5 Prosody in Articulatory Phonology

Jelena Krivokapić

5.1 Introduction

The linguistic term *prosody* has been conceived of as the structure of the utterance encoding prominence and hierarchical structure at the lexical and phrasal levels (e.g., Beckman 1996; Ladd [1996] 2008, 2001; Shattuck-Hufnagel and Turk 1996; Jun 2005). *Prominence* indicates a prominent syllable in a word or a prominent word in a phrase. The hierarchical structure groups gestures or segments into larger prosodic units, ranging from the smallest (e.g., the syllable in English) to the largest (e.g., the intonation phrase in English). This chapter examines phrase-level prosody as it is understood in the framework of Articulatory Phonology (Browman and Goldstein 1992, 1995; Goldstein, Byrd, and Saltzman 2006) and implemented in the Task Dynamics computational model (Saltzman and Munhall 1989; Saltzman 1995; Nam and Saltzman 2003; Saltzman et al. 2008). This is a relatively recent model of prosodic structure, with its first developments in Byrd et al. (2000). I start with a brief introduction to Articulatory Phonology.[1]

Articulatory Phonology posits that phonological representations and speech production events are isomorphic. The motivation for this approach is to overcome the division that is traditionally made between cognitive and physical aspects of speech (Fowler 1980; Browman and Goldstein 1995) by viewing these two aspects of speech not just as compatible, but as the macroscopic and microscopic aspect of the same representation, where microscopic properties refer to the physical characteristics (articulatory, acoustic) of the macroscopic cognitive units (the combinatorial units of speech). Importantly in this characterization, the microscopic and macroscopic properties of a system are expected to interact with and inform each other (Browman and Goldstein 1990b). The basic unit of speech is a gesture, a linguistically relevant constriction of the vocal tract, which is both an abstract, cognitive unit and at the same time allows for deriving the physical properties of speech. In other words, gestures are simultaneously cognitive units and "units of action" (Browman and Goldstein 1989).

Gestures are modeled as dynamical systems (specifically, constriction gestures are modeled as critically damped mass-spring systems).[2] The articulatory trajectories are lawfully derived by the dynamical systems of the cognitive units (gestures). Another fundamental characteristic of the model is that cognitive processes of speech are understood as dynamical:

> Thus, a dynamical perspective on coordinated movement not only reduces the conceptual distance between cognition on the one hand and mere bodily movement on the other, it forces reconceptualization of the nature of the inner cognitive processes themselves in dynamical terms. It thus turns out that cognition is not best thought of as

> something fundamentally distinct from movements of the body; rather, bodily coordination (and thereby interaction with the world) is really part of cognition itself. (Port and van Gelder 1995, 159)

Each gesture is specified by a set of spatial and temporal parameters (target, stiffness, and damping, where the target corresponds to the linguistic task) and is implemented by the coordinated movements of the articulators associated with each gesture. The *gestural score*, which represents the spatiotemporal properties of gestures in an utterance and their relative timing, drives the movements of the articulators.[3]

It needs to be pointed out that gestures are defined in terms of abstract linguistic tasks (e.g., a lip aperture constriction for [m] rather than the specific movements of lips) and realized through coordinative structures—functional groupings of specific articulators that perform the task (e.g., upper and lower lip and jaw for bilabial constrictions). As cognitive units, gestures are also contrastive, and the contrast can be (i) that gestures can be present or absent, (ii) in terms of their spatiotemporal properties (e.g., gestures using the same articulator can differ in constriction location), and (iii) determined through the temporal organization with other gestures.

Gestures are thus discrete units, specifying a task, but they result in continuous gesture trajectories, which in turn result in continuous movement of the vocal tract articulators. Because gestures are specified for both phonological and phonetic information, there is no need for a component mediating between phonology and phonetics.

Each gesture has associated with it a planning oscillator, or clock, that triggers the activation of that gesture (see Saltzman and Byrd 2000; Nam and Saltzman 2003; Nam, Goldstein, and Saltzman 2009). Thus the timing between two gestures is controlled by the coupling of their planning oscillators (note that the coupling specification is part of the lexical entry).

The model makes use of two prevalent modes of coupling in skilled action, namely in-phase (simultaneous) and anti-phase (sequential) coupling. Both modes are available to humans spontaneously and require no learning, with the in-phase mode being the preferred, more stable mode (Haken, Kelso, and Bunz 1985; Turvey 1990). These two modes have been shown in experimental and modeling work to give rise to syllable structure (Browman and Goldstein 2000; Nam and Saltzman 2003; Goldstein, Byrd, and Saltzman 2006; Nam 2007). When the planning oscillators are in-phase (as in the consonant [C] and vowel [V] gestures in a CV syllable), the associated gestures will be triggered simultaneously, and when the planning oscillators are anti-phase, as in the coda V and C gesture in a CVC syllable, the V and the coda consonant gesture are triggered sequentially.

In complex syllable onsets (but not in codas), for example, in CCV syllables, the planning oscillators for each consonant are in-phase coupled to the vowel and anti-phase coupled to each other (figure 5.1). This leads to competitive coupling, which results in the first C starting earlier and the second starting later (when compared to the CV timing) capturing the *c-center effect*, whereby the midpoint of the onset consonants has a stable temporal relationship to the vowel, regardless of the complexity of the onset sequence (Browman and Goldstein 1988, 2000; Goldstein et al. 2009). In syllable codas, consonants are timed anti-phase to each other, and the first consonant is timed anti-phase to the vowel. Because there is only one type of coupling involved, there is no competitive coupling and consequently no c-center effect.

In the Articulatory Phonology framework, intergestural timing thus emerges from the dynamics of the planning oscillator ensemble, and syllable structure is the result of temporal coordination among the gestures comprising the syllable. The two coupling

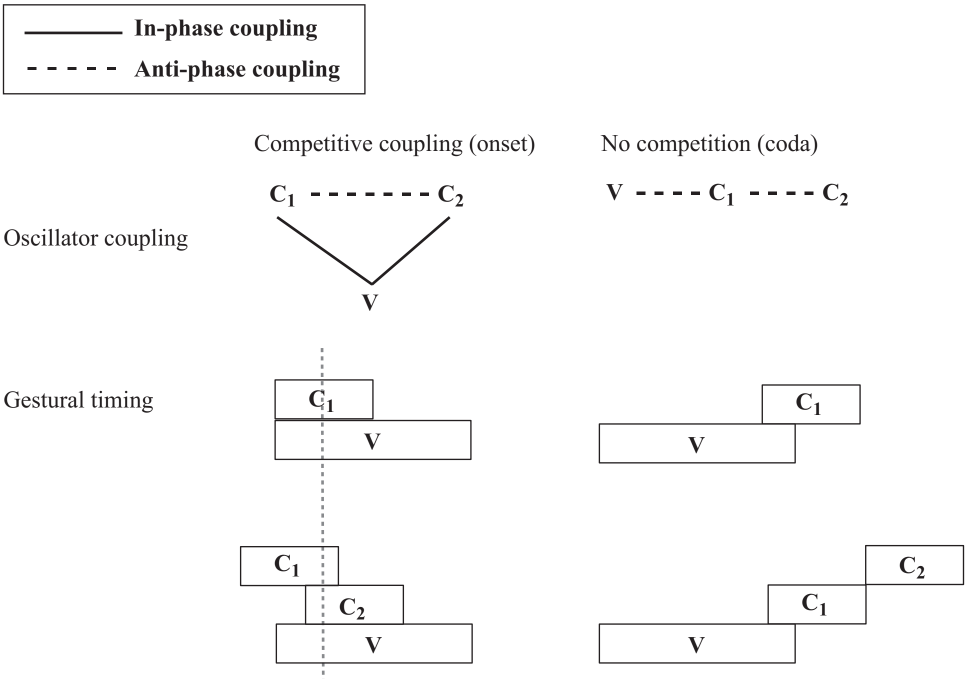

Figure 5.1
Schematic representation of oscillator coupling and the associated gestural timing relationships. Competitive coupling (which exists in syllable onsets) is shown in the left part of the figure, and the gestural timing can be, for example, for the syllable onsets in *pin* and *spin*. The dashed line in the gestural timing of the onsets represents the c-center, which stands in a stable relationship to the vowel regardless of the number of consonants in the onset. The left part of the figure shows coupling in syllable codas, and the examples of words having this gestural timing could be *tip* and *tips*. *Source*: Figure adapted from Marin and Pouplier (2010).

modes have been shown to provide a good account for empirically observed gestural timing in syllables (Browman and Goldstein 1988; Sproat and Fujimura 1993; Krakow 1999; Byrd et al. 2009), although there are cross-linguistic differences and exceptions, and the results are overall less consistent for codas than for onsets (see, e.g., Goldstein et al. 2009; Marin and Pouplier 2010; Hermes 2013; Brunner et al. 2014; and discussion in Pouplier 2011). This conceptualization of the syllable can also help explain a wide range of phenomena, such as typological properties of syllable structure and their acquisition (Goldstein, Byrd, and Saltzman 2006; Nam et al. 2009) and speech errors (Goldstein et al. 2007). (See the overview in Goldstein, Byrd, and Saltzman 2006; Gafos and Goldstein 2011; and Pouplier 2011.)

Finally, it should be mentioned that the specific predictions the model makes about articulation can be tested through the software implementation of the model, TADA (Task Dynamics Application). TADA, which also incorporates the Haskins articulatory synthesizer CASY (Configurable Articulatory Synthesis), takes a text string as the input and generates an intergestural coupling graph (which specifies the gestures of the utterance and the coupling between the planning oscillators of the gestures), the gestural score, the movements of the articulators, and a time-varying vocal tract shape and area function. TADA can be combined with the synthesizer HLsyn (Hanson and Stevens 2002) to synthesize the full acoustic signal. It can also be used without HLsyn, but it will not generate the full acoustic signal of the utterance, only its resonance pattern. For examples of use of TADA, see Gao 2008; Proctor 2009; and Mücke et al. 2012, who used it to verify their predictions for tones in Mandarin, for liquids, and for pitch accent alignment, respectively.[4]

Building on the outlined assumptions and foundations, prosodic structure is understood within Articulatory Phonology as arising through the coordination of a set of prosodic gestures. Like constriction gestures, these gestures present cognitive units and characterize the physical properties of speech. The prosodic gestures are temporal or clock-slowing gestures and tone gestures. We discuss these in sections 5.2 and 5.3, respectively, and prosodic gesture coordination in section 5.4. We turn to polysyllabic shortening and rhythm in section 5.5 and provide a summary and outlook in section 5.6.

5.2 Temporal Modulation Gestures and the π-Gesture Model

Byrd and Saltzman (2003; see also Byrd et al. 2000 for an earlier version of the model) extend the Articulatory Phonology approach to prosodic boundaries and propose that these boundaries are viewed as prosodic gestures (π-gestures). Note that the π-gesture is the cognitive representation of the prosodic boundary, not just its phonetic implementation, in the same spirit that constriction gestures are simultaneously units of representation and units of action. Like constriction gestures, the π-gesture extends in time. It is coordinated with other gestures, and it has a target.[5] Unlike constriction gestures, it does not have a constriction target; instead, it has a temporal target. The goal of the π-gesture is to locally slow the utterance clock that controls the time course of the constriction gestures. The effect of the π-gesture is to slow the time course of the constriction gestures that are active at the same time as the π-gesture. A slower course of activation leads to gestures being temporally longer.[6] Due to the slowdown of the clock, constriction gestures coactive with the π-gesture will also start temporally later compared to the same sequence of gestures without the π-gesture. As a consequence, gestures at boundaries become temporally longer and less overlapped. Both of these effects have been documented in the literature. (For less overlap, see McClean 1973; Hardcastle 1985; Jun 1993; Byrd et al. 2000; Bombien et al. 2010; Bombien, Mooshammer, and Hoole 2013; Byrd and Choi 2010. For lengthening in articulatory studies, see, e.g., Fougeron and Keating 1997; Byrd and Saltzman 1998; Byrd 2000; Keating et al. 2004; Cho 2006; Tabain 2003. For acoustic studies, see, e.g., Oller 1973; Wightman et al. 1992; Shattuck-Hufnagel and Turk 1998.)

The strength of activation of the π-gesture (indicated by the arrow in figure 5.2) determines how strong the boundary effects are. Hierarchically higher boundaries will have stronger activations, and consequently there will be more lengthening and less overlap than at lower boundaries, thus accounting for the empirical findings that boundary lengthening is cumulative. Whether the π-gesture also differs in scope at different boundaries is an empirical question, but the default assumption is that it does not (Byrd and Saltzman 2003; see also Byrd, Krivokapić, and Lee 2006). Note that the π-gesture extends both phrase-finally and phrase-initially; in other words, it accounts for both phrase-final and phrase-initial lengthening.

