ANNUAL REVIEWS

Annual Review of Linguistics

Cracking Prosody in Articulatory Phonology

2049

Dani Byrd[1] and Jelena Krivokapić[2,3]

[1]Department of Linguistics, University of Southern California, Los Angeles, California 90089, USA; email: dbyrd@usc.edu

[2]Department of Linguistics, University of Michigan, Ann Arbor, Michigan 48109, USA

[3]Haskins Laboratories, New Haven, Connecticut 06511, USA

Annu. Rev. Linguist. 2021. 7:31–53

The *Annual Review of Linguistics* is online at linguistics.annualreviews.org

https://doi.org/10.1146/annurev-linguistics-030920-050033

Keywords

Articulatory Phonology, prosody, dynamics, phrasal structure, prominence, tone

Abstract

Articulatory Phonology advances an account of phonological structure in which dynamically defined vocal tract tasks—gestures—are simultaneously and isomorphically units of cognitive representation and units of physical action. This paradigm has fundamentally altered our understanding of the linguistic representation of words. This article reviews the relatively recent incorporation of prosody into Articulatory Phonology. A capsule review of the Articulatory Phonology theoretical framework is presented, and the notions of phrasal and prominence organization are introduced as the key aspects of linguistic prosodic structure under consideration. Parameter dynamics, activation dynamics, and prosodic modulation gestures, such as the π-gesture, are outlined. The review is extended to touch on rhythm, intonation, and pauses and to consider innovations for integrating multiple aspects of prosodic structure under this dynamical approach. Finally, a range of questions emerges, crystallizing outstanding issues ranging from the abstract and theoretical to the interactive and functional.

1. A CAPSULE REVIEW OF ARTICULATORY PHONOLOGY

In the early 1980s, Cathe Browman, Louis Goldstein, and Elliot Saltzman found themselves together at Haskins Laboratories and mutually occupied with resolving a puzzle of spoken language. They saw that scholars of language had been pursuing, in parallel efforts, a cognitive representation of words as abstract, discrete symbolic entities and, simultaneously but without intersection, an account of spoken word realization in terms of characteristic quantitative acoustic properties. The chasm between the phonological and the physical had every appearance of arising from two incompatible descriptions of speech—an account of the mental lexicon that relied on an inventory of invariant recombinable symbols and an account of the speech signal—the consequence of realizing these symbolic representations in actual communication, requiring continuous time- and context-dependent variation of high dimensionality. Faced with this dissonance, Fowler, Browman, Goldstein, Saltzman, and like-minded others recognized that the dynamical descriptions being developed at that time to characterize a variety of ecological and psychological phenomena could potentially provide an avenue for resolving this duality (e.g., Fowler et al. 1980, Kelso et al. 1986, Saltzman & Munhall 1989). After all, why would a communication system evolve units that are shattered in the very process of being used for communication (Fowler 1980)?

Articulatory Phonology[1] was born as a theoretical framework for phonology that took seriously the notion that the combinatorial units composing words are simultaneously objects both of cognitive representation and of physical embodiment, and, as such, required no mediation module in the grammar to map between the two. To pursue this foundational assumption, scholars developing this framework conceived of these phonological primitives as defined dynamically and temporally overlapping and called these cognitive objects gestures.

Vocal tract constriction tasks, rather than movements of individual articulators, are the objects of cognitive control. At the ground level, gestures are modeled as critically damped point attractor systems defined in the space of vocal tract constriction tasks (Saltzman & Munhall 1989). These gestural tasks are realized when they drive the physical vocal tract plant according to the parameter values defined for their distinctive dynamical systems, such as their equilibrium position (constriction target) and stiffness (time to reach target) (Saltzman & Munhall 1989). The presence or absence of a particular gesture, in combination with the discrete gestural parameter values, forms the basis (or at least, a basis) for phonological contrasts, such as the traditional contrasts of place and manner and the distinction between vowels and consonants.

Crucially, when composed in words, gestural units are not synthesized as beads on a string but rather overlap in time, such that one gesture's activation can and will overlap with that of its neighbors (see **Figure 1** for example gestural scores) according to principled—and potentially language-specific—phonological systems of organization. Output articulation and consequent acoustics result from the combined demands of coactive gestural units over time.

Whereas a symbolic approach relies on abutting segments composed of symbolic features that are translated or mapped to an acoustic and/or articulatory implementation, a gestural approach relies on homomorphic units of information and action, similar in scale to traditional phonological features but with overlapping activations that wax and wane, defined in a vocal tract task space. The attractions of this approach are many, and the 1990s were an exciting period for illuminating the many types of phonological variation—such as assimilation, allophony, reduction

[1] This introduction is by no means intended to be a thorough review of Articulatory Phonology. For overviews, see Browman & Goldstein 1989, 1990, 1992, 1995; Goldstein et al. 2006; Gafos & Goldstein 2012; see also the Related Resource.

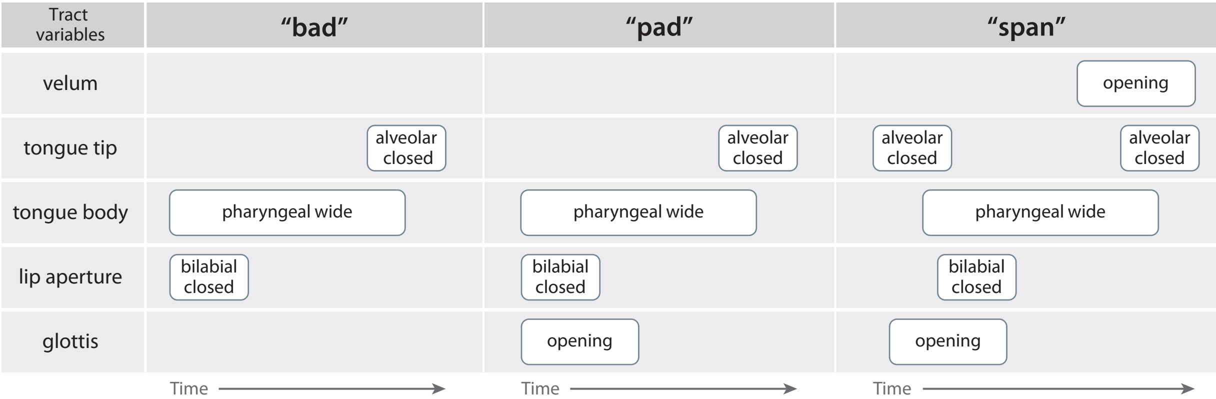

Figure 1

Three gestural scores—for the words "bad," "pad," and "span"—illustrating the notion of contrast within Articulatory Phonology as encoded in intergestural timing and the presence or absence of gestures. Boxes indicate gestural activation intervals (with parameter values of the gestures noted within the boxes); tract variables of the model on which gestures may be defined are shown on the left.

and deletion, and speech errors—that could be elegantly accounted for by postulating overlapping coactivation of subsegmental units accompanied by variation in their temporal coordination and magnitudes.[2]

Over the next decade, the Articulatory Phonology spotlight shone on studies of how gestures are coordinated, particularly in ways that reflect our traditional understanding of syllable structure. Gestures do not float freely but rather are organized via their coordination in time (L. Goldstein, personal communication). The naturally attractive modes of intergestural coordination (i.e., those available without learning) are understood to be in-phase (synchronous) and antiphase (sequential) coordination, and a deep understanding of phonological structure and variation leveraging these two basic modes of intergestural organization has been developed. Notions of gestural association implemented through dynamical coupling were developed and allowed for multiple pairwise coordination relations among gestural units to compete in a weighted way (Browman & Goldstein 2000, Nam & Saltzman 2003, Goldstein et al. 2006, Nam et al. 2009). Competition among pairwise gesture alignments was shown to yield a rich complexity of empirically observed coordination patterns from a constrained set of underlying coupling relationships. In this way, the dynamical systems approach was extended to syllable structure,[3] and much of this research has elegantly characterized crosslinguistic variation in syllable structuring (e.g., Gafos 2002; Goldstein et al. 2007a, 2009; Shaw et al. 2009, 2011; Pouplier & Beňuš 2011; Pouplier 2012; Shaw & Gafos 2015). Perhaps the best-explored phenomenon in this area is the C-Center effect, in which pairwise couplings between onset consonants with their tautosyllabic vowel yield an output organization in which the vowel initiates in the "center" of the onset consonant cluster (representative studies on

[2] There is an extensive literature on Articulatory Phonology accounts of segmental variability; for an overview, see Gafos & Goldstein (2012). Representative research includes publications by Browman & Goldstein (1990), Zsiga (2000), Gafos (2002), Chitoran et al. (2002), Goldstein et al. (2006, 2007b), Kochetov & So (2007), and Parrell & Narayanan (2018).

