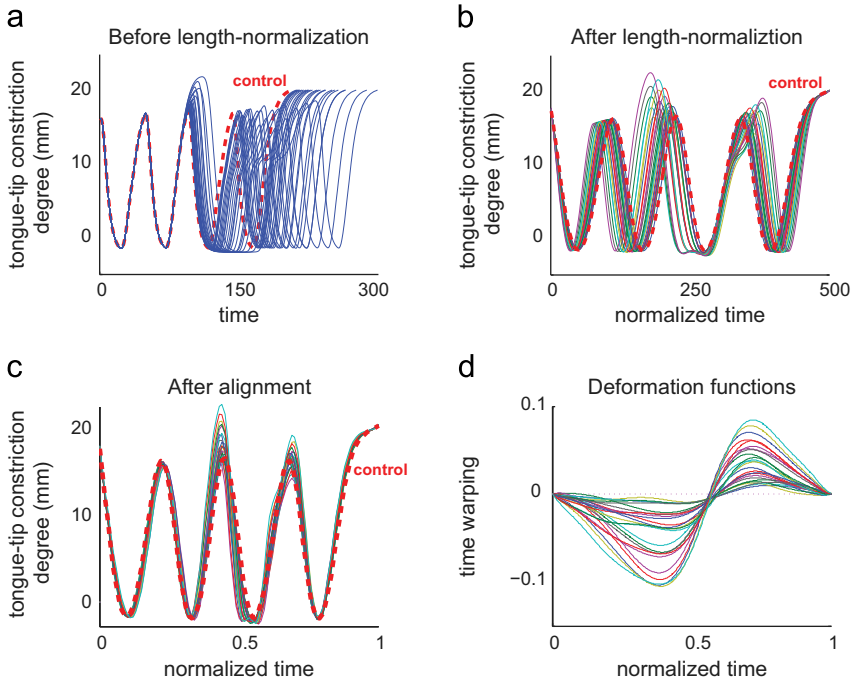

Fig. 2. (a) Outputs of TaDA articulatory synthesis based on systematically varied boundary conditions. In this and other plots, smaller constriction degree corresponds to less distance tongue tip and palate. (b) Linearly time normalized outputs by uniform sampling of 500 points. (c) Outputs after FDA time normalization procedure. It is clear that all extrema positions are well aligned. (d) Time deformation functions that visualize the differences in detailed time evolution pattern as a function of boundary condition. Figures show condition with coronal [t] and no coda consonant ([tata#tata]).

regularized FDA time alignment of the 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.

After time alignment, a time deformation function $F_{test}(t)$ is computed as follows:

$$F_{test}(t) = h_{test}(t) - h_{ref}(t)$$

where $h_{test}(t)$ represents the time-warping function generated from aligning the test to the reference signal and $h_{ref}(t)$ represents the clock time of the reference signal. $F_{test}(t)$ represents *delay* ($F_{test}(t)>0$) or *advance* ($F_{test}(t)<0$) of the internal clock time of a test signal with respect to the reference (Fig. 2d). This is equivalent to saying that when $F_{test}(t)$ is positive, the point on the curve at time t in the reference signal occurs relatively later in the test signal (it is *delayed* in time); conversely, when $F_{test}(t)$ is negative, the point on the curve at time t in the reference signal occurs relatively earlier in the test signal (it is *advanced* in time). 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 discernible.

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) due to the clock-slowing that the π-gesture instantiates. That is, we expect boundary adjacent lengthening to extend in both directions.

Recall that our intent here is to assess the suggestion in Lee et al. (2006) that the area under the time-deformation function could prove a valuable derived measure for capturing the effects of the prosodic juncture on the slowing of gestural timing. 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 Fig. 2b), the non-linear 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 Fig. 2d). Therefore, in order to compute the area under the deformation curve as the measure of the prosodic lengthening effect, we take the absolute value of the curve.

When using the automatic FDA alignment method, there are some cases that result in obvious mis-alignments of the position signals (Fig. 3a). In the example shown (taken from Study 2, Section 3), the automatic FDA alignment procedure aligns the second lip opening movement in the control signal (for the second syllable of "mama") to the first lip opening movement (for the first syllable of "mama") in the control signal.[2] This alignment problem can be fixed by preselecting relevant internal kinematic landmarks in each signal to ensure they are aligned properly by the FDA time-warping procedure (Lee et al., 2006). In this case, we chose the three LA maxima present in the signal, corresponding to the point of maximum aperture during the vowels as well as the local minimum preceding the last LA maximum. This alternate procedure provides accurate alignment of the two signals (Fig. 3b). In the following sections, we present two separate methods of dealing with this problem of misalignment. In the first case, we apply the FDA alignment method without internal landmarks. If this automatic method misaligns the kinematic signals, as it does on occasion, the internal-landmark method is used for the time-warping step for that token. In the second case, we use internal landmarks for all tokens. The results of the two methods for position will be compared to each other. In the current study, the four minima locations corresponding to consonant closure for /p/ or /t/ (see Fig. 2c)

[2] A reviewer points out that the misalignments found with the no-landmark method may be due to a larger number of maxima and minima in the test signals which fail to align properly compared to the reference signals (as can be seen on the left in Fig. 3a). Though the stimuli did not vary, slight variation in production may result in these additional peaks.