A potential consequence of the temporal effects are changes in the spatial properties of gestures.[7] Specifically, some of the phrase-initial articulatory effects that have

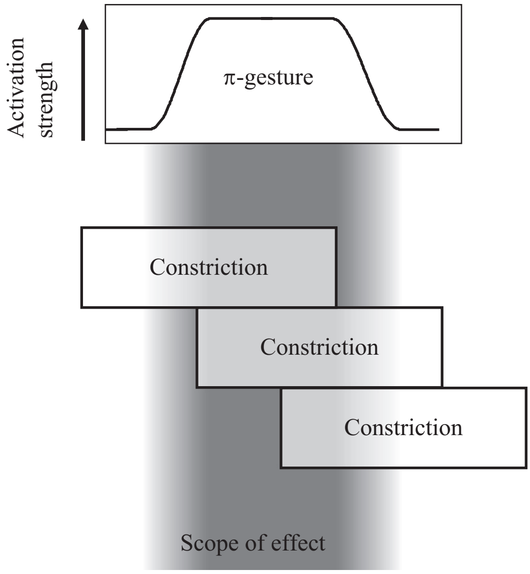

Figure 5.2
A schematic representation of the π-gesture model showing the overlap of the π-gesture and constriction gestures. The arrow indicates the strength of activation of the gesture, and the shading represents the scope of its effect, with darker shading indicating stronger activation. *Source*: Figure adapted from Byrd, Krivokapić, and Lee (2006).

been reported in the literature, including for example the increase in gestural amplitude in phrase-initial positions, could be a consequence of the temporal properties (for an overview of initial strengthening phenomena, see Fougeron and Keating 1997; Fougeron 2001; Byrd and Saltzman 2003). One possibility for this effect to arise is when two gestures share the same articulator and are both affected by the π-gesture. Due to the effect of the π-gesture, the two gestures are less overlapped, and there will be more time for the first gesture to reach its target (i.e., there will be detruncation), leading to larger amplitude of that gesture. In that case, the spatial changes can be understood as a consequence of the temporal changes, not as a main effect of prosodic structure. Note, however, that the π-gesture is not meant to capture all the effects of prosodic boundaries, only the temporal effects. For example, while voice onset time (VOT) increases at prosodic boundaries in Korean (Cho and Keating 2001), which could be understood as the result of a π-gesture–induced decrease in gestural overlap (as discussed in Byrd and Saltzman 2003), in Dutch the effect is opposite (VOT decreases at boundaries; Cho and McQueen 2005), indicating that processes other than lengthening take place at prosodic boundaries and interact with the lengthening processes.

We now turn to some of the predictions of the π-gesture model (as developed in Byrd and Saltzman 2003). Like constriction gestures, π-gestures extend over a certain period, and consequently the boundary effects are expected to *extend over a certain period*. The scope of the effect has not been fully established yet (we will return to this question), but the prediction of the model is that the effect will be *local*, because only gestures coactive with the π-gesture will be affected. A related prediction is that there will not be gestures that are "skipped" in the lengthening process. A further prediction of the model is that the boundary effect is strongest at the boundary and will *decrease* *with distance from it*, as the strength of activation of the π-gesture decreases (see figure 5.2). Thus gestures (and acoustic segments) closest to the boundary will lengthen most, and gestures farther away from the boundary will lengthen less, regardless of the specific gestures (and acoustic segments) involved. These predictions have been borne out in a number of articulatory (Byrd, Krivokapić, and Lee 2006; Krivokapić 2007; Katsika 2012 for Greek) and acoustic studies (Berkovits 1993a, 1993b for Hebrew, Cambier-Langeveld 1997 for Dutch).

An effect of the π-gesture is that gestures coactive with it are less overlapped. This is due to the assumption that gestures coactive with a π-gesture have a delayed activation in comparison to phrase-medial gestures. The prediction that arises is that the stronger the boundary, the less the gestural overlap will be. This prediction has been borne out so far (Byrd and Choi 2010; Bombien et al. 2010).

It should be pointed out that the effects of the boundary also depend on the shape of the π-gesture. That is, the strongest effects are predicted to be at the point where the activation of the π-gesture is strongest. In Byrd and Saltzman (2003) it was assumed that the strongest activation of the π-gesture is at the center, between two prosodic phrases. A different shape of the gesture would give different predictions as to where the temporal effects are the strongest. For example, if the shape of the π-gesture is skewed to the left, the effect would be strongest phrase-finally (see discussion and computational simulations in Byrd and Saltzman 2003). At this point, the shape of the π-gesture is an empirical question.

It is expected that any type of gesture coactive with the π-gesture will lengthen, but the possibility cannot be excluded that lengthening could interact with other factors (such as perceptibility or coarticulatory resistance). This prediction has not been extensively investigated, but research so far suggests that all gestures examined lengthen, even though there are some differences between them regarding the extent of lengthening. Thus, Fougeron (2001) finds, for French, postboundary lengthening for a whole range of consonants (/t, n, k, l, s/) and two different vowels; similarly, Cho and McQueen (2005) find final lengthening for /t, d, s, z/ for Dutch. Both velar and bilabial consonants have been found to lengthen as well in German (Bombien, Mooshammer, and Hoole 2013). On the other hand, Fougeron finds less lengthening on /s/ than for other consonants, and Cho and McQueen find that VOT in Dutch is shorter at higher prosodic boundaries than at lower boundaries (the opposite holds for Korean; e.g., see Cho and Keating 2001 and Jun 1993). These might be examples of how lengthening effects interact with other aspects of speech production.

A further prediction of the π-gesture model is that there is no qualitative difference between phrase-final and phrase-initial lengthening. In principle, there might be a difference in the *amount* of lengthening between phrase-final and phrase-initial position. As mentioned, the amount of lengthening depends on the strength of activation of the π-gesture, and therefore, where the effect is strongest (phrase-finally, phrase-initially, or distributed symmetrically) depends on the shape of the π-gesture and its coordination with other gestures (see section 5.4). The shape of the π-gesture is still an open empirical question, but importantly, there is no systematic, categorical, linguistically relevant difference between phrase-final and phrase-initial lengthening. Of course, listeners might be exploiting the lengthening arising from the π-gesture in different manners depending on whether the lengthening occurs at the beginning or at the end of a phrase. For example, phrase-initial lengthening might be more used for word recognition and phrase-final lengthening for syntactic disambiguation. Speakers also might use prosodic boundaries to plan the upcoming utterance, and the planning might be differently distributed across the boundary. But the phonetic and phonological characteristics of the boundary are not specified separately for phrase-initial and phrase-final lengthening.

Within this model, boundaries of different strength also do not differ in a linguistically relevant manner. In other words, there are no different types of π-gestures for different prosodic categories (e.g., for the intermediate or intonation phrase in Beckman and Pierrehumbert's 1986 model); the π-gestures differ only in strength, that is, in the amount of effect. A related issue is whether boundaries can be conceived of as being categorically different, reflecting a prosodic hierarchy with a small number of distinct categories, or whether boundaries are gradient, reflecting a more gradient view of the prosodic hierarchy. The dynamic specification of the π-gesture allows for a structurally gradient or categorical prosodic hierarchy (Byrd 2006). As discussed in Byrd (2006), within Articulatory Phonology, a gradient structure would arise if the π-gesture has a continuum of activation strength values, while a categorical structure could arise by specifying a small number of attractors for the activation strength of the π-gesture. Another way a categorical structure could arise is if the π-gesture has a continuum of activation strength values, but different prosodic categories are distinguished through tone gestures, such as phrase accents and boundary tones (see Byrd 2006; Krivokapić 2007). This is a fundamental question for prosodic theory, but it has not been addressed thoroughly and is in need of extensive research.

Phonetic evidence for a categorical view of prosodic boundaries would include the existence of specific phonetic properties marking each category. For example, the intermediate phrase (ip) and intonation phrase (IP) categories in Beckman and Pierrehumbert's (1986) model are distinguished by the presence of a boundary tone at the IP but not at the ip level (the increase in lengthening between ip and IP does not distinguish them, as this is a gradient, rather than a categorical difference; see also Frota 2000 on this point). Thus, in Beckman and Pierrehumbert's model, phonetic properties give rise to the hierarchy of word, ip, and IP. However, studies examining the production and perception of boundaries indicate that the boundary values form a continuum, not a small number of clusters that would be expected if the hierarchy consists of categorically distinct phrases (Ladd 1988; de Pijper and Sanderman 1994; Swerts 1997; Krivokapić 2007; Krivokapić and Byrd 2012; see also arguments in Ladd [1996] 2008; Frota 2000; and Wagner 2005). Importantly, evidence suggests that the continuum is driven by structural properties of the hierarchy (Ladd 1988; Frota 2000; Krivokapić 2007; Krivokapić and Byrd 2012; Ladd [1996] 2008; Wagner 2005). A second way to establish categorical differences between prosodic categories would be to show that different prosodic categories have different cognitive functions (similarly to phonemic contrasts, which distinguish lexical items). At this point, no evidence of such different functions exists. For example, prosodic phrases are known to disambiguate syntactically ambiguous utterances, but both ip and IP categories do that (e.g., Carlson, Clifton, and Frazier 2001; Clifton, Carlson, and Frazier 2002; Snedeker and Casserly 2010). A third way that a particular category could be justified is by the existence of phonological rules that apply only at that specific level of the prosodic hierarchy. However, claims about such rules have been made based on impressionistic evaluations, rather than on experimental evidence (as pointed out in Byrd 2006 and Wagner 2005). The few studies that have examined these rules have found that they apply in a gradient manner (see discussion in Byrd 2006, Wagner 2005, Krivokapić and Byrd 2012). Thus, questions about the structural properties of the prosodic hierarchy, specifically whether it is gradient or categorical, and if categorical, how many categories exist remain open. Importantly, however, the dynamic view of prosodic boundaries allows us to conceive of a different type of prosodic hierarchy, one that would possibly better fit the empirical evidence to date.

A question related to the structure of the prosodic hierarchy that has not been addressed directly in this model is the relationship between syntactic and prosodic structure. In general, there are two types of approaches to the prosody-syntax interface. In one, phonological and phonetic phenomena associated with the phrase level are mediated by prosodic structure (e.g., Nespor and Vogel [1986] 2007; Selkirk 1984, 2011), while in the other, they are defined directly over syntactic structure, rendering a separate prosodic structure unnecessary (Cooper and Paccia-Cooper 1980; Wagner 2005; see discussion of the approaches in Selkirk 2011). The relationship between syntax and prosody in the Articulatory Phonology model has largely been unexplored, and there is no theoretical reason for the model to prefer one or the other approach. (For a brief exploration of how the mapping between syntactic structure and prosodic structure could be approached within the model, see Krivokapić 2007.)

It has been suggested that the π-gesture model could be extended to account for lengthening in prominence (Byrd and Saltzman 2003; see also Bombien, Mooshammer, and Hoole 2013), and recent work has modeled lexical stress using a clock-slowing gesture, namely the μ-gesture (Saltzman et al. 2008).[8] Before the dynamics of the μ-gesture can be fully developed, a clearer understanding of the gestural timing and the kinematic properties of gestures in prominent syllables is needed. One difference between the μ-gesture and the π-gesture that can be expected lies in the coordination of the temporal modulation gesture to the constriction gestures. The evidence so far indicates that the strongest effect of the boundary is between two phrases and decreases with distance from that point, indicating that the strongest activation of the π-gesture is between two phrases. Prominence on the other hand seems to exert the strongest influence on the vowel of the prominent syllable (Cho and Keating 2009, Bombien et al. 2010; Bombien, Mooshammer, and Hoole 2013), indicating that the strongest activation of the temporal modulation gesture in prominence is at the vowel. This difference does not imply that the temporal modulation gestures differ qualitatively for boundaries and for prominence, just that their coordination with constriction gestures is likely to be different.

To summarize, the π-gesture model, which extends the gestural account of Articulatory Phonology to prosodic boundaries and to prominence, conceptualizes prosodic boundaries in a fundamentally different way from other models. Boundaries are understood to be gestures; they are inherently temporal. That is, their effect is to slow down gestures that are coactive with them; in this way increasing their duration (and in some cases their spatial extent) and reducing gestural overlap. The effect is thus local and continuous, and gradually decreases with the distance from the strongest activation point of the π-gesture. From this model, a set of specific predictions about the temporal properties of prosodic boundaries emerges, and these predictions have been tested in a large number of studies. These studies overall indicate that the gestural approach is a good way to systematically capture properties of boundaries. The model also raises questions about critical aspects of the prosodic hierarchy. A number of questions related to the properties of the π-gesture have been mentioned that still remain to be explored (see Byrd and Saltzman 2003 for additional questions).

5.3 Tone Gestures

Gao (2008; see also Gao 2009) further develops the gestural approach to prosodic structure and extends it to lexical tones. In her approach, tones are gestures that are, like other gestures, specified as dynamical systems (Saltzman 1986; Saltzman and Kelso 1987; Saltzman and Munhall 1989), extend in time, have targets, and are coordinated with other gestures. Gestures generally have specific articulators associated with them, but, as Gao points out, at this point it is not possible to model the articulators involved in the production of F0 due to a lack of understanding of all physiological processes involved in its production (see Hirose 2010 for an overview of the production of F0). However, the *abstract linguistic tasks* of tone gestures can be specified at the level of F0, and in that sense, F0 is an appropriate task variable, just like lip aperture, for example.

Tone gestures have as their goal linguistically relevant variations in fundamental frequency (Gao 2008; McGowan and Saltzman 1995). In modeling the four tones of Mandarin, Gao posits two tone gestures (with a high [H] and with a low [L] target). Tones 1 and 3 have a single gesture (tone 1 has an H target, tone 3 an L target), while tones 2 and 4 arise from a combination of two tonal targets. Tones, like constriction gestures, have planning oscillators associated with them, and the coordination of tones is determined by general coupling mechanisms. In tone 2, the L and H gesture are coordinated in-phase and the L gesture has a shorter activation interval than the H gesture, while in tone 4, the H and L gesture are coordinated anti-phase. For tone 2, the in-phase coordination results in an overlap of the L gesture with the H gesture while both gestures are active. This overlap results in the mid-level F0 in the first part of tone 2, followed by the rise to H. For tone 4, the anti-phase coordination results in the H tone followed by the L tone, that is, the falling contour of tone 4. Gao's computational implementation of these gestural representations in TADA (Nam et al. 2004) showed that these gestural scores yield F0 contours that match contours obtained in speech (Gao 2008, 44; Xu 1997, 67).