[3] Interestingly, a dynamical definition of multigesture structures at a segmental level of granularity—a segment-sized molecule—encountered more of a struggle in emerging, as attention was on the utility of the primitive gestural unit (though see Byrd 1996, Saltzman et al. 2000) and such a definition is still being developed (Tilsen 2016; see also Shih & Inkelas 2019).

the C-Center effect include Browman & Goldstein 1988; Byrd 1995; Honorof & Browman 1995; Browman & Goldstein 2000; Nam & Saltzman 2003; Kühnert et al. 2006; Nam 2007; Marin & Pouplier 2010; Hermes et al. 2013, 2017; Shaw & Gafos 2015).

In the current Articulatory Phonology framework, gestures are understood to be dynamic atomic units of encoding and production that are organized into larger structures, which are termed molecules (leveraging the chemistry analogy of binding). The conception of gestural molecules served to recognize the phonological vitality of groups of gestures that recur in many words, are systematically patterned, and are temporally cohesive.

At this point in the development of Articulatory Phonology, we broach the topic of this review—the incorporation of prosody into the dynamical framework of Articulatory Phonology beginning in the 2000s. Given the foundational understanding that words are composed of a small set of stable combinatory units, Articulatory Phonology tackled how to account for surface variability that indicates that the physical characteristics of those units are not fixed. The defining dynamical properties of gestural primitives offered an advantage and opportunity in this regard, as they rigorously accounted for surface variability in performance while assuming an underlying invariance at the level of control (e.g., Browman & Goldstein 1990, Saltzman 1995). For decades, this capacity has been explicitly deployed in Articulatory Phonology accounts of variation due to coarticulation and syllable position. Could it also be deployed to address variation due to the function and placement of words in larger informational structures?

2. DEFINING THE PROSODIC GROUND

Even within the terminologically rich domain of linguistics, the term prosody has referred to a wide variety of speech variation, ranging from affect to loudness, to fluency, to phrasing, and to prominence, including emphasis, contrast, and scope. For our purpose of considering how demonstrably important linguistic structure can be representationally and implementationally captured within Articulatory Phonology, herein we use the term prosody to refer to informational structure in language above the level of individual lexical entry. Prosody used in this sense refers to the grouping of information into meaningful phrases and to the prominence of particular words within those phrasal groups—as encoded by both the intonational (pitch) and durational (timing/rhythmic) properties of utterances. Consider how pitch, duration, and loudness serve to encode differences in meaning in saying these utterances:[4]

Prominence:	She didn't earn an A; she earned a B.
	She didn't earn an A; but she got one by cheating.
Phrasal organization:	When you make hollandaise slowly, it curdles.
	When you make hollandaise, slowly it curdles.

Prosody is critically important for conveying and understanding speech accurately, which is to say that the parity of the message between sender and receiver depends on far more than the individual words in sequence. Furthermore, prosodic structure differs from language to language and gives rise to language-specific signatures—systematic spatiotemporal variation—in the production of

[4]The tone excursion at or around these locations of prominence and boundaries (i.e., pitch accents, phrase accents, and boundary tones) can also have meaningful import. Consider, for example, the distinction among "You saw *Hamilton*." "You saw *Hamilton*?" and "You saw *Hamilton*!"

speech. The linguistic account of prosody requires an understanding of how human language represents or encodes this type of information and of how the structured variation in the actions of speaking physically embodies prosodic structure. Articulatory Phonology has made strides in extending its foundational assumptions regarding the dynamic nature of atomic cognitive units of language to a consideration of prosody. These developments are the topic of this review. For a more extensive review of this topic presented specifically for the specialist, we refer the reader to Krivokapić (2020); for other overviews of prosodic structure, see Fletcher (2010), Cole (2015), and Arvaniti (2020).

3. PROSODY AND PARAMETER DYNAMICS

Stable gestural representations form the basis for lexical contrast in Articulatory Phonology. This stability of word representation has two components: First, there exists a set of gestures, each with its stable internal activation gating pattern for the parameter values of its (point) attractor equation of motion, and second, there exists a stable intergestural organization of those units' activation functions with respect to one another. But words are situated in communicative utterances, and as articulatory tracking data grew accessible to the field, it became apparent that prosodic structure above the word level is associated with variation in these nominally fixed parameter values and/or activation functions. For example, at the juncture between words in different phrases—that is, at phrase edges or boundaries—gestures exhibit boundary-related articulatory lengthening, reduced temporal overlap, and spatial magnitude increases.[5] These effects scale, such that, for example, greater lengthening and greater pulling apart of gestures are observed at larger boundaries (e.g., Byrd & Saltzman 1998, Bombien et al. 2010, Byrd & Choi 2010, Hsieh 2017).

In order to address empirical observations of lengthening in the language of gestural dynamics, modulation of individual gestural parameter values was proposed by Beckman, Edwards, and colleagues (Edwards et al. 1991, Beckman et al. 1992, Beckman & Edwards 1994, Harrington et al. 1995). The seminal paper by Edwards et al. (1991, p. 378) highlights that "phrase final lengthening can be interpreted as an actual targeted slowing down at the phrase edge" and goes on to explicate that different prosodic sources of acoustic lengthening, such as prominence and boundaries, exhibit differing dynamic underpinnings. The proposal that the stiffness parameter value was lowered for gestures phrase-finally provided the first dynamical account of prosodic slowing in articulation. This advance into parameter dynamics—time-varying rather than fixed gestural stiffness values—directly altered a specific gestural parameter value in a specified prosodic environment (which was not itself specified dynamically). While successful at capturing kinematic properties of interest at prosodically significant locations, this approach also had limitations. An account of prosodic signatures relying on parameter dynamics had no means of unifying the spatial and temporal domains. That is, "picking out" a gesture's stiffness as the object of prosodic modulation lacked theoretical coherence—why that particular parameter; why that particular gesture; why a targeted prosodic position rather than a truly dynamical variation in time? Perhaps greater theoretical elegance and empirical advantages would be offered by an approach that could unify

[5] For lengthening, see, for example, Edwards et al. (1991), Beckman et al. (1992), Byrd & Saltzman (1998), Byrd (2000), Tabain (2003), Keating et al. (2004), Byrd et al. (2005, 2006), Cho (2006), Krivokapić (2007b), Bombien et al. (2013), Cho et al. (2014), and Katsika (2016). For spatial magnitude, see, for example, Fougeron & Keating (1997), Cho & Keating (2001), Fougeron (2001), Bombien et al. (2013), Georgeton & Fougeron (2014), and Georgeton et al. (2016). For reduced gestural overlap, see, for example, McClean (1973), Hardcastle (1985), Holst & Nolan (1995), Byrd & Saltzman (1998), Byrd (2000), Byrd et al. (2000), Bombien et al. (2006), and Byrd & Choi (2010).

the prosodic action across gestural control structures in prosodically sensitive regions in a truly dynamical realization.

The next step in the elaboration of Articulatory Phonology's consideration of prosody involved activation dynamics. Activation dynamics allows gestural activation trajectories to unfold in a flexible way that is sensitive to the prosodic environment, and offers the potential for simultaneously allowing the prosodic structure to itself be specified dynamically.