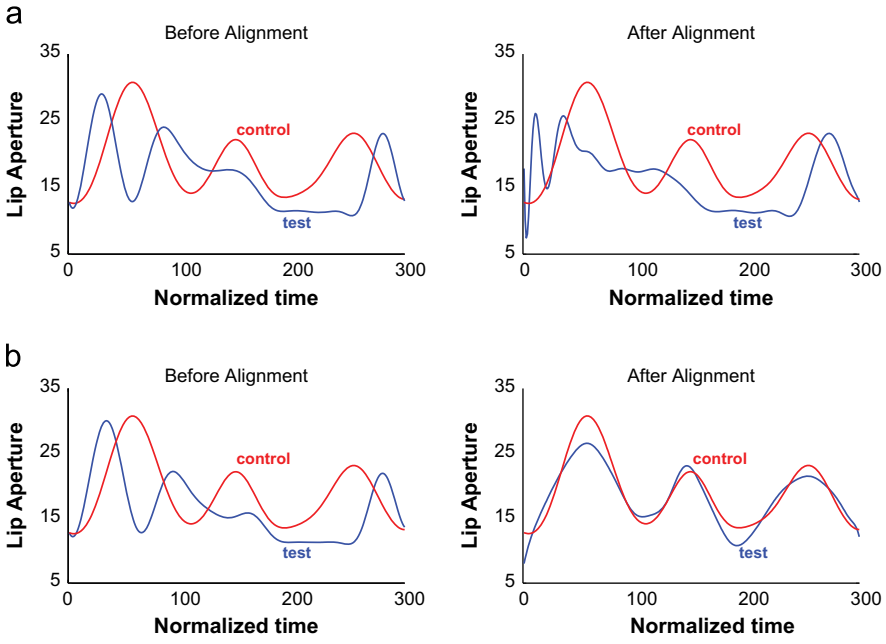

Fig. 3. When no internal landmarks are used in the FDA non-linear time warping procedure, there are some cases (a) where the results show obvious misalignments of relevant kinematic events (such as labial closure, which occurs near minima in the Lip Aperture signal). Selecting a small number of internal landmarks in both test and control signals ensures accurate alignment (b).

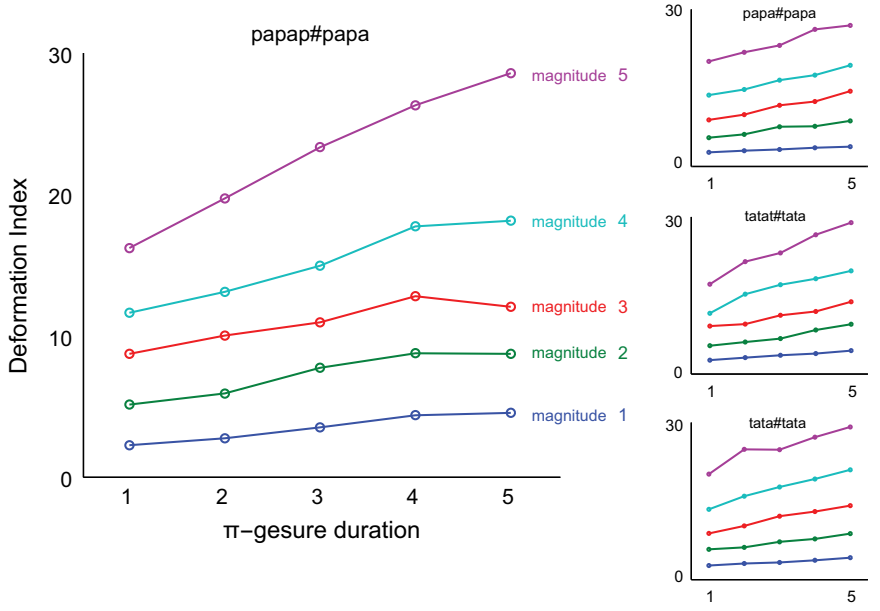

Fig. 4. Deformation Index results calculated from synthetic speech data using the automatic FDA alignment method. The five levels of π-gesture activation are clearly distinct at any given duration, and show little to no overlap across all durations. The magnitude of the π-gesture also has a much larger effect than its duration.

are chosen manually and used as internal break points during the time alignment procedure. For Study 1, there were no errors with the automatic alignment method.