Starting from this view of tones, Gao examines tonal alignment in syllables with different syllable structures (CV and CVC) and with different onset consonants. Kinematic data analysis showed two different coordination patterns of the four tones, but they could both be explained by positing that the tone gestures behave like consonant gestures in their coordination patterns, that is, that they are coordinated to vowel and consonant gestures like onset consonants are. For example, for tones 1 and 3, the onset consonant and the vowel are in-phase coupled as are the tone and the vowel gestures, while the consonant and the tone are anti-phase coupled. The multiple coupling specifications give rise to competitive coupling—for example, for onset consonant gestures, and consequently to the c-center effect. Thus, in tones 1 and 3, the order of the gesture onsets is consonant, vowel, tone, indicating that the consonant gesture shifted leftward and the tone gesture rightward from the vowel onset, as in onset consonant clusters. As Gao points out, the specified coupling relations for tones 1 and 3 are like the ones observed for the onset consonants in English *spin* where the two supralaryngeal consonant gestures (tongue tip for /s/ and lip aperture for /p/) are in-phase coupled to the vowel and anti-phase coupled to each other (see figure 5.1). In Gao's analysis then, tones pattern with consonants in how they are coordinated, and they affect the gestural timing of consonant and vowel gestures. Note that the view that tones pattern with consonants fits well with theories of tonogenesis arguing that at least some tones have their historic origin in consonants (e.g., Hombert, Ohala, and Ewan 1979).

In addition to the observed in-phase coordination of vowel and tone gestures, anti-phase coordinations of the tone with the vowel (patterning like coda consonants) have also been suggested. Thus, Hsieh (2011) suggests that tone 3 consists of an L tone coordinated in-phase with the vowel and anti-phase with the onset consonant and an H tone gesture that is coordinated anti-phase with the vowel, accounting in this way for the behavior of tone 3 in different contexts (Gao 2008 discusses tone 3 in sentence-medial positions only, a context also captured by Hsieh 2011; for another account of tone 3, see Yi and Tilsen 2014).

Gao's (2008) approach to tones has been used to investigate the coordination of tones with constriction gestures in pitch accents (this is referred to as "tonal alignment" in the autosegmental-metrical approach to intonation) in Catalan and Standard German (Mücke et al. 2012), and Italian and Viennese German (Niemann et al. 2011). These studies find that differences and similarities between languages arise from the specification of the coordination patterns of tone and vowel gestures; in other words, they are the result of phonological structure. For example, in a kinematic, acoustic, and modeling study examining pitch accent coordination with vowel and consonant gestures in a variety of syllable structures, Mücke et al. (2012) show that in Viennese German, an L and an H tone are coordinated anti-phase to each other and in-phase to the V gesture of the accented syllable. This leads to competitive coupling, which in turn gives rise to the H gesture shifting rightward, and the L gesture shifting leftward (as for onset consonant clusters; see figure 5.1). In Catalan on the other hand, Mücke et al. show that L and H are also coupled anti-phase to each other, but only the H gesture is coupled in-phase with the vowel. The result is that there is no competitive coupling and that the H gesture starts simultaneously with the V gesture. Importantly, as argued in Mücke et al., these results could be described as the tone gesture being aligned to "the left edge of the onset consonant" (in Catalan) as opposed to the "middle of the onset consonant" in German (221). However, an examination of this question that leverages the gestural approach and the coupling oscillator model can account for the observed differences between languages in a precise manner. Similarly, Niemann et al. (2011) argue that German and Italian differ in the gestures comprising rising tones and in the coupling of gestures. Thus, the rising tone in Italian comprises one gesture (with an H target) that is in-phase coordinated with the vowel gesture, but in German, the rising tone is composed of two gestures (with an L and an H target) that are in-phase coordinated to the vowel gesture and anti-phase to each other. This difference in coupling in the two languages leads to different F0 realizations. Niemann et al. also find that pitch accent tone gestures do not affect the lexically specified CV coordination patterns.

Mücke et al. (2012) observe an important distinction between lexical tone gestures and pitch accent tone gestures: tone gestures in lexical tones, but not in pitch accents, affect the lexically specified CV coordination patterns (cf. the effects of tone on coordination reported in Gao 2008 for Mandarin lexical tone and Niemann et al. 2011 for German and Italian pitch accents). They hypothesize that this could be due to the coupling of tone gestures for pitch accents not being specified lexically, but postlexically, and therefore not being fully integrated into the coupling model, as opposed to the coupling of tone gestures in tone languages, where tones are part of the lexical representation and therefore integrated into the coupling model. Thus, whether tone gesture coupling is specified at the lexical or at the postlexical level could reflect and lead to typological differences.[9]

Tonal alignment has been the subject of extensive research (for an overview, see Ladd 2008). So far, consistent alignment between tonal landmarks (such as dips and peaks) and acoustic events has not been found. Lack of consistent alignment could be due to the relevant alignment landmarks being articulatory rather than acoustic in nature (as suggested in D'Imperio et al. 2007 and Gao 2008), but more importantly, the search for alignment points has often focused on surface rather than on structural properties (as discussed in Ladd 2006 and Gao 2008). To account for tonal alignment, Gao (2008) has taken as the starting point of her examination the assumption that the relevant domain of examination of timing is in terms of the structural properties of speech and that tones are gestures (see Ladd 2006 and Prieto and Torreira 2007, who also suggest that gestures in the sense of Articulatory Phonology might provide a better account of alignment).

Much more work remains to be done to understand how the Articulatory Phonology model can account for different languages and the various factors affecting tonal alignment, such as phonological vowel length (e.g., Ladd, Mennen, and Schepman 2000), syllable structure (e.g., Prieto and Torreira 2007), speech rate (e.g., Prieto and Torreira 2007; Hsieh 2011), and upcoming word boundary/prosodic environment (D'Imperio 2001; Prieto, van Santen, and Hirschberg 1995; Silverman and Pierrehumbert 1990). However, the results so far indicate that systematic phonological differences between languages and dialects, exhibited in different tonal composition and different coupling graphs, can account for surface variability in F0 alignment. Viewing tones as gestures within Articulatory Phonology also has the advantage of presenting tonal information in the same manner as constriction and prosodic information, thus allowing for an integrated investigation of prosodic phenomena. The approach has recently been extended to boundary tones in Greek (Katsika 2012; Katsika et al. 2014).

5.4 The Coordination of Prosodic Events

The coordination of gestures is part of the phonological structure in Articulatory Phonology and is specified by the coupled oscillator model. This model has reconceptualized the notion of the syllable and has been shown to account for timing of lexical tones and pitch accents. The π-gesture has not yet been integrated into the coupled oscillator model, but the coordination of the π-gesture to constriction gestures and other prosodic gestures is critical for understanding prosodic structure within Articulatory Phonology, and recent work has been exploring this question.

One of the predictions of the π-gesture model is that lengthening is continuous; that is, it does not skip any segments or gestures. A related question is how the onset of the boundary (specifically, the π-gesture) is determined. This question is of interest for any theory of prosodic structure, but Articulatory Phonology has a consistent way of framing it, namely, in terms of onset of the π-gesture and its coordination with other gestures.

These issues have been examined in a number of recent studies, and a specific focus has been placed on how preboundary prominence interacts with lengthening (e.g., Turk and Shattuck-Hufnagel 2007; Byrd and Riggs 2008; Katsika 2012, 2016; Katsika et al. 2014). In an acoustic study of English, Turk and Shattuck-Hufnagel (2007; see also Rusaw 2013, who replicated these findings) examined the effect of stress on the onset of final lengthening, testing words with final, penultimate, and antepenultimate lexical stress. They found that final lengthening started earlier in the word when the stressed syllable was farther away from the boundary, indicating that the boundary effects shift leftward toward the stressed syllable. However, unstressed syllables between the stressed and the final syllable were not always lengthened or were lengthened less than the stressed syllable. White (2002) also finds that the scope of lengthening extends from the end of the stressed syllable (in antepenultimate, penultimate, or ultimate position) to the boundary, but that the intervening unstressed syllables lengthen as well, contrary to Turk and Shattuck-Hufnagel (2007) and Rusaw (2013). Evidence for an interaction of boundary and prominence also comes from an articulatory magnetometry study by Byrd and Riggs (2008), who find that the onset of final lengthening can be attracted to the prominent syllable (but the results are speaker-specific). Byrd and Riggs suggest two possible accounts: either that the π-gesture shifts toward the prominent syllable, thus keeping the scope of lengthening constant, or that the π-gesture extends toward the prominent syllable, thus increasing the scope of the π-gesture. Crucially, both accounts specifically link the onset of the boundary to another prosodic event (see also Byrd 2006; Krivokapić 2007), capturing the interdependence of boundaries and prominence that many studies have shown.

For Greek, Katsika (2012, 2016) found in an articulatory magnetometer study that the onset of final lengthening occurred during the final syllable in words with final stress, but that it started earlier when the stressed syllable occurred earlier (consistent with White 2002; Turk and Shattuck-Hufnagel 2007; Byrd and Riggs 2008; and Rusaw 2013). Like White (2002), she found that lengthening is continuous (once it starts, it does not skip any gestures), as predicted by the π-gesture model. Katsika (2012) and Katsika et al. (2014) combined this examination of temporal properties of prosodic boundaries with the examination of the onset of the boundary tone in IP boundaries. The onset of the boundary tone was always during the final vowel, but it occurred earlier within the vowel when the stressed syllable occurred earlier; in other words, it shifted toward the stressed syllable.

Katsika and Katsika et al. suggest that this interdependency of prominence and boundary arises through the coordination of the μ-gesture, the boundary tone gesture, and the π-gesture. The π-gesture has a dual coordination: it is coordinated both with the phrase-final vowel gesture and with the μ-gesture (the temporal modulation gesture on the stressed syllable) of the phrase-final word. The coordination with the μ-gesture is weaker, thus affecting the onset of the π-gesture less than the π-gesture's coordination with the phrase-final vowel gesture. In this way, the onset of lengthening will depend on the position of the stressed syllable; that is, the onset will shift slightly, in a gradient manner, toward the stressed syllable. Thus, the flexible scope of lengthening is accounted for by these two coordinations.[10]

Katsika and Katsika et al. also suggest that the variability in the alignment of the boundary tone can be accounted for if it is assumed that the boundary tone gesture is activated when the π-gesture reaches a certain high level of activation (see also Byrd 2006 on this relationship between π-gesture strength of activation and boundary tone). When lexical stress of the phrase-final word is earlier in the word, the π-gesture will start earlier, and it will reach the level of activation that triggers the boundary tone earlier, thus triggering the boundary tone earlier. This means that, like the π-gesture, the onset of the boundary tone will shift slightly toward the stressed syllable (for further discussion, see Katsika 2012 and Katsika et al. 2014). Thus, the variability in the onset of lengthening and onset of the boundary tone gesture arises from the coordination of the π-gesture and the μ-gesture. This account also indicates a further interdependency of boundary properties: that between the temporal and tonal properties. Note that the coordination of prosodic gestures (which gestures are coordinated, what the strength of the coordination is, what type of coordination) can be expected to be a source of cross-linguistic variation (see Katsika et al. 2014).

Katsika and Katsika et al. further identify pause postures, which are specific configurations of the vocal tract occurring at IP boundaries. They argue that pause postures are triggered by a very high activation of the π-gesture (higher than needed for the activation of the boundary tone gesture, thus achieving that only very strong boundaries have pause postures). While pause postures need to be further examined, some corroborating evidence for pauses as linguistic entities comes from the studies of Ramarayanan et al. (2009, 2010, 2013). These studies show that grammatical pauses, defined there as pauses between syntactic constituents, have specific kinematic properties: they exhibit less variability and a decrease in velocity of articulators in comparison to other types of silent intervals. The observation of these pause-specific configurations also indicates the need for further exploration of the relationship between final lengthening, pauses, and initial lengthening (see also discussions in Byrd and Saltzman 2003; Krivokapić 2014; and Katsika et al. 2014).

To summarize, a number of recent studies (Byrd 2006, Byrd and Riggs 2008, Katsika 2012, Katsika et al. 2014) argue that prosodic boundaries arise through a set of temporally specified prosodic gestures that interact in a dynamic way. This view allows for a characterization of the tonal and temporal properties of prosodic structure, including an account of the gradient effects that prosodic structure gives rise to. Recent work on articulation during pauses has revealed further aspects of prosodic boundaries.

5.5 Polysyllabic Shortening, Rhythm, and the Prosodic Hierarchy

A recent extension of the theory is the modeling of polysyllabic shortening and rhythm. Early researchers (Pike 1945; Abercrombie 1967) developed the notions of stress timing and syllable timing and proposed that all languages are stress-timed or syllable-timed. The phonetic correlates of these two rhythm types were suggested to be foot and syllable isochrony. Since then, a large body of research has examined rhythmic properties of languages but has not found evidence for isochrony in either production or perception (Bolinger 1965; Huggins 1972; Lehiste 1972, 1977; Nakatani, O'Connor, and Aston 1981), and it is now abundantly clear that strict isochrony does not exist (see the overview in Arvaniti 2012).

Despite the clear evidence against strict isochrony, it is not disputed that there are temporal regularities, as evidenced in durational aspects of speech, although it is likely that other aspects of speech, such as the timing of pitch accents and phrasal boundaries, are also relevant for characterizing rhythm (as argued in Jun 2014). Specifically, there is evidence of polysyllabic shortening—the shortening of stressed syllables with the addition of unstressed syllables to a word or a foot—that can be seen as a tendency toward foot isochrony (Lehiste 1972; Klatt 1973; Port 1981; Rakerd, Sennett, and Fowler 1987; Kim and Cole 2005; White and Turk 2010; Shattuck-Hufnagel and Turk 2011). Kim and Cole (2005), for example, find that the duration of the foot (defined in the Abercrombian sense of extending from one stressed syllable to the next) increases with the addition of unstressed syllables. In the process, the unstressed syllables do not shorten, but the stressed syllable does. In this way, the duration of the foot is somewhat constrained.