4. PROSODY AND ACTIVATION DYNAMICS

It was desirable both on independent scientific grounds and for the internal theoretical development and coherency of the Articulatory Phonology framework to advance an understanding of prosodic or information structure in a way consistent with the theory's dynamical model of lexical (word) structure. A successful step would be to unite phrase-final and phrase-initial lengthening and gestural overlap change at phrase junctures under the umbrella of a single mechanism of articulatory control dynamics. Byrd and colleagues (Byrd & Saltzman 1998, Byrd et al. 2000) proposed that the dynamical tool kit be extended to the level of gestural activation to account for this type of prosodic variability. Soon thereafter, Byrd & Saltzman (2003) presented an explicit computational model to do so. This research introduced the prosodic or π-gesture model.

The π-gesture model proposed that prosodic variability can emerge from the interaction of lexically specified dynamics of constriction gestures with prosodic gestures that represent phrase boundaries via time-varying modulation. Crucially, these π-gestures are not vocal tract constriction action gestures of any sort but rather are informational gestures that act vicariously on the constriction gestures with which they are coactive. For the first time, a dynamic gestural unit had no independent life as an action unit. And for the first time, prosody was associated with an independent compositional unit.

An alternative to this theoretical approach of encoding prominence and phrasing with locally acting abstract events is found in other linguistic accounts of prosody, especially in traditional grammatical approaches that are explicitly relational in their organization (as familiar, for example, from musical notation). These accounts offer their own rich insights into prosodic organization, particularly with regard to rhythm, prominence, and syntactic and dialogic structure (see, e.g., Wagner & Watson 2010, Cole 2015, Arvaniti 2020).

Within Articulatory Phonology, the π-gesture model views phrase boundaries as activation dynamics created by prosodic gestures that directly alter the time flow of gestural activation, thereby indirectly affecting vocal tract actions (for a thorough account, see Byrd & Saltzman 2003). Just as constriction gestures do, prosodic gestures have a discrete interval of activation, which waxes and wanes, and which must be coordinated in time with other gestures (**Figure 2*a***).

However, rather than representing an articulatory constriction event, this cognitive element acts to slow the time course of any constriction gestures that are "in play" or active during the scope or domain of activation of the prosodic gesture (**Figure 2*b***). The term transgestural is used to describe this type of dynamical effect across all concurrently active gestures; transgestural effects do not permit the "skipping" or "targeting" of specific gestures within the scope of the modulating gesture. This means that a phrase boundary is expected to affect all gestures coactive with the π-gesture. The prosodic slowing of the central clock pacing the unfolding of vocal tract constriction gestures means that gestures coactive with the π-gesture will start temporally later, in addition to exhibiting the gestural lengthening created by depressing the gating-in of the intrinsic stiffness parameter. Thus, as a consequence, the gestures become less overlapped. It may also be the case that a constriction will be detruncated (by the delay of an upcoming coarticulated gesture) and so exhibit less spatial undershoot, yielding a larger spatial magnitude. It is via this complex of

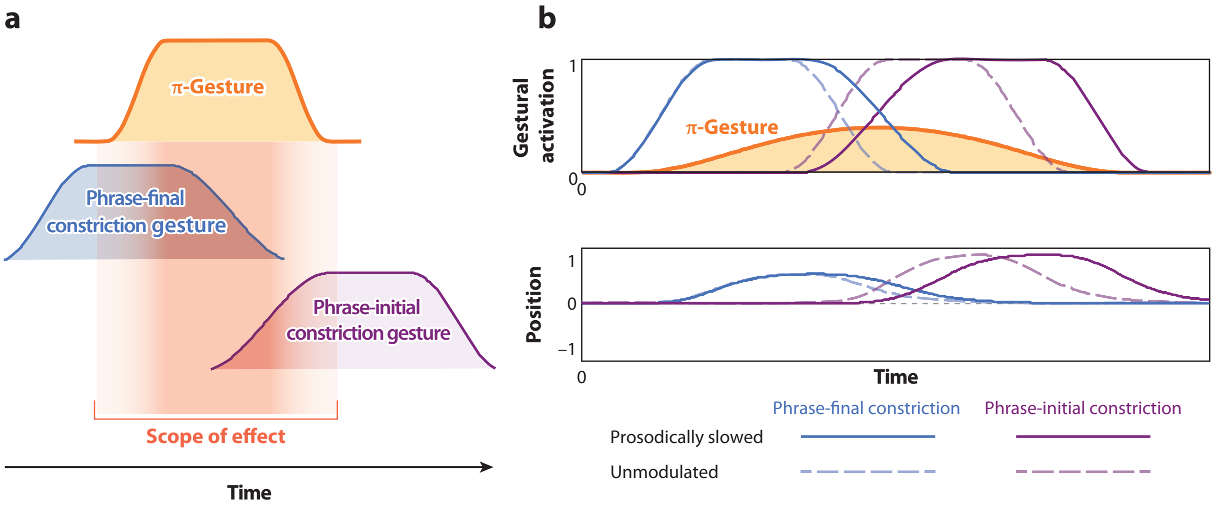

Figure 2

Two schemas illustrating Byrd & Saltzman's (2003) π-gesture model (inspired by their figures 3 and 6). (*a*) The arrangement of a phrase-final and a phrase-initial constriction gesture, spanned by a phrase boundary instantiated by a π-gesture. The scope of the π-gesture's effect is indicated as the orange shaded area. This modulation effect strengthens as the activation of the π-gesture waxes as the phrase edge approaches, and weakens as the π-gesture activation wanes as the phrase edge recedes. (*b*) Movement trajectories for two constriction gestures spanning a phrase boundary. The π-gesture is centered between the two constriction gestures, and the effect of its modulation on gestural activations, and consequently constriction position, is observed in the solid trajectories. Figure adapted with permission from Byrd & Saltzman (2003).

consequences that the multifaceted articulatory signatures of phrase boundaries are unified under the dynamical mechanism of the π-gesture.

Within the π-gesture model, boundaries are represented gradiently rather than categorically or symbolically. So, boundaries encoding greater disjuncture (for example, in traditional terms, at an Intonational Phrase versus an Intermediate Phrase) have greater activation and thereby exert a greater slowing or "stretching" of the concurrent gestural activations. The π-gesture model makes no distinction between final and initial phrase edges; a single π-gesture spans a boundary, reaching transgesturally into both final and initial phrase-edge material. (We return in Section 8, below, to the complicated issues of continuing investigation regarding one versus two π-gestures at a juncture and pausing.) Krivokapić (2020) notes that while there may be a "difference in the *amount* of lengthening between phrase-final and phrase-initial position [. . .due to the activation] shape of the π-gesture. . .there is no systematic, categorical, linguistically relevant difference between phrase-final and phrase-initial lengthening" (emphasis in original). Lastly, the more temporally distant a vocal tract constriction gesture is from a π-gesture's maximal activation, the smaller the influence the π-gesture will have on slowing its activation, predicting that, transgesturally, gestures closest to the boundary will lengthen more than those farther from it.[6] Since the inception of the π-gesture model, numerous empirical findings have supported its predictions regarding intragestural and intergestural timing at phrase boundaries. Predictions of the model regarding the symmetricality and temporal scope of phrase-boundary effects have been supported by experiments showing that gestures lengthen over a continuous stretch, that lengthening affects both

[6]Skewed π-gesture activation functions, as well as the attraction or extension of the π-gesture away from the phrase edge, remain poorly examined in the model.

phrase-final and phrase-initial gestures, that the lengthening effect decreases with distance from the boundary, and that gestures overlap less at boundaries (e.g., Byrd et al. 2005, 2006; Bombien et al. 2006; Byrd & Riggs 2008; Lee et al. 2006; Katsika 2012).

While the π-gesture model remains the only computationally implemented model of prosody (or of activation dynamics) within Articulatory Phonology, theoretical developments have continued to advance. Specifically, the development of a coupled oscillator planning model for choreographing constriction gestures in time and the initial integration of prominence (and rhythm) into an account of prosody have cooperated to lead to the next developments.