2.2. *Results of Study 1*

Results of examining the Deformation Index—the area under the time deformation function—from the four conditions are shown in Fig. 4. As there was no difference in the results calculated using the two methods (Mann–Whitney $U=5661$, $p>0.05$), only the results from the automatic method are detailed below. 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, although there is some slight overlap between the strengths at very extreme durations (*e.g.*, a π-gesture of magnitude 4 at durations 4–5 and a π-gesture of magnitude 5 at duration 1). One can also see in Fig. 4 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 12 arbitrary units, whereas at strength level 1, the varying durations of the π-gestures differ in their effect only by 2 units.

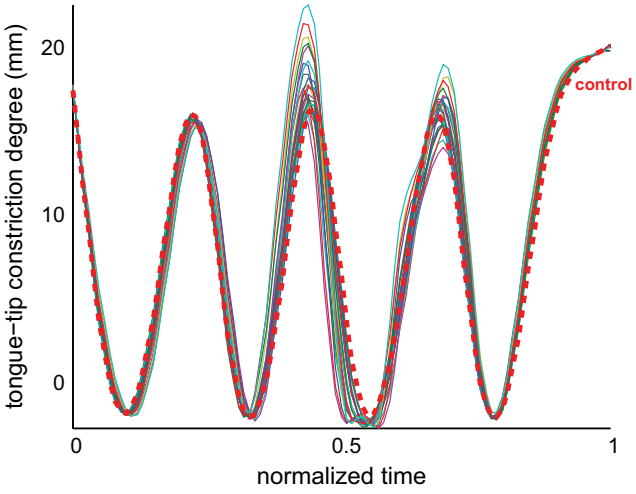

Fig. 5. When compared against the control signal with no π-gesture (dotted line) more extreme productions can be seen for vowels both preceeding and following the π-gesture and for the consonant at the π-gesture center. Figure shows condition with coronal [t] and no coda consonant ([tata#tata]), after FDA alignment.

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 do reinforce one another; the effects of activation duration are much more noticeable at high activation strengths. This makes sense: given a constant difference in strength between two given π-gestures, the overall difference in effect between the two should increase as they are active for longer and longer periods.

Differences due to the four segmental conditions (stop consonant and coda differences) were limited. There were no differences between conditions with [p] and those with [t]. This is the predicted pattern; since the π-gesture is active over all concurrent gestures, we do not expect differences in those gestures' particular active articulators to affect differences in overall slowing. There were, however, slight differences due to the presence or absence of a preboundary coda consonant gesture. We can see that the Deformation Index (the non-linear slowing effect) is slightly higher in [CV#CV] than in [CVC#CV] (Fig. 4). 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). Using the alignment technique described above, we can see in Fig. 5 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. Fig. 5 also shows that, with this alignment of the π-gesture to the constriction gestures, the magnitude of the preceding vowel is expanded to a greater degree than that of the following one. Though the π-gesture is active for the same duration on either side of the consonant gesture at the level of gestural control, this does not mean that it is active for the same duration on either side of consonant closure at the level of articulator motion. Because articulator closure is attained later than the onset of the control gesture, the π-gesture is effectively active for longer in the vowel preceding closure than following release, at the level of articulator movement. This may explain the vowel difference seen in Fig. 5. Particular spatial effects will be driven by precise temporal alignment of the π-gesture with the constriction gestures and in future work this may contribute to an understanding of how π-gestures are implemented in speech.

2.3. Discussion

Three main conclusions can be drawn from the results above. First, the FDA approach, using a new derived variable Deformation Index (area under time-deformation functions), can recover differences in prosodic lengthening due to activation strength and duration of a π-gesture in synthetic articulatory trajectories from a realistic control model. It is important to note, however, that the Deformation Index is sensitive to differences in the presence or absence of a preboundary coda constriction (though the particular vocal tract variable does not seem to matter in this case). This difference 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 post-boundary onset but not the pre-boundary coda. When there is a pre-boundary coda, the constriction gesture for the post-boundary onset—while active for the same duration—has a much smaller effect on the ultimate articulatory trajectory as the articulator starts close to or at its target; if there is little or no movement while the π-gesture is active, it will effectively lengthen the constriction plateau without affecting the dynamic portion of the movement into the closure. 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 are similar. While the absolute amount of lengthening is similar regardless of the constriction gestures affected by the π-gesture, it seems the FDA method may be more sensitive to changes in the portions of the kinematic trajectory that show changes, compared to those that are relatively more stable. This result indicates that the Deformation Index may be most useful when analyzing data with controlled segmental context, though it may still give reasonable results when small differences are present.