Within Articulatory Phonology, rhythmic properties are understood to emerge through the coupling of the foot oscillator, the syllable oscillator, and the temporal modulation gesture μ (Saltzman et al. 2008; Nam et al. 2008). The μ-gesture, which is associated with stressed syllables, slows the gestural activation of coactive gestures. As a consequence, stressed syllables are longer than unstressed syllables. The foot and the syllable oscillator (and, similarly, phrase-level oscillators, although this part of the model is not fully developed yet) extend the model of planning oscillators for gestures to prosodic structure.

As Saltzman et al. (2008, 180) stated, "the foot oscillator attempts to keep the duration of the foot constant, while the syllable oscillator attempts to keep the duration of the syllable constant." These opposite goals of the oscillators, and the resulting competition, are modeled with interoscillator coupling functions, where the foot and syllable oscillators are coupled to each other bidirectionally, and each oscillator has a weighting that specifies the influence of the oscillator in the function. Depending on which oscillator's weighting is stronger, the foot or the syllable will be the more dominant oscillator, and the language will display a tendency toward foot or syllable isochrony. In the case of foot dominance, an increase in the number of syllables in the foot will lead to the shortening of syllables (the syllables are "squeezed" in the foot), leading to polysyllabic shortening (the approach discussed so far is based on O'Dell and Nieminen 1999; for related approaches see also Barbosa 2002; Cummins and Port 1998; and Tilsen 2009). To account for the fact that only the stressed (that is, not the unstressed) syllable shortens (as found in Kim and Cole 2005), the foot oscillator's strength (weight) in the interoscillator coupling function is modeled to be weaker during the unstressed than during the stressed syllables. Thus, the syllable oscillator is dominant during the unstressed syllables. This leads to the foot oscillator's squeezing being weaker during the unstressed than during the stressed syllables, which results in shortening of the stressed but not the unstressed syllables.

The model of Saltzman et al. (2008; see also Nam et al. 2008 and O'Dell and Nieminen 1999) does not predict strict isochrony, but it captures polysyllabic shortening and the tendencies toward isochrony, thus allowing both the empirical facts and the hard-to-capture division between stress-timed and syllable-timed languages to be accounted for. Typological differences between languages can be understood as the difference in the dominance of the foot and syllable oscillator (compatible with a gradiently varying scale of stress-timedness, as Dauer 1983 suggested).

Extending the planning oscillators to prosodic structure provides the possibility of understanding the prosodic hierarchy as a network of oscillators, where oscillators with higher frequencies (for lower prosodic phrases) are embedded into oscillators of lower frequencies (higher prosodic phrases), as discussed in Saltzman et al. (2008; see also Cummins and Port 1998 and Goldstein 2012). This approach has not been fully implemented within the Task Dynamics model of Articulatory Phonology (but see Tilsen's 2009 dynamical model for this type of approach). The integration of this hierarchy of oscillators with other aspects of prosodic structure (such as temporal and tonal boundary events) is also a matter for future research. One immediately obvious consequence of this conceptualization of the prosodic hierarchy is that as in the π-gesture approach, prosodic phrases based on the planning oscillators do not differ qualitatively; corresponding to the evidence outlined in section 5.1, they differ only in oscillator frequencies.

Note that two possible views of prosodic structure have been outlined: the prosodic hierarchy can be understood as arising from a hierarchy of nested planning oscillators or through the coordination of the π-gesture (which can have a categorical or a gradient activation), other prosodic gestures, and constriction gestures. These approaches also entail two distinct views of prosodic boundaries. In one view (the π-gesture view), the boundary is a phonological unit, and in the other, the boundary is a consequence of phrases, without a separate phonological status. Future research will determine which of these directions represents a more accurate way of conceptualizing the prosodic hierarchy.

5.6 Conclusion

Within Articulatory Phonology, prosodic structure is understood in a gestural manner. Prosodic boundaries are viewed as arising through the coordination of temporal, tonal, and constriction gestures. Rhythmic structure is determined by the μ-gesture and the coupling between foot and syllable oscillators.

Within this model, and as opposed to other prosodic models, temporal information is part of the linguistic representation, and the coordination of gestures is guided by general principles of coupling. The clear, predictive nature of this model has sparked numerous studies. In particular, it has allowed the question of the scope of effect of the boundary and of gestural coordination to be formulated precisely and has led to a number of studies examining the temporal properties of boundaries and the relationship between temporal properties of boundaries, tonal gestures, and prominence. The dynamical nature of the model also makes it possible to account for both gradient and categorical prosodic phenomena (as shown, e.g., in the examination of boundary strength or in the properties of the foot) in a manner that would be difficult to capture in other models. Furthermore, the simultaneous examination of temporal and tonal prosodic events is crucial for any model of prosodic theory, but is facilitated in Articulatory Phonology by the understanding that all prosodic phenomena are gestures that can be coordinated with each other, and in that way can influence one another. This view also suggests an extension of the model to integrate body gestures. Recent studies have found strong evidence that body gestures are timed to prosodic structure (Krahmer and Swerts 2007; Esteve-Gibert and Prieto 2013; Rochet-Capellan et al. 2008). Speech production models, and models of prosodic structure in particular, need to uncover the principles of this coordination. Within the Articulatory Phonology model, body gestures could be understood as gestures—for example, a hand or head gesture. These, like constriction gestures, would be associated with planning oscillators and coordinated with constriction and prosodic gestures following the same principles of coordination as other gestures.

Many aspects of the model remain to be developed, among them many that are not well understood regardless of the theoretical framework. However, the basic assumption of the model—the dynamic, gestural approach in which prosodic units have temporal properties and are coordinated with each other by general coupling mechanisms—should continue to lead to fruitful research.

Acknowledgments

I would like to thank Louis Goldstein, Argyro Katsika, Hosung Nam, and especially Elliot Saltzman for many productive discussions over the past year about questions presented here; Harim Kwon and Kate Sherwood whose questions in my Prosody 2014 seminar provided guidelines in writing this chapter; and Pam Beddor, Susanne Fuchs, and Caterina Petrone for very helpful comments on an earlier version of the manuscript. This work was supported by National Institutes of Health grant DC002717 to Douglas Whalen.

Notes

1. For overviews of the model, see Browman and Goldstein (1989, 1990a, 1990b, 1991, 1992); Goldstein, Byrd, and Saltzman (2006); Pouplier (2011); Gafos and Goldstein (2011); and the

Haskins website: http://www.haskins.yale.edu/research/gestural.html. For an introduction to Task Dynamics, see Hawkins (1992), and for an overview of the application of dynamical systems theory to speech, see van Lieshout (2004).

2. A dynamical system is one in which future states of the system are determined by its present state and a set of laws that specify the forces changing the system (Saltzman and Kelso 1987; Saltzman 1995). The computational model is named "Task Dynamics" to capture that "(a) it deals with the performance of well-learned skilled movements or gestures designed to accomplish real-world tasks; and (b) it is defined with respect to the dynamics that underlie a given action's kinematics. Note that kinematics refers to a gesture's observable spatiotemporal properties (e.g., its position, velocity, and acceleration trajectories over time), while dynamics refers to the pattern of the underlying field of forces that gives rise to these kinematics" (Saltzman 1986, 130).

3. This view of phonology and phonetics does not lend itself to easy transcription, and this applies to the model of prosody as well.

4. TADA can be downloaded at http://www.haskins.yale.edu/tada_download/index.php and the manual for the model (Nam and Goldstein 2007) is available at http://www.haskins.yale.edu/tada_download/doc/TADA_manual_v09.pdf.

5. The π-gesture has not been integrated in the coupled oscillator model at this point, but the coordination of the π-gesture to constriction gestures is a central question of the theory, and I discuss approaches to this question in section 5.4.

6. Gestures also become slower (with lower peak velocities), but this interacts with gesture displacement (or amplitude), which means that the change in peak velocity might not always be observable. See Byrd and Saltzman (2003).

7. A spatial-modulation gesture, which modifies the spatial extents of constriction gestures, is discussed in Saltzman et al. (2008). At this point, however, it has not been used extensively in the model because it is not yet clear whether a separate spatial-modulation gesture is necessary to alter spatial extent, or whether such effects can be derived indirectly via the truncation/detruncation effects that are direct outcomes of applying a temporal modulation gesture (E. Saltzman, personal communication, 2015; see also Byrd and Saltzman 2003).

8. The π-gesture and the μ-gesture are conceptually similar, but the π-gesture has not yet been implemented in the coupled planning oscillator network. Both locally modulate the temporal properties of an utterance. For a further discussion of the entailed differences, see Saltzman et al. (2008). For lengthening under prominence, see, for example, Turk and Sawusch (1997), Turk and White (1999), Cho and Keating (2009), Mücke and Grice (2014), and the overview in Fletcher (2010).

9. While the typological predictions of the prosodic model in Articulatory Phonology have not been much discussed, the coordination between gestures can be expected to be a major source of typological variation. See discussions about typology of syllable structure, and how it relates to coupling graphs, in Nam, Goldstein, and Saltzman (2009).

10. For a related approach see Rusaw (2013), who modeled the interaction between boundary and prominence lengthening, as found in Turk and Shattuck-Hufnagel (2007) and replicated in Rusaw (2013), with an artificial central pattern generator neural network model.

References

Abercrombie, D. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Arvaniti, A. 2012. "The Usefulness of Metrics in the Quantification of Speech Rhythm." *Journal of Phonetics* 40:351–373.

Barbosa, P. A. 2002. "Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production." In *Proceedings of the first International Conference on Speech Prosody*, Aix-en-Provence, France. https://www.isca-speech.org/archive/sp2002/papers/sp02_163.pdf.

Beckman, M. E. 1996. "The Parsing of Prosody." *Language and Cognitive Processes* 11:17–67.

Beckman, M. E., and J. B. Pierrehumbert. 1986. "Intonational Structure in Japanese and English." *Phonology Yearbook* 3:255–309.

Berkovits, R. 1993a. "Progressive Utterance: Final Lengthening in Syllables with Final Fricatives." *Language and Speech* 36:89–98.

Berkovits, R. 1993b. "Utterance-Final Lengthening and the Duration of Final-Stop Closures." *Journal of Phonetics* 21:479–489.

Bolinger, D. L. 1965. "Pitch Accent and Sentence Rhythm." In *Forms of English: Accent, Morpheme, Order*, edited by I. Abe and T. Kanekiyo, 139–180. Cambridge, MA: Harvard University Press.

Bombien, L., C. Mooshammer, and P. Hoole. 2013. "Articulatory Coordination in Word-Initial Clusters of German." *Journal of Phonetics* 41:546–561.

Bombien, L., C. Mooshammer, P. Hoole, and B. Kühnert. 2010. "Prosodic and Segmental Effects on EPG Contact Patterns of Word-Initial German Clusters." *Journal of Phonetics* 38:388–403.

Browman, C. P., and L. Goldstein. 1988. "Some Notes on Syllable Structure in Articulatory Phonology." *Phonetica* 45:140–155.

Browman, C. P., and L. Goldstein. 1989. "Articulatory Gestures as Phonological Units." *Phonology* 6:201–251.

Browman, C. P., and L. Goldstein. 1990a. "Gestural Specification Using Dynamically-Defined Articulatory Structures." *Journal of Phonetics* 18:299–320.

Browman, C. P., and L. Goldstein. 1990b. "Representation and Reality: Physical Systems and Phonological Structure." *Journal of Phonetics* 18:411–424.

Browman, C. P., and L. Goldstein. 1991. "Tiers in Articulatory Phonology, with Some Implications for Casual Speech." In *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, edited by J. Kingston and M. E. Beckman, 341–376. Cambridge: Cambridge University Press.

Browman, C. P., and L. Goldstein. 1992. "Articulatory Phonology: An Overview." *Phonetica* 49:155–180.

Browman, C. P., and L. Goldstein. 1995. "Dynamics and Articulatory Phonology." In *Mind as Motion*, edited by R. F. Port and T. van Gelder, 175–193. Cambridge, MA: MIT Press.

Browman, C., and L. Goldstein. 2000. "Competing Constraints on Intergestural Coordination and Self-Organization of Phonological Structures." *Bulletin de la Communication Parlée* 5:25–34.

Brunner, J., Geng, C., Sotiropoulou, S., and Gafos, A. 2014. "Timing of German Onset and Word Boundary Clusters." *Laboratory Phonology* 5:403–454.

Byrd, D. 2000. "Articulatory Vowel Lengthening and Coordination at Phrasal Junctures." *Phonetica* 57:3–16.

Byrd, D. 2006. "Relating Prosody and Dynamic Events: Commentary on the Papers by Cho, Navas, and Smiljanić." In *Laboratory Phonology 8: Varieties of Phonological Competence*, edited by L. Goldstein, 549–561. Berlin: Mouton de Gruyter.

Byrd, D., and S. Choi. 2010. "At the Juncture of Prosody, Phonology, and Phonetics: The Interaction of Phrasal and Syllable Structure in Shaping the Timing of Consonant Gestures." In *Papers in Laboratory Phonology 10*, edited by C. Fougeron, B. Kühnert, M. Imperio, and N. Vallee, 31–59. Berlin: Mouton de Gruyter.