5. PROSODY AND MODULATION GESTURES

Saltzman et al. (2008) significantly extended the π-gesture model to capture prosodic events beyond boundaries, specifically those related to prominence. Previous research suggests some similarity between boundary and prominence effects, such that under prominence, gestures become larger and longer, and prominence effects extend over time. That said, in contrast to gestures at a boundary, gestures encoding prominence exhibit higher peak velocity than nonprominent gestures, and prominence effects, while occurring over an interval, are coordinated such that they are strongest at the nucleus of the stressed syllable of the prominent item (Beckman et al. 1992, Beckman & Edwards 1994, Cho 2006, Katsika 2016). This research suggests that the effects of boundaries and prominence could be modeled in a similar manner, albeit with their own particulars of coordination and transgestural modulation. Saltzman et al. (2008) postulated a temporal modulation gesture—rebranded a μ_T-gesture—that acts to modulate the gestural plan or score (e.g., **Figure 1**) at its creation rather than at its realization. While the π-gesture affects the gestural score once it is specified—that is, once words have been selected from the lexicon and composed into an utterance—the μ-gesture is incorporated into Articulatory Phonology's planning oscillator model (Saltzman et al. 2008) and acts to change the frequency parameters of the planning oscillators. Moreover, Saltzman et al. (2008) propose the utilization of two "flavors" of μ-gesture—a temporal μ-gesture (μ_T-gesture) and, more tentatively, a spatial μ-gesture (μ_S-gesture)—that modulate the utterance time flow or spatial target parameters, respectively, of coactive vocal tract constriction gestures. The μ_T-gesture has been used to model temporal effects of stress in American English (Saltzman et al. 2008). It has also been suggested that the μ_S-gesture can capture temporal and spatial effects of stress in Greek, with temporal effects being the secondary consequence of spatial effects (Katsika 2018).

The original π-gesture model applies to activation dynamics, locally time-warping a specific gestural score. In contrast, the μ-gesture approach modulates the ongoing values of the foot and syllable natural frequency parameters or the rate of phase flow (Saltzman et al. 2008) of the planning oscillators that govern the underlying pacing of an utterance. Despite this distinction, the μ-gesture carries on the primary foundational properties laid out for a dynamic prosodic gesture within Articulatory Phonology. It is active over a local discrete interval, it acts transgesturally, it must be coordinated in time, and it acts vicariously to modulate gestures that are concurrently active within its scope or domain.

While research to date on prosodic μ-gestures has focused on the μ_T-gesture (Saltzman et al. 2008, Katsika et al. 2014, Katsika 2016; but see Katsika 2018 for discussion of the role of the μ_S-gesture for modeling prominence), the possibility of a spatial modulation gesture, which presumably will exhibit distinct kinematic consequences from a temporal modulation gesture, offers the interesting possibility that μ_S-gestures could prove useful in capturing prominence in particular, given that constrictions under prominence exhibit kinematic properties (e.g., velocity profiles and patterns of overlap) different from those at boundaries. Furthermore, if there is more than one type of modulation gesture at work in prosody, the coordination patterning of μ_S-gestures could also differ from that of μ_T-gestures, perhaps capturing a coordination of the former with a lexically stressed syllable (along the lines of Saltzman et al. 2008, Katsika et al. 2014), while boundary gestures are anchored to a phrase edge in coordination with final/initial constriction gestures [and seemingly also coordinated with a nearby lexical stress in many languages (Katsika 2016)]. Indeed, from a cognitive or functional perspective there is no reason to suppose that realizing informational grouping (boundaries) must share the same cognitive underpinnings as realizing prominence.

6. GESTURAL MODULATION AND RHYTHM

In order to grasp how gestural modulation may encode informational prominence and phrasing in Articulatory Phonology, a consideration of linguistic rhythm is warranted. Despite decades of linguistic study, the definition and properties of linguistic rhythm remain disputed (e.g., Arvaniti 2012, Tilsen & Arvaniti 2013). The traditional crosslinguistic division into stress- and syllable-timed languages with the concomitant properties of foot isochrony and syllable isochrony is well known to be inaccurate (e.g., Dauer 1983; review in Arvaniti 2012). Yet, there is evidence that a tendency toward foot isochrony in American English exists, such as polysyllabic shortening, whereby stressed syllables shorten when unstressed syllables are added to the foot (Lehiste 1972, Klatt 1973, Port 1981, Rakerd et al. 1987). Specifically, Kim & Cole (2005) find that stressed but not unstressed syllables within a foot shorten with the addition of unstressed syllables to the foot. To model these findings with the gestural modulation approach, Saltzman et al. (2008) generalize the planning oscillator model, which has been used to drive gestural timing, to include a foot and a syllable planning oscillator that are bidirectionally coupled to each other (see also O'Dell & Nieminen 1999 and Barbosa 2007 for such an approach). Each oscillator has a weighting that determines how strong its influence is. In a stress-timed language, the foot oscillator will be more strongly weighted; thus, the foot oscillator will be more dominant. This will lead to the shortening of syllables within the foot (i.e., polysyllabic shortening) and thus toward foot isochrony. The other salient property of the foot, namely that its stressed syllable is longer than its unstressed syllables, is due to the μ_T-gesture, which, as discussed above, has the effect of lengthening the gestures coactive with it. Finally, the effect observed by Kim & Cole (2005), among others, that only stressed syllables, not unstressed ones, shorten in a foot is accounted for by adjusting the weight of the foot and syllable oscillators. While the foot oscillator's weight is stronger than the syllable oscillator's weight during the stressed syllable, it is weaker during the unstressed syllables. Consequently, only the stressed syllable will be affected by the "squeezing" driven by the foot oscillator.[7]

One pressing point, both for an Articulatory Phonology approach to prosodic structure and for prosodic research in general, is that there is an insufficient understanding of the kinematic properties of stress. Until diverse high-quality crosslinguistic articulatory data are acquired and kinematics better understood, it will not be possible to determine whether a μ_T-gesture or a μ_S-gesture provides a better model of stress. Similarly, without knowing the kinematic properties of

[7]It has also been suggested (Krivokapić 2013) that convergence between speakers of American and Indian English can arise as a result of the weighting of the oscillators changing, such that for speakers of American English the foot oscillator becomes less dominant in the course of an interaction with Indian English speakers. In general, changing of the oscillator weighting could be a way to model crosslinguistic differences, to account for the continuum of stress-timedness across languages.

the constrictions spanning a foot and how these interact with phrase boundaries, it will not be possible to definitively model the foot within Articulatory Phonology.

With these caveats in mind, as they are currently understood, the μ_T-gesture and the foot and syllable oscillators collaborate to account for the rhythmic proclivities of American English. The introduction of the foot and syllable planning oscillators also gives rise to the possibility of modeling the prosodic hierarchy as arising from nested planning oscillators. We return to this topic in Section 10, below.

7. INTONATION AND GESTURAL REPRESENTATION

In addition to the temporal phenomenon of boundary-related lengthening, prosodic structure is manifested via tonal properties and acoustic pauses. Tonal properties include marking of pitch accents and marking of phrase edges (for example, at Intermediate and Intonational Phrase boundaries in English). Analyzing kinematic and acoustic data, D'Imperio et al. (2007) were the first to demonstrate that intonational gestural events might be coordinated with constriction events rather than acoustic landmarks, and later research has provided further evidence of such coordination (Niemann et al. 2011, Mücke et al. 2012). This vantage point has opened the door for considering tonal attraction/crowding/repulsion in terms of (competing) coupling relations among the tonal (and constriction) gestures, and prosodic gestures are likely to have a role to play here as well.

Let us step back for a moment and consider the gestural representation of tone. Empirical and modeling research has shown that lexically specified tones can be understood as gestures (Gao 2008, Karlin & Tilsen 2015, Hu 2016, Lee 2018, Zhang et al. 2019; see also Ladd 2006), generally understood to be H and L primitives. Gestures have as their task linguistically relevant variations in vocal fold vibration frequency. Like constriction gestures, they are coordinated with the other gestures of the lexical representation via existing modes of coupling, typically in-phase and antiphase, and like consonant and vowel gestures, they can be affected by competition among the coupling relationships (Gao 2008, Hsieh 2017, Karlin & Tilsen 2015, Hu 2016). The representation of lexical tone in gestural terms will require that these tonal task gestures (whether defined articulatorily or acoustically) necessarily be coordinated with prosodic gestures. This modeling remains almost wholly unexplored (but see Lee 2018, Lee et al. 2018b, and Oh & Byrd 2019 for initial thoughts).