Second, we find that the activation strength of the π-gesture has a larger effect than its duration on articulatory lengthening in these synthetic utterances. Though not the primary purpose of the current study, this finding is a step toward understanding how parametric variation in local clock-slowing—*i.e.* local variation in the pacing of articulatory trajectories—can be connected with how articulatory trajectories unfold. In the future, similar approaches will allow us to more accurately model prosodic scope effects found in real-world articulatory experiments (*e.g.*, Krivokapic, 2007).

Lastly, while the new measure of Deformation Index is a specialized measure of temporal warping, the time standardization and alignment techniques of FDA offer a way to easily visualize and compare the magnitude of speech events across repetitions by neatly separating temporal and spatial variation between categories of prosodic boundary, just as previous work has used FDA to separate spatial and temporal variability (*e.g.*

Table 2
Stimuli sentences for study two. The sequence [mama#mimi] is placed at a different type of prosodic boundaries in each sentence, indicated by the word in the left column.

	Sentence
No boundary	Poppa-Pikt and Momma-Mimi tapped Cody.
List	Poppa, Pikt, Momma, Mimi, and Bibi tapped Cody.
Vocative	Quick Momma, Mimi tapped Cody.
Utterance	Poppa picked Momma. Mimi tapped Cody.

Koenig et al., 2008; Lucero, 2005). While we do not pursue examining the spatial effects of prosodic boundaries further in this paper, these effects should be visible in the residual differences between test and control signals after FDA alignment, an avenue we are currently investigating.

3. Study 2: natural speech

The results of Study 1 show that the Deformation Index can detect differences in boundary strength in synthetic speech. Natural speech, however, is much more variable than the synthetic speech signals used above. Subjects differ in the types and number of prosodic boundaries they use, and there is variation within productions of each type of boundary, as well as between types. In order to show the usefulness of the Deformation Index, it is necessary to demonstrate that it can detect and quantify differences between classes of prosodic boundary in natural speech; this is the aim of Study 2.

3.1. Methods

3.1.1. Stimuli

In order to test the ability of the Deformation Index to quantify different strengths of prosodic boundaries, it was necessary to create a series of sentences with identical segmental sequences across boundaries of varying types. We chose to use a series of sentences examining lip closing and opening movements for the target sequence [mama#mimi] embedded in four separate carrier sentences with differing syntactic structure, as originally used in Byrd and Saltzman (1998). These sentences, shown in Table 2, were selected for a number of reasons. First, the use of only the consonant [m] in the target series allows for the use of the same articulator, as was done in Study 1. Second, the sentences were designed to elicit a number of different prosodic boundaries from multiple subjects. A large body of work in phonological theory has shown evidence for multiple levels of prosodic boundary (including Word, Phonological Phrase, Intonational Phrase, and Utterance, though the precise number and type of categories is still disputed (Shattuck-Hufnagel & Turk, 1996). Additionally, experimental work has shown that speakers differ in the number and type of prosodic boundaries they employ for the same set of stimuli (Byrd & Saltzman, 1998; Shattuck-Hufnagel & Turk, 1996)). Because of this possible inter-subject variability in the choice of prosodic boundaries used in the production of a given set of stimuli, the stimuli sentences for this study were designed not to elicit particular phonological prosodic categories but rather a range of prosodic boundary strengths from each subject. We will discuss the results in terms of these atheoretic categories, or conditions ("no-boundary," "list," "vocative," "utterance"), rather than in terms of phonological prosodic categories of any particular theory. Lastly, there was no explicit manipulation of sentence-level focus or stress (as in *e.g.* Cho, 2006); therefore a possible confound of the phrase-level stress/accent cannot be ruled out since subjects were free to implement this spontaneously in their readings. That said, no notable variation was apparent in listening to the sentences.

3.1.2. Subjects and procedure

Four young adult subjects (TA, TB, TC, TD) participated in the Study 2. Subjects were seated approximately one meter in front of a computer screen off of which they read the test sentences. Sentences were blocked in the same order for all subjects: 'no-boundary', 'list', 'vocative', and 'utterance'. Subjects read each sentence 10 times, for a total of 40 sentences per subject. This data was collected at the first part of a larger session that included a second experiment.