Byrd, D., A. Kaun, S. Narayanan, and E. Saltzman. 2000. "Phrasal Signatures in Articulation." In *Papers in Laboratory Phonology V*, edited by M. B. Broe and J. P. Pierrehumbert, 70–87. Cambridge: Cambridge University Press.

Byrd, D., J. Krivokapić, and S. Lee. 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., and D. Riggs. 2008. "Locality Interactions with Prominence in Determining the Scope of Phrasal Lengthening." *Journal of the International Phonetic Association* 38:187–202.

Byrd, D., and E. Saltzman. 1998. Intragestural Dynamics of Multiple Phrasal Boundaries." *Journal of Phonetics* 26:173–199.

Byrd, D., and E. Saltzman. 2003. "The Elastic Phrase: Dynamics of Boundary-Adjacent Lengthening." *Journal of Phonetics* 31:149–180.

Byrd, D., S. Tobin, E. Bresch, and S. Narayanan. 2009. "Timing Effects of Syllable Structure and Stress on Nasals: A Real-Time MRI Examination." *Journal of Phonetics* 37:97–110.

Cambier-Langeveld, T. 1997. "The Domain of Final Lengthening in the Production of Dutch." In *Linguistics in the Netherlands*, edited by J. Coerts and H. de Hoop, 13–24. Amsterdam: John Benjamins.

Carlson, K., C. Clifton Jr., and L. Frazier. 2001. "Prosodic Boundaries in Adjunct Attachment." *Journal of Memory and Language* 45:58–81.

Cho, T. 2006. "Manifestation of Prosodic Structure in Articulation: Evidence from Lip Kinematics in English." In *Laboratory Phonology 8: Varieties of Phonological Competence*, edited by L. Goldstein, 519–548. New York: Walter de Gruyter.

Cho, T., and P. Keating. 2001. "Articulatory and Acoustic Studies of Domain-Initial Strengthening in Korean." *Journal of Phonetics* 29:155–190.

Cho, T., and P. Keating. 2009. "Effects of Initial Position versus Prominence in English." *Journal of Phonetics* 37:466–485.

Cho, T., and J. McQueen. 2005. "Prosodic Influences on Consonant Production in Dutch: Effects of Prosodic Boundaries, Phrasal Accent, and Lexical Stress." *Journal of Phonetics* 33: 121–157.

Clifton, C. J., K. Carlson, and L. Frazier. 2002. "Informative Prosodic Boundaries." *Language and Speech* 45:87–114.

Cooper, W. E., and J. Paccia-Cooper. 1980. *Syntax and Speech*. Cambridge, MA: Harvard University Press.

Cummins, F., and R. Port. 1998. "Rhythmic Constraints on Stress Timing in English." *Journal of Phonetics* 26:145–171.

Dauer, R. M. 1983. "Stress-Timing and Syllable-Timing Reanalyzed." *Journal of Phonetics* 11:51–69.

de Pijper, J. R., and A. A. Sanderman. 1994. "On the Perceptual Strength of Prosodic Boundaries and Its Relation to Suprasegmental Cues." *Journal of the Acoustical Society of America* 96:2037–2047.

D'Imperio, M. 2001. "Focus and Tonal Structure in Neapolitan Italian." *Speech Communication* 33:339–256.

D'Imperio, M., R. Espesser, H. Loevenbruck, C. Menezes, N. Nguyen, and P. Welby. 2007. "Are Tones Aligned with Articulatory Events? Evidence from Italian and French." In *Papers in Laboratory Phonology 9*, edited by J. Cole and J. I. Hualde, 577–608. Berlin: Mouton de Gruyter.

Esteve-Gibert, N., and P. Prieto. 2013. "Prosodic Structure Shapes the Temporal Realization of Intonation and Manual Gesture Movements." *Journal of Speech, Language, and Hearing Research* 56:850–864.

Fletcher, J. 2010. "The Prosody of Speech: Timing and Rhythm." In *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle, J. Laver, and F. E. Gibbon, 523–602. Hoboken, NJ: Wiley-Blackwell.

Frota, S. 2000. *Prosody and Focus in European Portuguese: Phonological Phrasing and Intonation*. New York: Garland.

Fougeron, C. 2001. "Domain-Initial Articulatory Variations in French." *Journal of Phonetics* 29:109–135.

Fougeron, C., and P. Keating. 1997. "Articulatory Strengthening at Edges of Prosodic Domains." *Journal of the Acoustical Society of America* 101:3728–3740.

Fowler, C. A. 1980. "Coarticulation and Theories of Extrinsic Timing." *Journal of Phonetics* 8:113–133.

Gafos, A., and L. Goldstein. 2011. "Articulatory Representation and Organization." In *The Handbook of Laboratory Phonology*, edited by A. C. Cohn, C. Fougeron, and M. K. Huffman, 220–231. Oxford: Oxford University Press.

Gao, M. 2008. "Mandarin Tones: An Articulatory Phonology Account." PhD diss., Yale University.

Gao, M. 2009. "Gestural Coordination among Vowel, Consonant and Tone Gestures in Mandarin Chinese." *Chinese Journal of Phonetics* 2:43.

Goldstein, L. 2012. "Dynamical Models of the Prosodic Hierarchy." Paper presented at the Workshop on Prosody of the Wider World, Cologne, Germany, May 30–June 1.

Goldstein, L., D. Byrd, and E. Saltzman, E. 2006. "The Role of Vocal Tract Gestural Action Units in Understanding the Evolution of Phonology." In *Action to Language via the Mirror Neuron System*, edited by M. Arbib, 215–249. Cambridge: Cambridge University Press.

Goldstein, L., H. Nam, E. Saltzman, and I. Chitoran. 2009. "Coupled Oscillator Planning Model of Speech Timing and Syllable Structure." In *Frontiers in Phonetics and Speech Science*, edited by G. Fant, H. Fujisaki, and J. Shen, 239–250. Beijing: Commercial Press.

Goldstein, L., M. Pouplier, L. Chen, E. Saltzman, and D. Byrd. 2007. "Dynamic Action Units Slip in Speech Production Errors." *Cognition* 103:386–412.

Haken, H., J. A. S. Kelso, and H. Bunz. 1985. "A Theoretical Model of Phase Transitions in Human Hand Movements." *Biological Cybernetics* 51:347–356.

Hanson, H., and K. N. Stevens. 2002. "A Quasi-Articulatory Approach to Controlling Acoustic Source Parameters in a Klatt-Type Formant Synthesizer Using HLSyn." *Journal of the Acoustical Society of America* 112:1158–1182.

Hardcastle, W. J. 1985. "Some Phonetic and Syntactic Constraints on Lingual Coarticulation during /kl/ Sequences." *Speech Communication* 4:247–263.

Hawkins, S. 1992. "An Introduction to Task Dynamics." In *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, edited by G. J. Docherty and D. R. Ladd, 9–25. Cambridge: Cambridge University Press.

Hermes, A. 2013. "Articulatory Coordination and Syllable Structure in Italian." In *Speech Production and Perception*, edited by S. Fuchs and P. Perrier, 9–25. Frankfurt am Main: Peter Lang Verlag.

Hirose, H. 2010. "Investigating the Physiology of Laryngeal Structures." In *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle, J. Laver, and F. E. Gibbon, 130–152. Hoboken, NJ: Wiley-Blackwell.

Hombert, J.-M., J. J. Ohala, and W. G. Ewan. 1979. "Phonetic Explanations for the Development of Tones." *Language* 55:37–58.

Hsieh, F.-Y. 2011. "A Gestural Account of Mandarin Tone 3 Variation." In *Proceedings of the Seventeenth International Congress of Phonetic Sciences*, 890–893. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/index.htm.

Huggins, A. W. F. 1972. "On the Perception of Temporal Phenomena in Speech." *Journal of the Acoustical Society of America* 51:1279–1290.

Jun, S.-A. 1993. "The Phonetics and Phonology of Korean Prosody." PhD diss., Ohio State University.

Jun, S.-A. 2005. "Prosodic Typology." In *Prosodic Typology: The Phonology of Intonation and Phrasing*, edited by S.-A. Jun, 430–458. Oxford: Oxford University Press.

Jun, S.-A. 2014. "Prosodic Typology: By Prominence Type, Word Prosody, and Macro-rhythm." In *Prosodic Typology II: The Phonology of Intonation and Phrasing*, edited by S.-A. Jun, 520–540. Oxford: Oxford University Press.

Katsika, A. 2012. "Coordination of Prosodic Gestures at Boundaries in Greek." PhD diss., Yale University.

Katsika. A. 2016. "The Role of Prominence in Determining the Scope of Boundary-Related Lengthening." *Journal of Phonetics* 55:149–181.

Katsika, A., J. Krivokapić, C. Mooshammer, M. Tiede, and L. M. Goldstein. 2014. "The Coordination of Boundary Tones and Their Interaction with Prominence." *Journal of Phonetics* 44:62–82.

Keating, P., T. Cho, C. Fougeron, and C. Hsu. 2004. "Domain-Initial Articulatory Strengthening in Four Languages." In *Phonetic Interpretation*, edited by J. Local, R. Ogden, and R. Temple, 143–161. Cambridge: Cambridge University Press.

Kim, H., and J. Cole. 2005. "The Stress Foot as a Unit of Planned Timing: Evidence from Shortening in the Prosodic Phrase." In *Proceedings of Interspeech 2005*, 2365–2368. https://www.isca-speech.org/archive/archive_papers/interspeech_2005/i05_2365.pdf.

Klatt, D. H. 1973. "Interaction between Two Factors That Influence Vowel Duration." *Journal of the Acoustical Society of America* 54:1102–1104.

Krahmer, E., and M. Swerts. 2007. "The Effects of Visual Beats on Prosodic Prominence: Acoustic Analyses, Auditory Perception, and Visual Perception." *Journal of Memory and Language* 57 (3): 396–414.

Krakow, R. A. 1999. "Physiological Organization of Syllables: A Review." *Journal of Phonetics* 27:23–54.

Krivokapić, J. 2007. "The Planning, Production, and Perception of Prosodic Structure." PhD diss., University of Southern California.

Krivokapić, J. 2014. "Gestural Coordination at Prosodic Boundaries and Its Role for Prosodic Structure and Speech Planning Processes." *Communicative Rhythms in Brain and Behaviour. Philosophical Transactions of the Royal Society B (Biology)* 369:20130397.

Krivokapić, J., and D. Byrd. 2012. "Prosodic Boundary Strength: An Articulatory and Perceptual Study." *Journal of Phonetics* 40:430–442.

Ladd, D. R. 1988. "Declination 'Reset' and the Hierarchical Organization of Utterances." *Journal of the Acoustical Society of America* 84:530–544.

Ladd, D. R. (1996) 2008. *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladd, D. R. 2001. "Intonation." In *Language Typology and Language Universals: An International Handbook*, edited by M. Haspelmath, E. König, W. Oesterreicher, and W. Raible, 1380–1390. Berlin: Mouton de Gruyter.

Ladd, D. R. 2006. "Segmental Anchoring of Pitch Movements: Autosegmental Association or Gestural Coordination?" *Italian Journal of Linguistics* 18 (1): 19–38.

Ladd, D. R., I. Mennen, and A. Schepman. 2000. "Phonological Conditioning of Peak Alignment of Rising Pitch Accents in Dutch." *Journal of the Acoustical Society of America* 107:2685–2696.

Lehiste, I. 1972. "The Timing of Utterances and Linguistic Boundaries." *Journal of the Acoustical Society of America* 51:2018–2024.

Lehiste, I. 1977. "Isochrony Reconsidered." *Journal of Phonetics* 5:253–263.

Marin, S., and M. Pouplier. 2010. "Temporal Organization of Complex Onsets and Codas in American English: Testing the Predictions of a Gestural Coupling Model." *Motor Control* 14:380–407.

McClean, M. 1973. "Forward Coarticulation of Velar Movement at Marked Junctural Boundaries." *Journal of Speech, Language and Hearing Research* 16:286–296.

McGowan, R. S., and E. Saltzman. 1995. "Incorporating Aerodynamic and Laryngeal Components into Task Dynamics." *Journal of Phonetics* 23:255–269.

Mücke, D., and M. Grice. 2014. "The Effect of Focus Marking on Supralaryngeal Articulation: Is It Mediated by Accentuation?" *Journal of Phonetics* 44:47–61.

Mücke, D., H. Nam, A. Hermes, and L. Goldstein. 2012. "Coupling of Tone and Constriction Gestures in Pitch Accents." In *Consonant Clusters and Structural Complexity*, edited by P. Hoole, M. Pouplier, L. Bombien, C. Mooshammer, and B. Kühnert, 205–230. Berlin: Mouton de Gruyter.

Nakatani, L. H., K. O'Connor, and C. H. Aston. 1981. "Prosodic Aspects of American English Speech Rhythm." *Phonetica* 38:84–106.

Nam, H. 2007. "Syllable-Level Intergestural Timing Model: Split-Gesture Dynamics Focusing on Positional Asymmetry and Moraic Structure." In *Laboratory Phonology 9*, edited by J. Cole and J. I. Hualde, 483–506. Berlin: Walter de Gruyter.

Nam, H., and L. Goldstein. 2007. *TADA (TAsk Dynamics Application) Manual*. http://www.haskins.yale.edu/tada_download/doc/TADA_manual_v09.pdf.

Nam, H., L. Goldstein, and E. Saltzman, E. 2009. "Self-Organization of Syllable Structure: A Coupled Oscillator Model." In *Approaches to Phonological Complexity*, edited by F. Pellegrino, E. Marisco, and I. Chitoran, 299–328. Berlin: Mouton de Gruyter.