The gestural approach to lexical tones has been extended to intonation gestures, such as pitch accents, accentual phrases, and boundary tones. The general spirit of this approach is initially laid out in the segmental anchoring hypothesis as first developed in the context speech rate variability by Ladd et al. (1999) and Prieto & Torreira (2007), and by D'Imperio et al.'s seminal research integrating the potential of articulatory anchoring for phrase-edge tones (e.g., D'Imperio et al. 2007), advanced by Ladd (2006) for pitch broadly. Specifically, this research line probes the possibility that gestural tonal primitives for prosodic information encoding are coordinated or coupled with constrictions at prominent lexical items and phrasal junctures. For examples of such research, see Mücke et al. (2012) and Niemann et al. (2011) for pitch accents, Katsika et al. (2014) for boundary tone, Niemann (2016) for phrase accents, and Lee (2018) and Lee et al. (2018b) for accentual phrases.

One interesting finding from this line of research is that tone gestures in lexical representations and tone gestures deployed postlexically as pitch accents exhibit different coordination patterns. While lexical tones affect the coordination of consonants and vowels of a syllable (e.g., Gao 2008, Zhang et al. 2019), pitch accents do not appear to, as would be consistent with their postlexical status (Mücke et al. 2012). The advantage of the Articulatory Phonology framework is that the coordination among tone, consonant, and vowel gestures can be specified precisely, using linguistically relevant kinematic landmarks (such as gesture onsets) rather than requiring reference to somewhat approximate acoustic landmarks (see Ladd 2006 for issues arising when acoustic landmarks are used as anchoring sites for tones). Initial results also suggest that crosslinguistic variation might be accounted for by different ways in which intonation gestures are coupled to one another and/or to constriction gestures (Niemann et al. 2011, Mücke et al. 2012). Lastly, note that the incorporation of intonation into Articulatory Phonology via the integration of tonal gestures into the gestural score offers the advantage of having all the major aspects of both lexical and prosodic structure modeled as gestures—speech production gestural tasks and the modulation of those tasks via informational tasks. Critically, the linguistic structuring of languages is represented such that all components of prosody and of lexical items have a discrete activation interval coordinated within an utterance.

The dynamical tool kit, as cogently reviewed by Roessig et al. (2019) and Roessig & Mücke (2019), additionally offers the utility of phase transitions between attractors to account for prosodic (and other) speech variability. Such transitions can emerge as control parameters vary and have been postulated to account for resyllabification (de Jong 2001), gestural intrusion in speech errors (Goldstein et al. 2007b), and phonological alternations (e.g., Parrell 2012). The dynamical tool kit has been extended to the phonological grammatical level by Gafos and colleagues (e.g., Gafos & Beňuš 2006) by deploying shifts between attractors as a function of biasing parameters to account for both categorical and continuous phonological variation. This dynamical view of phonological grammar is extended by Lee (2018) and Lee et al. (2018b) to model the interaction of register tone and phrase accent and by Nava & Tepperman (2010) to model an interplay between rhythmic variability and relative prominence placement. Nava & Tepperman (2010) propose that the native versus nonnative (Spanish-L1) prosodic patterns in producing English can be modeled as discrete modes of a dynamical system by manipulating the control parameter of a tilted anharmonic oscillator (Tuller et al. 1994) to bias attractor selection and attractor values. Most recently, while not specifically integrating articulatory constriction gestures as deployed in Articulatory Phonology, Roessig et al. (2019) and Roessig & Mücke (2019) present an elegant dynamical systems approach to what can be understood as pitch accent timing variation. Roessig & Mücke (2019) account for pitch accent timing differences observed among broad, narrow, and contrastive focus categories via a single underlying dynamical mechanism, observing that "prosodic dimensions contribute differently to the complex shape of the compositional attractor landscape and respond differently to the scaling of the system."

8. PAUSES AND THE GESTURAL SCORE

Phrases do not always seamlessly transition from one phrase to the next. Thus, an Articulatory Phonology account of phrase boundaries must understand how such a seam or pause occurs between phrases. Early in the account of prosodic signatures in articulation, Byrd & Saltzman (2003) speculated as to whether a π-gesture's action might account for articulatory properties during pauses. They allowed for the possibility that two π-gestures might be present in such circumstances:

> . . .one delimiting the end of the first informational unit and one initiating the next?. . . [I]t is possible that a cessation or initiation of speaking requires the stopping or starting of the clock. Such stops and starts are typically coincident with the edges of informational units and, therefore, would typically be accompanied by a π-gesture under our account. (Byrd & Saltzman 2003, p. 176)

At first blush, it might seem that an overt clock-starting gesture would be required to initiate an utterance, along the lines stated in the above quotation. But within the planning oscillator framework, starting an utterance is done by selecting the gestures (with their intrinsic parameters) that are linked to the appropriate planning oscillators in the ensemble (whose coupling graph reflects the lexical entry); then, the planning ensemble is allowed to reach its steady state, after which articulation is governed by the triggering of each gesture's activation wave with reference to the ongoing activity of the ensemble—speaking starts without an explicit clock-starting gesture.

Recently, researchers have begun to investigate articulatory kinematic patterns occurring during acoustic pauses and have found systematic differences in behavior for different types of pauses, such as pauses at prosodic boundaries in comparison to pauses during resting positions (e.g., Ramanarayanan et al. 2009). In particular, there is some evidence of targeted, controlled articulation during pauses, termed pause postures (Katsika et al. 2014, Krivokapić et al. 2020; see also Gick et al. 2004, Tilsen et al. 2016, Rasskazova et al. 2018, 2019). It is unclear whether this distinct postural appearance is a consequence of an independently active gesture or the result of a neutral attractor pulling the articulators back toward their neutral position (see Krivokapić et al. 2020). If a pause posture is created by a separate gesture, we would expect this gesture to have a specific function. For example, a pause gesture might provide time for a speaker's planning of an utterance, or it could signal to a listener an upcoming long utterance or indicate a type of utterance, such as a rhetorical question. In contrast, if a pause posture is the consequence of "exposing" the neutral attractor, there is no expectation of specific function—while the exposed gesture could yield planning time and/or be informative to a listener, the only necessary goal of the gesture would be that of returning the articulators to their default articulatory setting or posture. Research on the existence, properties, and functions of pause postures is an area of immediate interest among speech production researchers, and evidence to date points to a systematic relation between pausing and the prosodic gestures understood to characterize phrasal junctures and prominence. We turn next to the integration of the various prosodic gestures.

9. AN INTEGRATED PROSODIC GESTURAL APPROACH

Recently, Katsika and colleagues (Katsika 2012, Katsika et al. 2014) have elaborated a prosodic gestural approach within the Articulatory Phonology framework that integrates the modeling of boundaries as π-gestures, prominence as μ-gestures, and pause postures. Katsika and colleagues conceive of a set of prosodic gestures that interact in a dynamic way to give rise to the temporal and tonal properties that characterize boundaries. Within this approach, the π-gesture has a dual coordination: It is coordinated to the phrase-final vowel constriction gesture and weakly to the μ_T-gesture associated with stress, resulting in a shift of the phrasal π-gesture toward the stressed syllable when stress occurs earlier in the word (**Figure 3*a***) in comparison to when it occurs later (**Figure 3*b***). This proposal accounts for findings showing that lengthening starts earlier in words with earlier stress for Greek (Katsika et al. 2014), as well as in a number of other languages (Katsika 2016).[8]

An additional feature of the Katsika model is that a phrasal tone gesture at a boundary will be triggered when the π-gesture reaches a designated level of activation. Since the π-gesture shifts toward a stressed syllable, this means that the threshold for boundary tone activation will

[8]It might also account for the shortening found on the stressed syllable in phrase-final words compared with phrase-medial words. Katsika (2016) suggests that the shortening is due to the μ_T-gesture being pulled slightly toward the π-gesture, thus reducing the amount of lengthening of the stressed syllable in phrase-final compared with phrase-medial words (see also Kim et al. 2017 for a similar shortening effect in American English).