3.1.3. Data collection and analysis

Kinematic articulator data was collected using an electromagnetic articulograph (Carstens AG500), which allows for three-dimensional tracking of transducers adhered to the articulators. For Study 2, transducers were adhered to the upper and lower lips, and the tip of the right index finger (the last sensor was used for the other study in the same experimental session). Reference sensors were attached to the nose ridge and behind each ear. Articulatory data was collected at 200 Hz and acoustic data at 16 kHz. After collection, the articulatory data was smoothed with a 9th-order Butterworth low pass filter with a cut-off frequency of 15 Hz, rotated to match the subject's occlusal plane (measured using a bite plate with two fixed sensors at the start of the experimental session), and corrected for head movement using the reference sensors.

Lip motion was calculated via the derived measurement Lip Aperture (LA), calculated by taking the two-dimensional Euclidean distance in the midsagittal plane between the sensors placed on the upper and lower lips. This gives a comparable measure to the tract variable Lip Aperture used for analysis of the synthetic utterances with /p/ in Study 1.

We applied the same FDA method described in Section 2.1.4 to the LA signal.[3] Recall that the general procedure is as follows: first, linearly time-normalize a test and a reference signal using the same number of samples; second, align the resulting signals using FDA non-linear time warping; lastly, integrate the resulting time-deformation function (representing local advancing or local slowing of the system time of a test signal with respect to clock time of the control signal) to give the Deformation Index.

[3] We also tested the FDA method using the velocity signal. However, nearly half (37/80) of the velocity signals showed misalignment using the automatic method. Because of both that issue and the fact that the velocity signal is derived directly and entirely from the position signal, we chose not to present that full analysis here, but the overall pattern of results was similar; namely, all subjects showed a significant effect of the boundary condition, with only two pairwise comparisons differing between position and velocity.

A few specific differences from study one exist with regard to procedure in study two. First, all signals were restricted to the time between maximum closure for the first and last [m] in [mamə#mimi] for LA (Fig. 6). In both cases, this was done in order to restrict the area of analysis to just that shared between all four target sentences. Second, time-normalization for all signals used 300, rather than 500, equal spaced samples. While absolutely smaller than the number used in Study 1, this number still gives more than twice the amount of samples in the original signals. Lastly, the reference signal for the FDA method was calculated as the average of the time-normalized no-boundary signals.

As in Study 1 (Section 2), we present two alternate ways of non-linearly aligning the data: automatically, and with the use of internal landmarks. For the second method, kinematic landmarks were chosen as the points of maximum opening between the first and last [m] constrictions in [mamə#mimi]. This was done in order to provide more landmarks (three versus only two possible for the [m] constrictions themselves), so that the alignment could be achieved with a better accuracy in the least-square sense. Automatic alignment failed on 5/80 tokens, which were then aligned using these same landmarks.

3.2. Results of Study 2

Following Byrd and Saltzman (1998), we do not expect to find that all subjects necessarily employ the same phonological prosodic boundary in producing the same test sentence. Moreover, it is not clear at this point what the arbitrary units of the Deformation Index may represent nor how consistent they may be between subjects. For those reasons, we analyze each subject separately. For each subject, separate non-parametric Kruskal–Wallis tests were conducted for each measurement (alignment without landmarks or with landmarks) with condition type (no-boundary, list, vocative, or utterance) as the factor. Results are presented in Table 3. All subjects showed a significant main effect of condition ($p<0.0001$) for all measurements. As can be seen in Fig. 7, the overall pattern of values was the same for both alignment techniques.

While the Deformation Index showed significant distinctions in prosodic boundary strength between different test conditions for all subjects, individuals differed in which conditions they distinguished, just as observed for subjects in Byrd and Saltzman (1998). Post-hoc tests using Mann–Whitney tests with Bonferroni correction were used to analyze which conditions differed for each subject. Results are presented in Table 4. It can be seen that all subjects distinguished between the no-boundary and utterance condition, and subjects differed in how the list condition and, especially, vocative conditions patterned. This variation indicates the production of the same syntactic context with either larger or smaller prosodic boundary strengths.

It should also be noted that the same spatial variability found in Study 1 is present in the current study. An example, from the vocative condition from subject TC is shown in Fig. 8. This result agrees with previous work that has found spatial expansion of speech movements near prosodic boundaries.

3.3. Discussion

The results from this study indicate that the Deformation Index is able to distinguish boundary strength in non-synthetic articulatory data, building on similar results from Study 1 using synthetic speech. Subject differed in how many different levels of prosodic junctures they used, and which syntactic test conditions were grouped together in temporal patterning. This was consistent with previous findings using the same stimuli (Byrd & Saltzman, 1998). That study found one speaker who split the stimuli no-boundary<list<vocative, utterance; one who split them no-boundary<list, vocative<utterance; and one who split them no boundary, vocative<list, utterance. While the details of the groupings found here are somewhat different, all subjects use more than one level of prosodic boundary and, moreover, the variability is similar.