Nam, H., L. Goldstein, E. Saltzman, and S. Byrd. 2004. "TADA: An Enhanced, Portable Task Dynamics Model in MATLAB." *Journal of the Acoustical Society of America* 115:2430.

Nam, H., and E. Saltzman, E. 2003. “A Competitive, Coupled Oscillator of Syllable Structure.” Paper presented at the Fifteenth International Congress of Phonetic, Barcelona, August 3.

Nam, H., E. Saltzman, J. Krivokapić, and L. Goldstein. 2008. “Modeling the Durational Difference of Stressed vs. Unstressed Syllables.” In *Proceedings of the Eighth Phonetic Conference of China*. https://sites.lsa.umich.edu/jelenakrivokapic/wp-content/uploads/sites/201/2014/12/Nam_Saltzman_Krivokapic_Goldstein_PCC_2008.pdf.

Nespor, M., and I. Vogel. (1986) 2007. *Prosodic Phonology*. Dordrecht, the Netherlands: Foris.

Niemann, H., D. Mücke, H. Nam, L. Goldstein, and M. Grice. 2011. “Tones as Gestures: The Case of Italian and German.” In *Proceedings of the Seventeenth International Congress of Phonetic Sciences*, 1486–1489. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/index.htm.

O’Dell, M. L., and T. Nieminen. 1999. “Coupled Oscillator Model of Speech Rhythm.” In *Proceedings of the Fourteenth International Congress of Phonetic Sciences*, edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey, vol. 2, 1075–1078. New York: American Institute of Physics.

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.

Pike, K. 1945. *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Port, R. 1981. “Linguistics Timing Factors in Combination.” *Journal of the Acoustical Society of America* 69:262–274.

Port, R. F., and T. van Gelder, eds. 1995. *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press.

Pouplier, M. 2011. “The Atoms of Phonological Representations.” In *The Blackwell Companion to Phonology*, edited M. van Oostendorp, K. Rice, B. Hume, and C. Ewen, 107–129. Hoboken, NJ: Wiley-Blackwell.

Prieto, P., and F. Torreira. 2007. “The Segmental Anchoring Hypothesis Revisited: Syllable Structure and Speech Rate Effects on Peak Timing in Spanish.” *Journal of Phonetics* 35:473–500.

Prieto, P., J. van Santen, and J. Hirschberg. 1995. “Tonal Alignment Patterns in Spanish.” *Journal of Phonetics* 23:429–451.

Proctor, M. 2009. “Gestural Characterization of a Phonological Class: The Liquids.” PhD diss., Yale University.

Rakerd, B., W. Sennett, and C. A. Fowler 1987. “Domain-Final Lengthening and Foot Level Shortening in Spoken English.” *Phonetica* 44:147–155.

Ramanarayanan, V., E. Bresch, D. Byrd, L. Goldstein, and S. Narayanan. 2009. “Analysis of Pausing Behavior in Spontaneous Speech Using Real-Time Magnetic Resonance Imaging of Articulation.” *Journal of the Acoustical Society of America Express Letters* 126:EL160–EL165.

Ramanarayanan, V., D. Byrd, L. Goldstein, and S. Narayanan. 2010. “Investigating Articulatory Setting—Pauses, Ready Position, and Rest—Using Real-Time MRI.” In *Proceedings of Interspeech 2010*, 1994–1997. https://www.isca-speech.org/archive/archive_papers/interspeech_2010/i10_1994.pdf.

Ramanarayanan, V., L. Goldstein, D. Byrd, and S. Narayanan. 2013. “A Real-Time MRI Investigation of Articulatory Setting across Different Speaking Styles.” *Journal of the Acoustical Society of America* 134:510–519.

Rochet-Capellan, A., R. Laboissière, A. Galván, and J. L. Schwartz. 2008. "The Speech Focus Position Effect on Jaw-Finger Coordination in a Pointing Task." *Journal of Speech Language and Hearing Research* 51:1507–1521.

Rusaw, E. 2013. "Modeling Temporal Coordination in Speech Production Using an Artificial Central Pattern Generator Neural Network." PhD diss., University of Illinois.

Saltzman, E. L. 1986. "Task Dynamic Coordination of the Speech Articulators: A Preliminary Model." In *Generation and Modulation of Action Patterns*, edited by H. Heuer and C. Fromm, 129–144. New York: Springer.

Saltzman, E. L. 1995. "Dynamics and Coordinate Systems in Skilled Sensorimotor Activity." In *Mind as Motion: Explorations in the Dynamics of Cognition*, edited by R. F. Port and T. van Gelder, 149–173. Cambridge, MA: MIT Press.

Saltzman, E. L., and D. Byrd. 2000. "Task-Dynamics of Gestural Timing: Phase Windows and Multifrequency Rhythms." *Human Movement Science* 19 (4): 499–526.

Saltzman, E. L., and J. A. S. Kelso. 1987. "Skilled Actions: A Task-Dynamic Approach." *Psychological Review* 94:84–106.

Saltzman, E. L., and K. G. Munhall. 1989. "A Dynamical Approach to Gestural Patterning in Speech Production." *Ecological Psychology* 1:333–382.

Saltzman, E., H. Nam, J. Krivokapić, and L. Goldstein. 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, 175–184. http://www.isca-speech.org/archive/sp2008.

Selkirk, E. O. 1984. *Phonology and Syntax*: *The Relation between Sound and Structure*. Cambridge, MA: MIT Press.

Selkirk, E. O. 2011. "The Syntax-Phonology Interface." In *The Handbook of Phonological Theory*, edited by J. Goldsmith, J. Riggle, and A. Yu, 435–484. Oxford: Blackwell.

Shattuck-Hufnagel, S., and A. Turk. 1996. "A Prosody Tutorial for Investigators of Auditory Sentence Processing." *Journal of Psycholinguistic Research* 25:193–247.

Shattuck-Hufnagel, S., and A. Turk. 1998. "The Domain of Phrase-Final Lengthening in English.", *Proceedings of the Sixteenth International Congress on Acoustics and 135th Meeting of the Acoustical Society of America*, Vol 2, 1235–1236.

Shattuck-Hufnagel, S. A., and A. Turk. 2011. "Durational Evidence for Word-Based vs. Prominence-Based Constituent structure in limerick speech." In *Proceedings of the Seventeenth International Congress of Phonetic Sciences*, 1806–1809. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/index.htm.

Silverman, K., and J. Pierrehumbert. 1990. "The Timing of Prenuclear High Accents in English." In *Papers in Laboratory Phonology I*, edited by J. Kingston and M. Beckman, 72–106. Cambridge: Cambridge University Press.

Snedeker, J., and E. Casserly. 2010. "Is It All Relative? Effects of Prosodic Boundaries on the Comprehension and Production of Attachment Ambiguities." *Language and Cognitive Processes* 25:1234–1264.

Sproat, R., and O. Fujimura. 1993. "Allophonic Variation in English /1/ and Its Implications for Phonetic Implementation." *Journal of Phonetics* 21:291–311.

Swerts, M. 1997. “Prosodic Features at Discourse Boundaries of Different Strength.” *Journal of the Acoustical Society of America* 101:514–521.

Tabain, M. 2003. “Effects of Prosodic Boundary on /aC/ Sequences: Articulatory Results.” *Journal of the Acoustical Society of America* 113:2834–2849.

Tilsen, S. 2009. “Multitimescale Dynamical Interactions between Speech Rhythm and Gesture.” *Cognitive Science* 33:839–879.

Turk, A. E., and J. R. Sawusch. 1997. “The Domain of Accentual Lengthening in American English.” *Journal of Phonetics* 25:25–41.

Turk, A., and S. Shattuck-Hufnagel, S. 2007. “Phrase-Final Lengthening in American English.” *Journal of Phonetics* 35:445–472.

Turk, A. E., and L. White. 1999. “Structural Influences on Accentual Lengthening.” *Journal of Phonetics* 27:171–206.

Turvey, M. T. 1990. “Coordination.” *American Psychologist* 45:938–953.

van Lieshout, P. 2004. “Dynamical Systems Theory and Its Application in Speech.” In *Speech Motor Control in Normal and Disordered Speech*, edited by B. Maassen, R. Kent, H. Peters, P. van Lieshout, and W. Hulstijn, 51–82. Oxford: Oxford University Press.

Wagner, M. 2005. “Prosody and Recursion.” PhD diss., MIT.

White, L. S. 2002. “English Speech Timing: A Domain and Locus Approach.” PhD diss., University of Edinburgh.

White, L., and A. E. Turk. 2010. “English Words on the Procrustean Bed: Polysyllabic Shortening Reconsidered.” *Journal of Phonetics* 38:459–471.

Wightman, C. W., S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price. 1992. “Segmental Durations in the Vicinity of Prosodic Phrase Boundaries.” *Journal of the Acoustical Society of America* 91:1707–1717.

Xu, Y. 1997. “Contextual Tonal Variations in Mandarin.” *Journal of Phonetics* 25:61–83.

Yi, H., and S. Tilsen. 2014. “A Gestural Account of Mandarin Tone Sandhi.” In *Proceedings of Meetings on Acoustics*, 060003. http://dx.doi.org/10.1121/2.0000029.

Commentary on Chapter 5: Time to Consider Nonoscillatory, Phonology-Extrinsic Timing Approaches to Prosody in Speech Motor Control

A. E. Turk

Introduction

Articulatory phonology/task dynamics (hereafter AP/TD) is the theory of speech production that provides the most comprehensive account of how prosodic structure influences the surface phonetic form of an utterance (Browman and Goldstein 1992; Browman and Goldstein 1995; Goldstein, Byrd, and Saltzman 2006; Saltzman and Munhall 1989; Nam, Goldstein, and Saltzman 2010; Saltzman et al. 2008). In this chapter, Krivokapić describes the mechanisms AP/TD uses to model coordination among speech articulators and a wide range of prosodic influences on coordination patterns. In this approach, prosodic gestures (*point-attractor oscillators*, i.e., damped oscillators that approximate a target and go no further) are used to model boundary-related lengthening, prominence-related lengthening, lexical tone, and pitch accent.

A second mechanism, a hierarchy of limit cycle planning oscillators (nondamped oscillators that actually oscillate), is used to model coordination and polysyllabic shortening. Because each gesture is assumed to be associated with a limit cycle planning oscillator, which freely oscillates, interoscillator coupling (entrainment) mechanisms can be used to model coordination among all types of prosodic gestures and segmental gestures and to model relationships among levels in a hierarchy of syllable, foot, and phrase oscillators. As Krivokapić shows, AP/TD is impressive in its coverage, and it is unrivaled among existing models of speech production in its accounts of temporal and prosodic phenomena.

The AP/TD view has been developed over many years and presented in many separate research articles, each dealing with a specific aspect of prosodic modeling. A major contribution of Krivokapić's essay is that it presents a synthesized view of the current AP/TD approach, drawn from a number of published sources. As a result, her review will be extremely useful to anyone working within the AP/TD approach, as well as to others working in other approaches who wish to understand it better.

In this commentary, I focus on one particular aspect of the theory: AP/TD's coupled oscillator approach to polysyllabic shortening (see next section), with some additional comments on the overall approach to speech timing. This aspect has been selected because it is a critical core assumption of the theory, which I argue is not compatible with a range of findings in the motor control literature. The additional comments address two further core assumptions: that gestures are coordinated with respect to their onsets, and that speech timing is controlled by phonology-intrinsic timing mechanisms. The first of these further assumptions is addressed in the "Challenges to AP/TD's Onset-Based Coordination" section, which presents evidence that motivates an alternative to onset-based coordination: coordination based on a goal-related part of movement (often the movement endpoint). The second is addressed in the "Challenges to AP/TD's Phonology-Intrinsic Timing" section, which presents evidence that motivates an alternative to phonology-intrinsic timing: phonology-extrinsic, general-purpose timing mechanisms. The discussion in these two sections is necessarily short; see Turk and Shattuck-Hufnagel 2020) for a more detailed treatment.

Accounts of Poly-subconstituent Shortening: A Hierarchy of Syllable, Foot, and Phrase Oscillators versus a Means of Signaling Word Boundaries

As Krivokapić described, AP/TD's approach to speech timing is based on gesture onset coordination via coupling between sets of planning oscillators for syllables, feet, and phrases. In this approach, the relative timing of gesture onsets is controlled via stable entrainment relationships (in-phase and antiphase) of the hierarchy of gestural coupling oscillators. The lengthening associated with prosodic prominence and boundaries is accounted for using π and μ_T gestures, which stretch the default gestural activation intervals with which they overlap.

The main motivation for the hierarchy of coupled syllable, foot, and phrase oscillators in AP/TD has come from findings of poly-segmental and polysyllabic shortening, where more subconstituents in a larger constituent (e.g., segments or syllables in a cross-word foot) results in durational compression of these segments or syllables. The duration of the larger constituent is shorter than expected based on the number of its subconstituent segments and/or syllables and their duration when produced in isolation. The constituent whose subconstituents undergo compression is often proposed to be a cross-word foot, that is, an interstress interval[1] that can, but does not have to, include multiple words or word fragments. On this view, for example, [*bake avo-*] in *bake avocadoes* and [*bake*] in *bake apples* would both be considered cross-word feet.