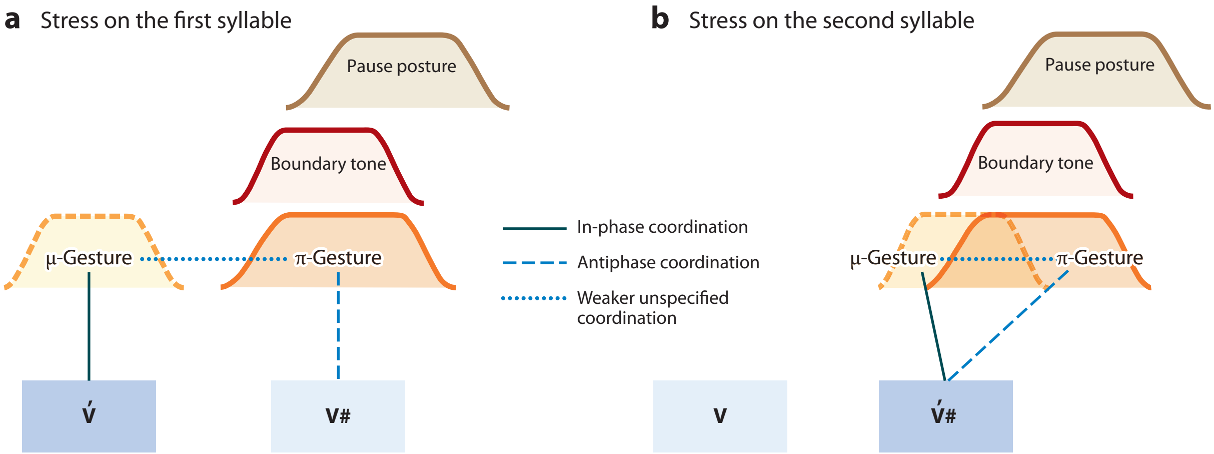

Figure 3

A schematic representation of boundary events, as conjectured by Katsika et al. (2014), for words with stress (*a*) on the first syllable and (*b*) on the second syllable. (Only the parts of the model described in the text are represented.) V́ indicates stressed vowels; lines indicate coordination between gestures. Figure adapted with permission from Katsika et al. (2014).

be reached earlier (and, thus, the boundary tone will start earlier) when stress is earlier in the word (as in **Figure 3*a***). This corresponds to Katsika et al.'s (2014) findings for Greek. Finally, the pause posture identified by Katsika et al. is triggered by an even stronger activation of a phrasal π-gesture. The predictions of this account are that boundary tones should occur only at boundaries with lengthening and that pause postures should occur only at very strong boundaries—namely those that have both lengthening and boundary tones.

The Katsika model provides a comprehensive account of temporal and tonal properties at boundaries and makes specific predictions about the interactions among a set of prosodic gestures. The model introduces two mechanisms: (*a*) the idea that there are different degrees of coupling strength among prosodic elements—an idea also explored and computationally implemented by Mücke et al. (2019) to examine differences in pitch accent coordination across languages, and (*b*) the idea that a sufficient level of activation of a π-gesture can trigger another (prosodic) gesture. This is a novel mechanism of gesture coordination within Articulatory Phonology and one that still needs to be modeled and better understood.

Some aspects of Katsika et al.'s account remain tentative and have not yet been computationally implemented. However, the integration of multiple prosodic gestures is an exciting development for Articulatory Phonology as an approach that integrates both tonal and temporal aspects of prosodic structure. This integrated account provides an avenue for modeling effects of prominence on boundary-related lengthening and for increased clarity in the temporal triggering of a π-gesture's onset. Future research will need to test the explicit predictions the account makes, as well as explore how typologically different languages do or do not necessitate innovations.

10. OUTSTANDING FRONTIERS

With the substantial progress made in incorporating prosody into the Articulatory Phonology framework, a range of questions has emerged, crystallizing outstanding issues. In large measure, these fall at the extremes of the spectrum ranging from the abstract and theoretical to the communicative and functional. We outline a subset of these food-for-thought items below.

The prosodic gesture model of Byrd and Saltzman introduced a novel way of understanding the prosodic hierarchy, namely as independent cognitive (or phonological) events (e.g., π-gestures) of varying strength. This boundary strength—in the form of π-gesture activation level—is, at least in potential, gradient (though perhaps it would not be expected to be smoothly so). Evidence for the view that, fundamentally, prosodic categories do not differ qualitatively but rather quantitatively is outlined by, for instance, Wagner (2005), Krivokapić & Byrd (2012), Beňuš & Šimko (2014), and Krivokapić (2020). More recent developments of foot and syllable planning oscillators (e.g., O'Dell & Nieminen 1999, Saltzman et al. 2008) allow for the possibility of modeling a prosodic hierarchy as arising from nested planning oscillators, as also suggested by Tilsen (2019). Note, though, that under neither approach does the prosodic hierarchy consist of qualitatively different symbolic categories (such as Intonation Phrase or Intermediate Phrase). Under the π-gesture model, boundaries of different strengths arise through varying activation strength (either smoothly distributed or clustered), whereas the nested oscillator approach deploys oscillators of varying frequency. These are two fundamentally different views of the prosodic hierarchy, though both leverage the dynamical tool kit. Theoretical, empirical, and computational research will be necessary to make a principled decision as to which account is superior.

Moreover, given that foot and syllable oscillators seem by their very definition to intrinsically capture rhythmic (e.g., stress and moraic) properties of lexical items and presumably of lexical items composed into longer utterances, it is incumbent on Articulatory Phonology to explicate an account of phonological rhythm in these terms. Saltzman et al. (2008) have approached the question of isochrony and compensatory shortening in this way. But a typological approach to crosslinguistic rhythm from a coupled oscillatory perspective remains remote (though see Port et al. 1995, Port 2003, Nam 2007, and Tilsen 2019, as well as research in the domain of music, e.g., Large et al. 2002 and Large 2008). As described above, modulation of these same planning oscillators has been proposed to account for prosodic structure; thus, an interaction of rhythm and prosody would seem inevitable. Katsika and colleagues (Katsika 2012, Katsika et al. 2014) have outlined a view that modulation gestures associated with lexical stress and those associated with boundaries are explicitly coupled with one another (see also Krivokapić 2007b, Byrd & Riggs 2008, Cho et al. 2016). Articulatory Phonology will ultimately need to provide both theoretical and computation integration of prosodic boundaries and prominence (and their associated pitch/phrase accents/tones) with rhythm that can successfully yield a typology of metrical structure and account for the empirically observed interactions of these linguistic phenomena.

At the functional end of the prosody spectrum lies the application of the dynamical approach to interactive speech. To date, most prosody research that has been able to provide the high-quality kinematic data serving to move Articulatory Phonology forward has been drawn from solo-speaker, read speech. Scholars developing the frontiers of Articulatory Phonology have shown an interest in leveraging innovative novel instrumental techniques such as real-time magnetic resonance imaging (rtMRI) (Byrd et al. 2009; Ramanarayanan et al. 2009, 2013; Toutios et al. 2016, 2019; Oh & Lee 2018; Parrell & Narayanan 2018; Chen et al. 2019; Hagedorn et al. 2019; Lim et al. 2019) and in the larger communicative circumstances of spoken language, particularly in (nonread) conversational interactions, to which we turn next.

While some functions of prosodic structure have been well examined from various psycholinguistic perspectives—such as the relationship between intonation and meaning for prominence and the role of prosodic phrasing in syntactic disambiguation—Articulatory Phonology has barely begun to incorporate this work on speech planning and communicative function. One line of research has begun to investigate the role of prosodic phrases in planning, suggesting that phrases chunk an upcoming utterance into planning units for phonological and phonetic encoding

(Keating & Shattuck-Hufnagel 2002; Krivokapić 2007a, 2012). The size of these prosodic phrases is flexible and depends on various factors, many of which are related to cognitive load, but it also depends on the working memory of the speaker (Ferreira & Swets 2002; Krivokapić 2007a, 2012; Swets et al. 2007, 2017; Bishop 2016b). Relatedly, it remains an open question as to what the function of articulatory patterns during pauses (Ramanarayanan et al. 2013) and at prosodic boundaries may be. It has been suggested that pause postures are related to the planning of an upcoming utterance (Krivokapić et al. 2020) or to breathing (Rasskazova et al. 2019), but this research has only just begun. Understanding the functional interplay between prosodic structure and cognitive and planning constraints will help us come closer to answering some of the fundamental questions of prosodic structure, such as the nature of the prosodic hierarchy. At this point, however, Articulatory Phonology as a theory of linguistic representation makes essentially no contact with higher-level planning and information structuring. While it may be that this is beyond the scope of Articulatory Phonology, the most desirable state of affairs would be for the dynamical representations deployed in Articulatory Phonology to be able to interact relatively seamlessly with other cognitive processes active in shaping spoken-language communication.