While we have so far examined the Deformation Index as simply an ordinal variable without meaningful units, it is interesting that the ranges for the Deformation Index are roughly similar between three of the four subjects (see Fig. 7). TA, TB, and TC all showed a value of around 25 for the category with the most prosodic slowing, with an intermediate category (for TB) with a value around 15. Additionally, the subject who showed a much smaller maximum value (TD) also showed much less distinction between the different categories. It is, of course, impossible to say from the current data

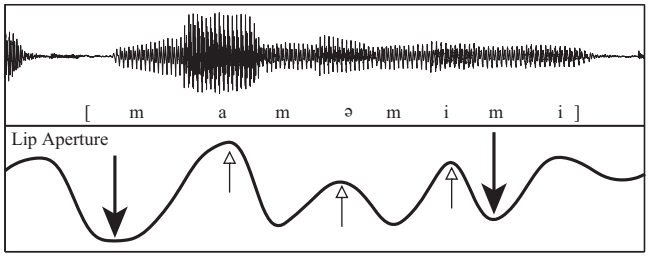

Fig. 6. Schematic showing position of boundaries (shown with open arrows) used in the FDA registration method. For Lip Aperture (LA), boundaries were placed at the maximum lip closure for the initial and final [m] in [mamə#mimi]. Closed arrows indicate the locations of the anchors used to define the portion of the utterance used in the calculation of the Deformation Index. The acoustic waveform and rough locations of the various segments are shown in the top panel for orientation. Greater LA signifies more distance between the upper and lower lips; hence, LA minima correspond to points of maximal consonantal constriction for /m/.

Table 3
Results of Kruskal–Wallis non-parametric tests by measurement (alignment without landmarks or with landmarks) with boundary (none, list, vocative, utterance) as the factor. As subjects showed different groupings of boundary strength, separate tests were conducted for each subject. All subjects show a significant result for both measurements.

Speaker	Alignment without landmarks	Alignment with landmarks
TA	$\chi^2(3)=30.3$, $p<0.0001$	$\chi^2(3)=30.6$, $p<0.0001$
TB	$\chi^2(3)=33.5$, $p<0.0001$	$\chi^2(3)=30.4$, $p<0.0001$
TC	$\chi^2(3)=25.2$, $p<0.0001$	$\chi^2(3)=26.6$, $p<0.0001$
TD	$\chi^2(3)=18.0$, $p<0.001$	$\chi^2(3)=18.9$, $p<0.001$

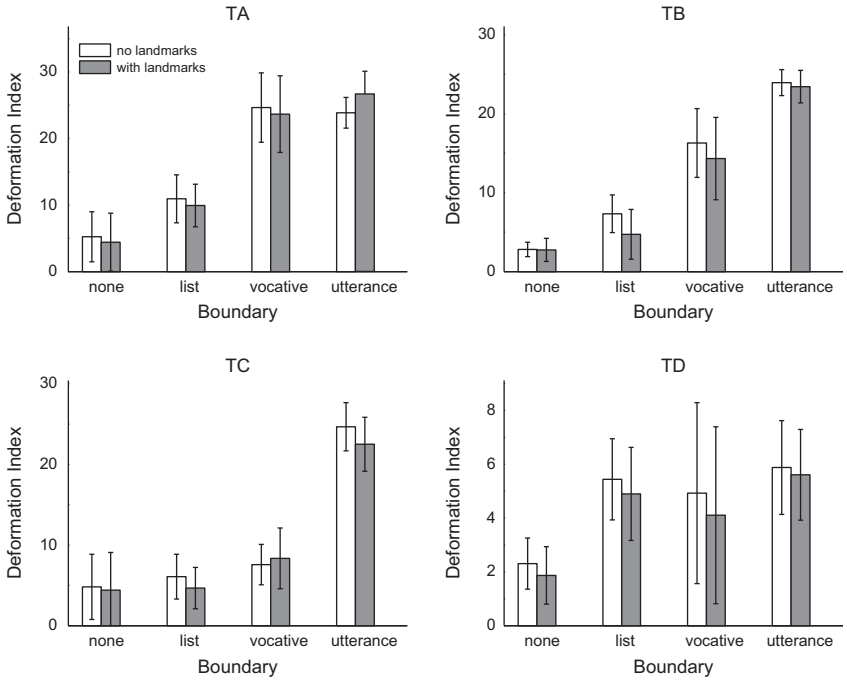

Fig. 7. Bar charts showing means and standard deviations of the Deformation Index by boundary condition. Subjects are each shown separately; clockwise from upper left: TA, TB, TC, TD. Note that all subjects show at least two distinct groups of boundary conditions, and that differences between the alignment methods are generally much smaller than those between boundary conditions.