English and Swedish have been claimed to have polysyllabic shortening within cross-word feet. Campbell (1988, cited in Williams and Hiller 1994), Eriksson (1991), Williams and Hiller (1994), Kim and Cole (2005), and Kim (2006) have shown shorter stressed syllable durations when additional syllables follow the stressed syllable within a cross-word foot, even though the effect size is small (e.g., 10 to 15 percent per additional syllable, in Williams and Hiller's 1994 study). Moreover, the total duration of the cross-word foot increases linearly with additional syllables and segments (see Dauer 1983 for a summary of such findings from the literature). Although compression does not yield cross-word foot isochrony, these findings suggest a *tendency* toward isochrony that might be due to periodic control at this level. As Classe (1939, 87) puts it, the surface shortening patterns might arise from a "rhythmic tendency," which "has to contend with other factors which obscure its effects"; for example, as proposed in O'Dell and Nieminen (1999), Barbosa (2007), and AP/TD (Saltzman et al. 2008), from the interaction of coupled suprasegmental planning oscillators at multiple levels of prosodic constituency (syllable, cross-word foot, and phrase), which yields tendencies toward isochrony on the surface.

However, the interpretation of these findings as unambiguous evidence for syllable, cross-word foot, and phrase oscillators is not conclusive for two reasons. First, in most cases, the researchers did not control for the locations of word and/or phrase boundaries in their experimental materials. Thus, it is difficult to be sure whether apparent shortening in cross-word feet might be a spurious effect of word- or phrase-final lengthening. Word- and especially phrase-final syllables are known to be longer than non–word-final syllables. Thus, in corpus studies, it is possible that most monosyllabic cross-word feet correspond to single words that might exhibit word- or phrase-final lengthening; final lengthening might therefore account for the difference in duration between monosyllabic (e.g., *bake*) versus polysyllabic (e.g., *bake avo-*) cross-word feet, without requiring a separate oscillator-based polysyllabic shortening mechanism (Windmann, Šimko, and Wagner 2015).

Second, the interpretation of findings as evidence for syllable, foot, and phrase oscillators is made difficult because alternative possibilities for the constituents that might be responsible for the compression effects were often not considered. For example, whether polysyllabic shortening might operate in intervals other than cross-word feet, such as words or word-based constituents. With respect to this second question, Kim (2006) tested whether shortening occurs within words, rather than cross-word feet, for one of two speakers in the study, but did not test the possibility that it occurs within word-based content + function word units, as opposed to within interstress intervals. If the polysyllabic shortening effects described are spurious, then the main motivation for using suprasegmental oscillators to model such effects would be removed. Moreover, if polysyllabic shortening operates within intervals other than those proposed in AP/TD (that is, in units other than the syllable, cross-word foot, and phrase), then the theory would need to be modified to accommodate these different units.

Windmann, Šimko, and Wagner (2015) tested for these possibilities in a corpus study of 5.5 hours of automatically segmented broadcast speech produced by fifty-three speakers of British English. Like others in the literature, they found a clear compression effect of number of syllables on syllable duration in cross-word feet defined on the basis of word-level stress, especially for syllables that were phrasally prominent. However, this effect was due almost entirely to word-final lengthening; when the position of the measured syllable with respect to the word boundary was taken into account, that is, when separate analyses were conducted for word-final and nonfinal syllables, polysyllabic shortening effects within cross-word feet largely disappeared. This result suggests that compression mechanisms for syllables might not be needed as long as word- and phrase-final lengthening mechanisms are available.

A few effects consistent with polysyllabic shortening within cross-word feet nevertheless remained after position with respect to word boundaries was controlled for. For example, Windmann, Juraj, and Wagner (2015) found longer word-final syllables before a following word-initial stressed syllable than before a following word-initial unstressed syllable; this occurred both for unstressed syllables when preceded by a stressed syllable and for stressed syllables themselves. Effects of this type, which have been called *stress-adjacent lengthening effects*, have been found elsewhere, by, for example, Fowler (1977); also Rakerd, Sennett, and Fowler (1987); van Lancker, Kreiman, and Bolinger (1988); and Fant, Kruckenberg and Nord (1991), discussed in White (2014).

These effects are consistent with polysyllabic shortening within cross-word feet, but it is still unclear which type of constituent is responsible. If the unstressed syllable that followed the word boundary consisted of a function word (e.g., *a, us, the, him*), and the effect did not occur when the following unstressed syllable formed part of a following word, the results would be consistent with polysyllabic shortening within a word-based constituent that includes a following function word (sometimes called a clitic group) rather than within a cross-word foot that includes word fragments. It is true that, either way, these findings suggest that there are durational effects additional to boundary-related lengthening and prominence effects that need to be accounted for in some way. Critically, however, it is still unclear whether the observed effects involve cross-word feet that may contain word fragments, as proposed by AP/TD and Kim (2006), or whether they involve a word-based constituent such as the word-based clitic group.

One preliminary study that did test for the clitic group, word, and cross-word foot as possible domains of polysyllabic shortening found more support for word-based prosodic units (words and clitic groups) than for cross-word feet that may contain word fragments. Shattuck-Hufnagel and Turk (2011) contrasted these three types of candidate units in metrically regular poetic contexts (i.e., reciting limericks), where cross-word foot periodicity involving word fragments would be most likely to surface if they play a role in speech. They found that the rhyme interval durations of, for example, *bake* in *baking apples, bake us apples, bake an apple*, and *bake us an apple* were reliably shorter than *bake* in *bake apples*, for all three participants in their study. This result is consistent with polysyllabic shortening either within metrically defined inter-stress intervals (from the stressed onset syllable in *bake* to the stressed onset of *apples*), or with polysyllabic shortening within word-based clitic groups (including the content word *bake* plus the function words *us* and/or *an*).

The same study reported additional results that help to resolve the question of word-based versus metrically based constituents as the domain of polysyllabic shortening. Comparison of the rhyme interval duration of *bake* in, for example, *bake elixirs* and *bake avocadoes* versus in, for example, *baking, bake us, bake an*, and *bake us an* showed that the rhyme interval of *bake* was significantly shorter in the latter, word-based constituents. This suggests that word-based clitic groups were more influential than metrically based cross-word feet (such as *bake e-* and *bake avo-*). Only one of the three participants showed shorter rhyme interval durations for the metrically based constituents. These results suggest a stronger role for word-based constituents as compared to cross-word feet containing word fragments. They put polysyllabic shortening primarily in the domain of word-based structure, rather than of structure based on cross-word feet, suggesting that the timing control system based on syllables, feet, and phrases proposed in AP/TD would need to be modified to accommodate word-based constituents close in size to the lexical word.

A separate set of findings provides a further challenge to the use of oscillators to implement these shortening effects. White and Turk (2010) and Turk and Shattuck-Hufnagel (2000) found that polysyllabic shortening does not occur in all contexts; instead it occurs more often in phrasally stressed words, where it is also greatest in magnitude. These results challenge an oscillator-based implementation system because they suggest that even in the same utterances, not all words are affected by the suprasegmental compression mechanism. This reduces the motivation for a periodicity-based compression mechanism. Instead, these results support the idea that rather than being a reflection of a periodic control mechanism tending toward surface isochrony, polysyllabic shortening may be one of a set of mechanisms that speakers use to signal the locations of word-based prosodic constituent boundaries in an utterance (Turk 2012).

This section has described some of the challenges to AP/TD's oscillator-based approach to modeling systematic variation in speech timing that are raised by its account of poly-subconstituent shortening. Additional challenges to other aspects of oscillator-based timing control are described in the following two sections, which address, respectively, (i) the assumption of coordination based on movement onsets and (ii) the assumption of duration lengthening using phonology-intrinsic clock-slowing gestures.

Challenges to AP/TD's Onset-Based Coordination, by Evidence of Coordination Based on Movement Endpoints

AP/TD's interplanning-oscillator-coupling approach to coordination is based on gesture onsets. In this approach, the relative timing of gesture onsets is controlled via stable entrainment relationships (in-phase and antiphase) of gestural coupling oscillators. The lengthening associated with prominence and boundaries is accounted for using π and μ_T gestures, which stretch the default gestural activation intervals with which they overlap. This approach is challenged by findings suggesting that coordination can be based on parts of movement other than the onset, notably the part of movement that is most "behaviourally meaningful,"often the endpoint (Shaffer 1982; Semjen 1992). For example, Perkell and Matthies (1992) found that lip protrusion timing in relation to voicing onset was much more consistent at the end of the protrusion movement, as compared to a point near the beginning of protrusion. (See also Leonard and Cummins 2011 for related results for speech-accompanying gestures; Gentner, Grudin, and Conway 1980 for typing; Bootsma and van Wieringen 1990 and Katsumata and Russell 2012 for hitting balls; and Billon, Semjen, and Stelmach 1996 and Spencer and Zelaznik 2003 for finger-tapping.)

These findings do not find an obvious account in AP/TD because it does not have a representation of the time of gestural target approximation, that is, of mass-spring settling time. This is because in AP/TD the time of target approximation is an emergent property of mass-spring (gestural) stiffness. Although the time of gestural target approximation/mass-spring gestural settling corresponds to a fixed proportion of a gestural planning oscillator cycle at a default speaking rate, this correspondence changes when a μ_T gesture stretches the gestural planning oscillator cycles at particular locations in an utterance, and at fast speaking rates, when gestural planning oscillator cycles are shortened. Findings of movement-endpoint–based coordination motivate alternative approaches that make it possible to refer to the (surface) timing of movement endpoints, and in which endpoint times serve as reference points for coordination (Shaffer 1982, Semjen 1992). For example, Lee's (1998, 2009) general tau theory proposes that endpoint coordination can be accomplished if movement tau, that is, the time remaining to movement endpoint attainment at the current movement rate, is kept in constant proportion either to the tau of another movement, or to the tau of an internally generated tau guide. This mechanism ensures synchronous endpoint achievement, as long as two movements (or a movement with the tau guide) are tau-coupled before the end of the movement.

In sum, evidence for movement coordination at goal-related parts of movements, e.g., endpoints, presents challenges to the onset-based coordination mechanisms of AP/TD. The second set of challenges is raised by AP/TD's use of a phonology-intrinsic clock rather than a general-purpose timekeeping mechanism.

Challenges to AP/TD's Phonology-Intrinsic Timing, by Evidence for General-Purpose Timekeeping Mechanisms

This section discusses the fact that AP/TD's account of prosodically governed duration lengthening relies on the slowing of a phonology-intrinsic clock by π/μ_T gestures and contrasts this mechanism with the use of general-purpose, phonology-extrinsic timekeeping mechanisms that meter out units of solar (i.e., surface) time. As noted herein and as described by Krivokapić, in AP/TD the lengthening associated with prominence and boundaries is accounted for using π and μ_T gestures, which stretch the default gestural activation intervals with which they overlap in time. This reliance on the adjustment of default timing is challenged by several sets of findings, including (i) findings showing that timing variability is greater for longer duration intervals, (ii) findings relating to prominence and boundaries suggestive of constraints on surface durations, and

(iii) the fact that π and μ_T gestures warp phonological time in relation to solar time, in a nonlinear manner across an utterance.

In relation to the first of these findings, many investigators have shown that timing variability for a movement grows with its duration. Results of this kind have been reported for nonspeech movements by Treisman (1963); Gibbon (1977); Schmidt et al. (1979); Rosenbaum and Patshnik (1980a, 1980b); Wing (1980); Hancock and Newell (1985); Wearden (1991); Ivry and Corcos (1993); Ivry and Hazeltine (1995); Spencer and Zelaznik (2003); Merchant et al. (2008); and Merchant, Zarco, and Prado (2008); see Malapani and Fairhurst (2002) for a review. For speech movements, see studies by Byrd and Saltzman (1998); Edwards, Beckman and Fletcher (1991); Remijsen and Gilley (2008); Chen (2006); Nakai et al. (2012); and Lefkowitz (2017).

These findings are challenging for AP/TD, because in that framework, activation intervals are stretched not by adding extra time to the activation intervals, but instead by a "clock slowing" mechanism. That is, the μ_T gestures slow the planning + suprasegmental ensemble clock during intervals overlapped by the μ_T gesture. In this approach, an interval that has been stretched by a μ_T gesture therefore has the same number of AP/TD planning + suprasegmental ensemble clock units as a corresponding interval that has not been stretched by a π or μ_T gesture. The π/μ_T approach therefore does not provide an account of the greater variability for longer surface duration intervals, because these longer surface duration intervals are not longer in AP/TD planning + suprasegmental ensemble clock time. The findings appear to instead require the representation of surface durations in solar time units, where longer duration intervals contain more surface time intervals (that is, they are longer in solar time) than shorter duration intervals. Greater variability of these longer duration intervals can be explained by a "noisy" timekeeper, in which variability correlates with interval duration (Gallistel 1999; Gallistel and Gibbon 2000; Jones and Wearden 2004; Shouval et al. 2014). It is not clear how AP/TD can account for these findings, because surface durations are only emergent in this approach, and are not represented.

With respect to the second point, findings suggestive of constraints on surface durations in prominent and boundary-adjacent positions provide another challenge to the use of π/μ_T gestures. For example, in Northern Finnish, a quantity language in which short versus long quantities are not signaled by vowel quality but by vowel duration, phonemically short vowels show smaller amounts of prominence- and boundary-related lengthening than phonemically long vowels do (Nakai et al. 2009; Nakai et al. 2012; Remijsen and Gilley 2008). These findings suggest that speakers of these languages avoid large amounts of lengthening on phonemically short vowels to maintain their surface durational distinction with phonemically long vowels. These findings are difficult to explain if (as in AP/TD) surface durations cannot be represented, because prominence- and boundary-related lengthening are emergent properties of default activation interval durations plus π/μ_T adjustments, without the possibility of representing the surface duration results as a goal.

With respect to the third point, AP/TD's slowing of the phonology-intrinsic clock in prosodically prominent and phrase-final positions, and at slow speaking rates, results in a lack of linear correspondence between AP/TD's planning + suprasegmental clock time and solar time. Because AP/TD does not make use of phonology-extrinsic timekeepers, this "time-warping" may have unintended and undesirable consequences. This may occur, for example, in cases where singers or speakers need to interact with external events, as when singing to instrumental music (where matching events in solar time is critical) or timing an oral presentation to end at a particular point in solar time.