Individual differences also exist in the phonetic manifestations of prosodic boundaries and prominence and in their perception, and have recently been the subject of systematic investigation (e.g., Cangemi et al. 2015; Bishop 2016a,b; Grice et al. 2017; Roy et al. 2017; Kim 2020). While the patterning of these differences has been argued to be evidence for an approach in which phonological and phonetic aspects of phrase boundaries are understood as isomorphic (Grice et al. 2017), systematic individual differences in prosodic organization have received relatively little attention within the Articulatory Phonology framework.

Lastly, in contemplating the realm of spoken-language communication behaviors as considered within Articulatory Phonology (or highly compatible approaches), innovative first steps include kinematic research on manual cospeech actions (Danner 2017), examination of cospeech movement of brow and head at floor exchanges (Danner et al. 2019), and examination of the relationship between deictic gestures and prosodic structure (Parrell et al. 2014, Krivokapić et al. 2016, 2017); see Wagner et al. (2014) for an overview of the relationship between cospeech gestures and prosody. In addition to cospeech movements, fascinating research has recently provided evidence for prosodic convergence between individuals engaged in synchronous speech or in conversation. Krivokapić (2013) and Lee et al. (2018a) have shown convergence in prosodic parameterization at prosodically sensitive positions, and Reichel et al. (2018) have shown convergence in global time-varying patterns associated with pitch between like dialogue acts (see also Reichel & Cole 2016 on categorical pitch accent convergence as well as Tiede & Mooshammer 2013 and Vatikiotis-Bateson et al. 2014 for other studies of articulatory convergence). These types of empirical findings suggest that cognitive primitives of linguistic representation can be active in interindividual convergence processes and that convergence research may provide a potential probe of atomic gestural units, both lexical and prosodic. A dynamical account of prosodic convergence may ultimately deploy notions of cross-speaker alignment or entrainment of prosodic gestures, but how exactly this "dynamical attraction" will play out in terms of interaction of oscillatory units at long timescales across individuals is wholly unclear.

In summary, prosody must be modeled with an eye to its social context. In natural interactive communication between speakers, elements falling under our admittedly narrow definitional umbrella of prosody—phrasing and prominence—most certainly interact with the functional aspects of a more encompassing consideration of prosody, such as tempo, affect, speech planning, stylistics, social dynamics, and conversational structuring. And within the larger field of cognitive science, we expect future research on prosody within Articulatory Phonology to expand the coherency between Articulatory Phonology as a dynamical framework for language and dynamical understandings of other cognitive domains, such as music, memory, vision, and systems neuroscience, such as in the work of E. Chang and colleagues (including Bouchard et al. 2013, Tang et al. 2017, Chartier et al. 2018).

11. SUMMARY

Over the last three decades, the revolutionary tenets of Articulatory Phonology have become increasingly mainstream in theoretical phonology. Browman, Goldstein, Saltzman, and their colleagues have unequivocally demonstrated that linguistic representations that provide for the incorporation of time, coordination, and dynamically defined abstract tasks can illuminate and unify a wide variety of phonological patterning across diverse languages. Structure above the gestural level, such as syllable structure and, to a lesser degree, stress and rhythm, became increasingly integrated into the dynamical conception of lexical representation. However, information encoding above the word level is a more recent arrival to the party.

This article has reviewed the incorporation of prosodic structure within the Articulatory Phonology framework. Under this dynamical systems approach, prosodic structure arises from a small set of abstract informational prosodic gestures, as conceived by Byrd, Saltzman, and colleagues, that capture the grouping of lexical items into informational units and the prominence of individual items within those groups. These prosodic gestures are coordinated with the (likewise limited) set of phonological gestures composing the words of an utterance and serve to modulate those phonological primitives and/or add intonational actions, which, in turn, are also coordinated with prosodic and lexical actions. The prosodic gestures that have been postulated over the past two decades include π-gestures, tone gestures, μ-gestures, and pause postures, in interaction with syllable- and foot-level rhythmic oscillators. Some of these abstract prosodic entities, such as π-gestures, have been computationally modeled within the Task Dynamics framework, but a great deal of further computational research remains in order for Articulatory Phonology to integrate a wider array of prosodic gestures into its simulation capability.

Phrasal structure and prominence structure, requisite in composing words into utterances, were first captured in Articulatory Phonology's theoretical and dynamical framework in the early 2000s. Phrasal boundaries were conceived of as gestural in nature, parallel to the way constriction events are understood (Byrd & Saltzman 1998, 2003; Byrd et al. 2000). The π-gesture model postulated gestures that extend in time and act transgesturally to modulate the activation dynamics of coproduced vocal tract constriction gestures. π-Gestures warp the unfolding of articulation in the vicinity of a phrase boundary. No longer is the gestural score a player-piano roll (à la Saltzman & Munhall 1989); rather, the fabric of an utterance is elastic—stretchable or compressible. This model made a number of testable hypotheses that, over the years, have generated a substantial body of work, which has largely supported its predictions and led to further developments of an account of prosody leveraging the incorporation of dynamics into phonological representation. The main further developments of the model have included the instantiation of tone gestures, the inclusion of a foot and syllable oscillator, the introduction of spatial and temporal modulation gestures, and a consideration of how the properties of prosodic boundaries may emerge via the coordination of a set of prosodic gestural events or cognitive primitives.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the National Institutes of Health (grant DC003172 to D.B.) and the National Science Foundation (grant 1551513 to J.K.). We thank Louis Goldstein, Cynthia Lee, and Elliot Saltzman for their time and fruitful conversations; the organizers and participants of the Workshop on Articulatory Phonology at Asilomar, July 2019, for their invigorating discussion; and an anonymous reviewer.

LITERATURE CITED

Arvaniti A. 2012. The usefulness of metrics in the quantification of speech rhythm. *J. Phon.* 40:351–73

Arvaniti A. 2020. The autosegmental metrical model of intonational phonology. In *Prosodic Theory and Practice*, ed. S Shattuck-Hufnagel, J Barnes. Cambridge, MA: MIT Press. In press

Barbosa PA. 2007. From syntax to acoustic duration: a dynamical model of speech rhythm production. *Speech Commun*. 49:725–42

Beckman ME, Edwards J. 1994. Articulatory evidence for differentiating stress categories. In *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, ed. PA Keating, pp. 7–33. Cambridge, UK: Cambridge Univ. Press

Beckman ME, Edwards J, Fletcher J. 1992. Prosodic structure and tempo in a sonority model of articulatory dynamics. In *Papers in Laboratory Phonology II: Segment, Gesture, Prosody*, ed. GJ Docherty, DR Ladd, pp. 68–86. Cambridge, UK: Cambridge Univ. Press

Beňuš Š, Šimko J. 2014. Emergence of prosodic boundary: continuous effects of temporal affordance on inter-gestural timing. *J. Phon.* 44:110–29

Bishop J. 2016a. Individual differences in top-down and bottom-up prominence perception. In *Proceedings of the 8th International Conference on Speech Prosody*, ed. J Barnes, A Brugos, S Shattuck-Hufnagel, N Veilleux, pp. 668–72. Boston: ICSA

Bishop J. 2016b. *Does working memory capacity predict individual differences in implicit and explicit prosodic phrasing?* Poster presented at Framing Speech: Celebrating 40 Years of Inquiry with Stefanie Shattuck-Hufnagel, MIT/Boston Univ., Boston, MA