Table 4
Results of post-hoc Mann–Whitney *U* tests with Bonferroni correction between boundary conditions for each subject. While subjects differ in the number of boundary groups, and which boundaries are grouped together, all show a difference at least between the no boundary and utterance conditions. Note that for speaker TD, the vocative condition did not differ significantly from any of the other conditions.

Speaker	Alignment without landmarks	Alignment with landmarks
TA	none, list < vocative, utterance	none, list < vocative, utterance
TB	none < list < vocative < utterance	none, list < vocative < utterance
TC	none, list, vocative < utterance	none, list, vocative < utterance
TD	none < list, utterance	none < list, utterance

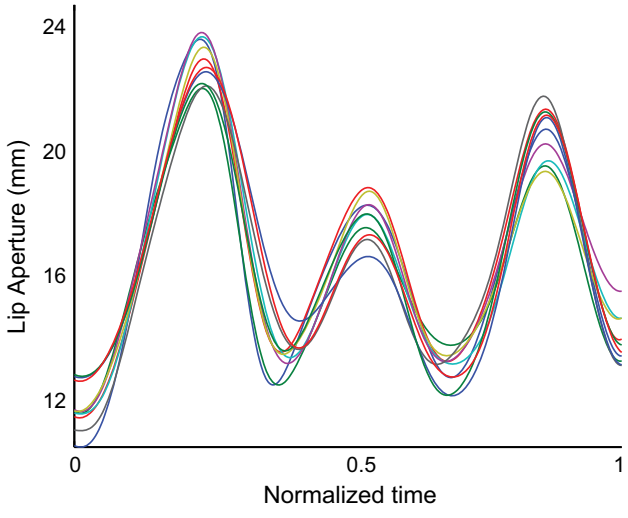

Fig. 8. Spatial variability remains after non-linear FDA landmark alignment in the LA signal. By removing the temporal effects of the prosodic boundary, the FDA alignment allows for easy visualization of spatial variability. Data shown is from the vocative boundary condition for subject TC.

whether this seeming correspondence indicates that the Deformation Index could be used to compare directly *between* subjects, but it points toward that possibility.

In both studies presented here, some spatial variability exists in the articulatory trajectories after the non-linear FDA time alignment. While the Deformation Index captures the temporal variability in the signal, this separation of spatial and temporal variation is an interesting and useful consequence of

the FDA procedure (cf. Lucero, 2005). Just as the FDA deformation function allows visualization of the waxing and waning of temporal effects near a prosodic boundary, comparison of residual differences between test and control signals after FDA time warping may, in future work, allow for analysis and modeling of spatial prosodic effects as speech approaches and recedes from a phrase boundary.

4. Conclusion

Articulatory studies examining piecewise durations between kinematic landmarks have shown that differences exist in prosodic boundary strength as a function of linguistic and communicative context. We have not, however, previously had a mechanism to quantify such changes for entire continuous trajectories that are varying in shape throughout the prosodically affected interval. Until now, there has been no way to accurately and automatically measure boundary strength in articulation, nor to distinguish with one quantitative measure between boundaries of different phonological categories. We have shown here that the FDA-based measure Deformation Index provides that capability. The Deformation Index captures prosodically-conditioned slowing across the interval of boundary-adjacent articulatory trajectories, rather than relying on selection of only a few points in time. Additionally, the process used to generate the Deformation Index automatically accounts for differences in the overall duration of these trajectories, such as may be due to speech rate within or across speakers, which interval-based analyses do not account for. (A similar global approach has been used to examine the degree of asynchrony in speech produced in time with another speaker or a recording (Cummins, 2009).) The results from our Study 1 indicate that the Deformation Index is useful in assessing the strength of a boundary despite variation in segmental and syllabic structure.

In the future, we will explore whether the Deformation Index can be used to directly compare the strength of prosodic boundaries between subjects, as the results from Study 2 indicate may be possible. If that is indeed the case, it may be possible to relate the Deformation Index values to phonological boundary classes such as Utterance, Intonational Phrase, Intermediate Phrase, and Word boundaries, which have so far proven elusive to classify quantitatively between speakers.

The Deformation Index is a single measure that captures the overall amount of lengthening in the vicinity of a prosodic boundary. In future work, further information can be mined from the deformation function created with this implementation of FDA time warping. For example, differences in the shape of the deformation function could reflect a differential timecourse of slowing or advancing of articulatory gestures. Though the shapes of the deformation function in the current study are very similar (*e.g.* Fig. 2c), it is possible that certain linguistic contexts could yield such shape differences; for example, deformation function shape differences could be correlated with linguistically significant information such as turn-taking or interruptions. Such differences could be quantified by computing instantaneous accumulated area values along the time axis. Thus, even if cumulative strength of lengthening (the Deformation Index) is not distinct, the evolution in time of the FDA deformation function could capture linguistically significant differences. Further research will address possible variation in the shape of the deformation function and methods to quantify such differences.