Taken together, these findings present a challenge to AP/TD's use of phonology-intrinsic timing mechanisms and motivate the development of models in which speech timing is controlled instead by general-purpose timing mechanisms that are extrinsic to the phonology.

Conclusion

While AP/TD has successfully modeled many aspects of segmental coordination and prosody, as demonstrated in Krivokapić's essay, there are a number of lines of evidence that are difficult to reconcile with its coupled oscillator approach to coordination; (i) its default-adjustment approach to longer durations at prosodic boundaries and in prominent positions; and (ii) its system of syllable-, cross-word foot- and phrase-level oscillators. These observations highlight the importance of developing alternative models of speech production and testing the predictions of these models against predictions of AP/TD. Turk and Shattuck-Hufnagel (2020) provide additional evidence that supports phonology-extrinsic timing-based approaches. Considerable work will be required to bring such models to the state of experimentally tested implementation that has made the AP/TD approach so influential and widely discussed.

Note

1. Note that the term *cross-word foot* does not specify the type of stress foot, that is, the level of prominence that delimits the foot. For example, Abercrombie's (1991) cross-word feet were delimited by "chest pulses" and were envisioned to be delimited by phrasal prominences. In AP/TD, cross-word feet are defined on the basis of word-level stress. See also Williams and Hiller (1984), who performed tests of polysyllabic shortening within feet of different types.

References

Abercrombie, D. 1991. "Some Functions of Silent Stress." In *Fifty Years in Phonetics*, edited by D. Abercrombie. Edinburgh: Edinburgh University Press.

Barbosa, P. A. 2007. "From Syntax to Acoustic Duration: A Dynamical Model of Speech Rhythm Production." *Speech Communication* 49 (9): 725–742.

Billon, M., A. Semjen, and G. E. Stelmach. 1996. "The Timing Effects of Accent Production in Periodic Finger-Tapping Sequences." *Journal of Motor Behavior* 28 (3): 198–210.

Bootsma, R. J., and P. C. van Wieringen. 1990. "Timing an Attacking Forehand Drive in Table Tennis." *Journal of Experimental Psychology: Human Perception and Performance* 16 (1): 21–29.

Browman, C. P., and L. Goldstein. 1992. "Articulatory Phonology: An Overview." *Phonetica* 49 (3–4): 155–180.

Browman, C. P., and L. Goldstein. 1995. "Dynamics and Articulatory Phonology." In *Mind as Motion*, edited by R. F. Port and T. v. Gelder, 175–193. Cambridge, MA: MIT Press.

Byrd, D., and E. Saltzman. 1998. "Intragestural Dynamics of Multiple Prosodic Boundaries." *Journal of Phonetics* 26:173–199.

Campbell, W. N. 1988. "Foot-Level Shortening in the Spoken English Corpus." Paper presented at the Seventh Forensic Anthropology Society of Europe Symposium, Edinburgh, UK, September 1988.

Chen, Y. 2006. "Durational Adjustment under Corrective Focus in Standard Chinese." *Journal of Phonetics* 34 (2): 176–201.

Classe, A. 1939. *The Rhythm of English Prose*. Oxford: Basil Blackwell.

Dauer, R. M. 1983. "Stress-Timing and Syllable-Timing Re-Analysed." *Journal of Phonetics* 11:51–62.

Edwards, J., M. E. Beckman, and J. Fletcher. 1991. "The Articulatory Kinematics of Final Lengthening." *Journal of the Acoustical Society of America* 89 (1): 369–382.

Eriksson, A. 1991. "Aspects of Swedish Speech Rhythm." PhD diss., University of Gothenburg.

Fant, G., A. Kruckenberg, and L. Nord. 1991. "Stress Patterns and Rhythm in the Reading of Prose and Poetry with Analogies to Music Performance." Paper presented at the Music, Language, Speech, and Brain International WennerGren Symposium, Stockholm.

Fowler, C. A. 1977. *Timing Control in Speech Production*. Bloomington: Indiana University Linguistics Club.

Gallistel, C. R. 1999. "Can a Decay Process Explain the Timing of Conditioned Responses?" *Journal of the Experimental Analysis of Behavior* 71:264–271.

Gallistel, C. R., and J. Gibbon. 2000. "Time, Rate, and Conditioning." *Psychological Review* 107 (2): 289–344.

Gentner, D. R., J. Grudin, and E. Conway. 1980. *Finger Movements in Transcription Typing*. Technical Report 8001. San Diego: Center for Human Information Processing, University of California, San Diego.

Gibbon, J. 1977. "Scalar Expectancy Theory and Weber's Law in Animal Timing." *Psychological Review* 84 (3): 279–325.

Goldstein, L., D. Byrd, and E. Saltzman. 2006. "The Role of Vocal Tract Gestural Action Units in Understanding the Evolution of Phonology." In *Action to Language via the Mirror Neuron System*, edited by M. A. Arbib, 215–249. Cambridge: Cambridge University Press.

Hancock, P. A., and K. M. Newell. 1985. "The Movement Speed-Accuracy Relationship in Space-Time." In *Motor Behavior: Programming, Control, and Acquisition*, edited by Herbert Heuer, U. Kleinbeck, and K.-H. Schmidt, 153–185. Berlin: Springer-Verlag.

Ivry, R., and D. M. Corcos. 1993. "Slicing the Variability Pie: Component Analysis of Coordination and Motor Dysfunction." In *Variability and Motor Control*, edited by K. Newell and D. Corcos, 415–447. Champaign, IL: Human Kinetics Publishers.

Ivry, R. B., and R. E. Hazeltine. 1995. "Perception and Production of Temporal Intervals across a Range of Durations: Evidence for a Common Timing Mechanism." *Journal of Experimental Psychology—Human Perception and Performance* 21 (1): 3–18.

Jones, L. A., and J. H. Wearden. 2004. "Double Standards: Memory Loading in Temporal Reference Memory." *Quarterly Journal of Experimental Psychology* 57B (1): 55–77.

Katsumata, H., and D. M. Russell. 2012. "Prospective versus Predictive Control in Timing of Hitting a Falling Ball." *Experimental Brain Research* 216 (4): 499–514.

Kim, H. 2006. "Speech Rhythm in American English: A Corpus Study." PhD diss., University of Illinois at Urbana-Champaign.

Kim, H., and J. Cole. 2005. "The Stress Foot as a Unit of Planned Timing: Evidence from Shortening in the Prosodic Phrase." In *Proceedings of Interspeech/Eurospeech 2005*, 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, September 4–8, 2005; ISCA Archive. http://www.isca-speech.org/archive/interspeech_2005.

Lee, D. N. 1998. "Guiding Movement by Coupling Taus." *Ecological Psychology* 10 (3–4): 221–250.

Lee, D. N. 2009. "General Tau Theory: Evolution to Date." *Perception* 38:837–858.

Lefkowitz, L. M. 2017. "Maxent Harmonic Grammars and Phonetic Duration." PhD diss., UCLA.

Leonard, T., and F. Cummins. 2011. "The Temporal Relation between Beat Gestures and Speech." *Language and Cognitive Processes* 26 (10): 1457–1471.Malapani, C., and S. Fairhurst. 2002. "Scalar Timing in Animals and Humans." *Learning and Motivation* 33:156–176.

Merchant, H., W. Zarco, R. Bartolo, and L. Prado. 2008. "The Context of Temporal Processing Is Represented in the Multidimensional Relationships between Timing Tasks." *PLoS One* 3 (9): e3169.

Merchant, H., W. Zarco, and L. Prado. 2008. "Do We Have a Common Mechanism for Measuring Time in the Hundreds of Millisecond Range? Evidence from Multiple-Interval Timing Tasks." *Journal of Neurophysiology* 99:939–949.

Nakai, S., S. Kunnari, A. Turk, K. Suomi, K., and R. Ylitalo. 2009. "Utterance-Final Lengthening and Quantity in Northern Finnish." *Journal of Phonetics* 37:29–45.

Nakai, S., A. E. Turk, K. Suomi, S. C. Granlund, R. Ylitalo, and S. Kunnari,. 2012. "Quantity Constraints on the Temporal Implementation of Phrasal Prosody in Northern Finnish." *Journal of Phonetics* 40:796–807.

Nam, H., L. Goldstein, and E. Saltzman. 2010. "Self-Organization of Syllable Structure: A Coupled Oscillator Model." In Approaches to Phonological Complexity, edited by François Pellegrino, E. Marisco, and I. Chitoran, 99–328. Berlin: Mouton de Gruyter.

O'Dell, M. L., and T. Nieminen. 1999. "Coupled Oscillator Model of Speech Rhythm." In *Proceedings of the Fourteenth International Congress of Phonetic Sciences*, edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey, vol. 2, 1075–1078. College Park, MD: American Institute of Physics.

Perkell, J. S., and M. L. Matthies. 1992. "Temporal Measures of Anticipatory Labial Coarticulation for the Vowel /u/: Within-Subject and Cross-Subject Variability." *Journal of the Acoustical Society of America* 91 (5): 2911–2925.

Rakerd, B., W. Sennett, and C. A. Fowler. 1987. "Domain-Final Lengthening and Foot-Level Shortening in Spoken English." *Phonetica* 44:47–155.

Remijsen, B., and L. Gilley. 2008. "Why Are Three-Level Vowel Length Systems Rare? Insights from Dinka (Luanyjang Dialect)." *Journal of Phonetics* 36:318–344.

Rosenbaum, D. A., and O. Patashnik. 1980a. "A Mental Clock Setting Process Revealed by Reaction Times." In *Tutorials in Motor Behavior*, edited by G. E. Stelmach and J. Requin, 487–499. Amsterdam: North-Holland.

Rosenbaum, D. A., and O. Patashnik. 1980b. "Time to Time in the Human Motor System." In *Attention and Performance*, vol. 8, edited by R. S. Nickerson. Hillsdale, NJ: Erlbaum.

Saltzman, E. L., and K. G. Munhall. 1989. "A Dynamical Approach to Gestural Patterning in Speech Production." *Ecological Psychology* 1 (4): 333–382.

Saltzman, E., H. Nam, J. Krivokapić, and L. Goldstein. 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, 175–184.

Schmidt, R. A., H. Zelaznik, B. Hawkins, J. S. Frank, and J. T. Quinn. 1979. "Motor-Output Variability: A Theory for the Accuracy of Rapid Motor Acts." *Psychological Review* 86 (5): 415–451.

Semjen, A. 1992. "Determinants of Timing in Serial Movements." In *Time, Action and Cognition*, vol. 66, edited by F. Macar, V. Pouthas, and W. J. Friedman, 247–261. Berlin: Springer.

Shaffer, L. H. 1982. "Rhythm and Timing in Skill." *Psychological Review* 89 (2): 109–122.

Shattuck-Hufnagel, S., and A. Turk. 2011. "Durational Evidence for Word-Based vs. Prominence-Based Constituent Structure in Limerick Speech." In *Proceedings of the Seventeenth International Congress of Phonetic Sciences*. Hong Kong. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/index.htm.

Shouval, H. Z., M. G. Hussain Shuler, A. Agarwal, and J. P. Gavornik. 2014. "What Does Scalar Timing Tell Us about Neural Dynamics?" *Frontiers in Human Neuroscience* 8:438.

Spencer, R. M. C., and H. N. Zelaznik. 2003. "Weber (Slope) Analyses of Timing Variability in Tapping and Drawing Tasks." *Journal of Motor Behavior* 35 (4): 371–381.

Treisman, M. 1963. "Temporal Discrimination and the Indifference Interval: Implications for a Model of the Internal Clock." *Psychological Monographs* 77 (13): 1–31.

Turk, A. 2012. "The Temporal Implementation of Prosodic Structure." In *The Oxford Handbook of Laboratory Phonology*, edited by A. C. Cohn, C. Fougeron, and M. K. Huffman, 242–253. New York: Oxford University Press.

Turk, A. E., and S. Shattuck-Hufnagel. 2000. "Word-Boundary-Related Duration Patterns in English." *Journal of Phonetics* 28:397–440.

Turk, A., and S. Shattuck-Hufnagel. 2020. *Speech Timing: Implications for Theories of Phonology, Phonetics, and Speech Motor Control*. New York: Oxford University Press.

Van Lancker, D., J. Kreiman, and D. Bolinger. 1988. "Anticipatory Lengthening." *Journal of Phonetics* 61:339–347.

Wearden, J. H. 1991. "Do Humans Possess an Internal Clock with Scalar Timing Properties?" *Learning and Motivation* 22:59–93.

White, L. 2014. "Communicative Function and Prosodic Form in Speech Timing." *Speech Communication* 63–64:38–54.

White, L., and A. E. Turk. 2010. "English Words on the Procrustean Bed: Polysyllabic Shortening Reconsidered." *Journal of Phonetics* 38:459–471.

Williams, B., and S. M. Hiller. 1994. "The Question of Randomness in English Foot Timing: A Control Experiment." *Journal of Phonetics* 22:423–439.

Windmann, A., J. Šimko, and P. Wagner. 2015. "Polysyllabic Shortening and Word-Final Lengthening in English." In *Proceedings of Interspeech* 2015, 16th Annual Conference of the International Speech Communication Association, Dresden, Germany, September 6–10; ISCA Archive. http://www.isca-speech.org/archive/interspeech_2015.

Wing, A. M. 1980. "The Long and Short of Timing in Response Sequences." In *Tutorials in Motor Behavior*, edited by G. E. Stelmach and J. Requin, 469–484. Amsterdam: North-Holland.