In sum, the current study demonstrates a means of quantifying boundary strength within and, perhaps, across speakers. This initial presentation of the Deformation Index has allowed for a new approach to quantifying local prosodically-conditioned temporal changes of entire continuous articulatory trajectories, abstracting away from global speech rate differences. And it has opened the door in future work for more nuanced quantification of the detailed time evolution of local slowing in prosodically important intervals.

Acknowledgments

The authors gratefully acknowledge the support of NIH Grant DC03172 and Louis Goldstein.

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 linguistic structure* (pp. 359–375). Tokyo: Ohmsha, Ltd.

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

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

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

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* (pp. 70–87). Cambridge: Cambridge University Press.

Byrd, D., Krivokapic, J., & Lee, S. (2006). How far, how long: On the temporal scope of phrase boundary effects. *Journal of the Acoustical Society of America*, *120*, 1589–1599.

Byrd, D., Lee, S., & Campos-Astorkiza, R. (2008). Phrase boundary effects on the temporal kinematics of sequential tongue tip consonants. *Journal of the Acoustical Society of America*, *123*, 4456–4465.

Byrd, D., Lee, S., Riggs, D., & Adams, J. (2005). Interacting effects of syllable and phrase position on consonant articulation. *Journal of the Acoustical Society of America*, *118*, 3860–3873.

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.

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in English. *Journal of the Acoustical Society of America*, *117*, 3867–3878.

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

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

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, *37*, 16–28.

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

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *106*, 3728–3740.

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action in units understanding the evolution of phonology. In: M. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 215–249). Cambridge: Cambridge University Press.

Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In: G. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in phonetics and speech science* (pp. 239–250). Beijing: The Commercial Press.

Koenig, L. L., Lucero, J. C., & Perlman, E. (2008). Speech production variability in fricatives of children and adults: Results of functional data analysis. *Journal of the Acoustical Society of America*, *124*, 3158.

Krivokapic, J. (2007). *The planning, production, and perception of prosodic structure*. University of Southern California ([Ph.D. dissertation]).

Lee, S., Byrd, D., & Krivokapic, J. (2006). Functional data analysis of prosodic effects on articulatory timing. *Journal of the Acoustical Society of America*, *119*, 1666–1671.

Lucero, J. (2005). Comparison of measures of variability of speech movement trajectories using synthetic records. *Journal of Speech, Language, and Hearing Research*, *48*, 336–344.

Lucero, J., & Koenig, L. (2000). Time normalization of voice signals using functional data analysis. *Journal of the Acoustical Society of America*, *108*, 1408–1420.

Lucero, J., & Löfqvist, A. (2005). Measures of articulatory variability in VCV sequences. *Acoustic Research Letters Online*, *6*, 80–84.
Lucero, J., Munhall, G. K., Gracco, V., & Ramsay, O. J. (1997). On the registration of time and the patterning of speech movements. *Journal of Speech, Language and Hearing Research*, *40*, 1111–1117.
Nam, H., Goldstein, L., Saltzman, E., & Byrd, D. (2005). TADA: An enhanced, portable task dynamics model in MATLAB (A). *Journal of the Acoustical Society of America*, *115*, 2430.
Oller, K. D. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, *54*, 1235–1247.
Parrell, B., Lee, S., & Byrd, D. (2010a). Quantifying prosodic boundary strength using functional data analysis of articulatory movement. *Journal of the Acoustical Society of America 128*, 2289.
Parrell, B., Lee, S., & Byrd, D. (2010b). Evaluation of juncture strength using articulatory synthesis of prosodic gestures and functional data analysis. In *Proceedings of speech prosody V*. Chicago, Illinois.
Ramsay, O. J., Munhall, G. K., Gracco, V., & Ostry, D. (1996). Functional data analyses of lip motion. *Journal of the Acoustical Society of America*, *99*, 3718–3727.
Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). New York: Springer-Verlag.
Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*, 333–382.
Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In P. A. Barbosa, S. Madureira, & C. Reis (Eds.), *Proceedings of the speech prosody 2008 conference* (pp. 175–84). Associação Luso-Brasileira de Ciências da Fala.
Shattuck-Hufnagel, S., & Turk, E. A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*, 193–247.
Tabain, M. (2003). Effects of prosodic boundary on /aC/ sequences: Articulatory results. *Journal of the Acoustical Society of America*, *113*, 2834–2849.
Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, *91*, 1707–1717.