Bombien L, Mooshammer C, Hoole P. 2013. Articulatory coordination in word-initial clusters of German. *J. Phon.* 41:546–61

Bombien L, Mooshammer C, Hoole P, Kühnert B. 2010. Prosodic and segmental effects on EPG contact patterns of word-initial German clusters. *J. Phon.* 38:388–403

Bombien L, Mooshammer C, Hoole P, Kühnert B, Schneeberg J. 2006. An EPG study of initial /kl/ clusters in varying prosodic conditions in German. In *Proceedings of the 7th International Seminar on Speech Production*, ed. HC Yehia, D Demolin, R Laboissière, pp. 35–42. Pampulha, Braz.: CEFALA

Bouchard KE, Mesgarani N, Johnson K, Chang EF. 2013. Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495:327–32

Browman CP, Goldstein L. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45(2–4):140–55

Browman CP, Goldstein L. 1989. Articulatory gestures as phonological units. *Phonology* 6:201–51

Browman CP, Goldstein L. 1990. Representation and reality: physical systems and phonological structure. *J. Phon.* 18:411–24

Browman CP, Goldstein L. 1992. Articulatory phonology: an overview. *Phonetica* 49(3/4):155–80

Browman CP, Goldstein L. 1995. Dynamics and articulatory phonology. In *Mind as Motion: Dynamics, Behavior, and Cognition*, ed. RF Port, T van Gelder, pp. 175–94. Cambridge, MA: MIT Press

Browman CP, Goldstein L. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bull. Commun. Parlée* 5:25–34

Byrd D. 1995. C-Centers revisited. *Phonetica* 52:285–306

Byrd D. 1996. A phase window framework for articulatory timing. *Phonology* 13(2):139–69

Byrd D. 2000. Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57:3–16

Byrd D, Choi S. 2010. At the juncture of prosody, phonology, and phonetics—the interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In *Laboratory Phonology 10*, ed. C Fougeron, B Kuehnert, M D'Imperio, N Vallee, pp. 31–60. Berlin: Mouton de Gruyter

Byrd D, Kaun A, Narayanan S, Saltzman E. 2000. Phrasal signatures in articulation. In *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, ed. MB Broe, JB Pierrehumbert, pp. 70–87. Cambridge, UK: Cambridge Univ. Press

Byrd D, Krivokapić J, Lee S. 2006. How far, how long: on the temporal scope of phrase boundary effects. *J. Acoust. Soc. Am.* 120:1589–99

Byrd D, Lee S, Riggs D, Adams J. 2005. Interacting effects of syllable and phrase position on consonant articulation. *J. Acoust. Soc. Am.* 118:3860–73

Byrd D, Riggs D. 2008. Locality interactions with prominence in determining the scope of phrasal lengthening. *J. Int. Phon. Assoc.* 38:187–202

Byrd D, Saltzman E. 1998. Intragestural dynamics of multiple phrasal boundaries. *J. Phon.* 26:173–99

Byrd D, Saltzman E. 2003. The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *J. Phon.* 31:149–80

Byrd D, Tobin S, Bresch E, Narayanan S. 2009. Timing effects of syllable structure and stress on nasals: a real-time MRI examination. *J. Phon.* 37:97–110

Cangemi F, Krüger M, Grice M. 2015. Listener-specific perception of speaker-specific production in intonation. In *Individual Differences in Speech Production and Perception*, ed. S Fuchs, D Pape, C Petrone, P Perrier, pp. 123–45. Frankfurt, Ger.: Lang

Chartier J, Anumanchipalli GK, Johnson K, Chang EF. 2018. Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex. *Neuron* 98:1042–54.e4

Chen W, Byrd D, Narayanan S, Nayak KS. 2019. Intermittently tagged real-time MRI reveals internal tongue motion during speech production. *Magn. Reson. Med.* 82(2):600–13

Chitoran I, Goldstein L, Byrd D. 2002. Gestural overlap and recoverability: articulatory evidence from Georgian. In *Laboratory Phonology* 7, ed. C Gussenhoven, N Warner, pp. 419–48. Berlin: Mouton de Gruyter

Cho T. 2006. Manifestation of prosodic structure in articulation: evidence from lip movement kinematics in English. In *Laboratory Phonology 8*, ed. L Goldstein, DH Whalen, CT Best, pp. 519–48. Berlin: Mouton de Gruyter

Cho T, Keating PA. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *J. Phon.* 29:155–90

Cho T, Minjung S, Kim S. 2016. Articulatory reflexes of the three-way contrast in labial stops and kinematic evidence for domain-initial strengthening in Korean. *J. Int. Phon. Assoc.* 46:129–55

Cho T, Yeomin Y, Kim S. 2014. Effects of prosodic boundary and syllable structure on the temporal realization of CV gestures in Korean. *J. Phon.* 44:96–109

Cole J. 2015. Prosody in context: a review. *Lang. Cogn. Neurosci.* 30(1/2):1–31

Danner SG. 2017. *Effects of speech context on characteristics of manual gesture*. PhD Thesis, Univ. South. Calif., Los Angeles

Danner SG, Krivokapić J, Byrd D. 2019. *Co-speech movement behavior in conversational turn-taking*. Poster presented at the 178th Meeting of the Acoustical Society of America, San Diego, Dec. 2–6

Dauer RM. 1983. Stress-timing and syllable-timing reanalyzed. *J. Phon.* 11:51–62

de Jong KJ. 2001. Rate-induced resyllabification revisited. *Lang. Speech* 44:197–216

de Jong KJ, Beckman ME, Edwards J. 1993. The interplay between prosodic structure and coarticulation. *Lang. Speech* 36:197–212

D'Imperio M, Espesser R, Loevenbruck H, Menezes C, Nguyen N, Welby P. 2007. Are tones aligned with articulatory events? Evidence from Italian and French. In *Laboratory Phonology 9*, ed. J Cole, JI Hualde, pp. 577–608. Berlin: Mouton de Gruyter

Edwards J, Beckman ME, Fletcher J. 1991. The articulatory kinematics of final lengthening. *J. Acoust. Soc. Am.* 89:369–82

Ferreira F, Swets B. 2002. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *J. Mem. Lang.* 46:57–84

Fletcher J. 2010. The prosody of speech: timing and rhythm. In *The Handbook of Phonetic Sciences*, ed. WJ Hardcastle, J Laver, FE Gibbon, pp. 521–602. New York: Wiley-Blackwell

Fougeron C. 2001. Articulatory properties of initial segments in several prosodic constituents in French. *J. Phon.* 29:109–35

Fougeron C, Keating PA. 1997. Articulatory strengthening at edges of prosodic domains. *J. Acoust. Soc. Am.* 101:3728–40

Fowler CA. 1980. Coarticulation and theories of extrinsic timing. *J. Phon.* 8:113–33

Fowler CA, Rubin PE, Remez RE, Turvey MT. 1980. Implications for speech production of a general theory of action. In *Language Production*, Vol. 1: *Speech and Talk*, ed. B Butterworth, pp. 373–420. New York: Academic

Gafos A. 2002. A grammar of gestural coordination. *Nat. Lang. Linguist. Theory* 20(2):269–337

Gafos A, Beňuš Š. 2006. Dynamics of phonological cognition. *Cogn. Sci.* 30(5):905–43

Gafos A, Goldstein L. 2012. Articulatory representation and organization. In *The Oxford Handbook of Laboratory Phonology*, ed. AC Cohn, C Fougeron, MK Huffman, pp. 220–31. Oxford, UK: Oxford Univ. Press

Gao M. 2008. *Mandarin tones: an Articulatory Phonology account*. PhD Thesis, Yale Univ., New Haven, CT

Georgeton L, Fougeron C. 2014. Domain-initial strengthening on French vowels and phonological contrasts: evidence from lip articulation and spectral variation. *J. Phon.* 44:83–95

Georgeton L, Kocjančič AT, Fougeron C. 2016. Effect of domain initial strengthening on vowel height and backness contrasts in French: acoustic and ultrasound data. *J. Speech Lang. Hear. Res.* 59(6):S1575–85

Gick B, Wilson I, Kock K, Cook C. 2004. Language-specific articulatory settings: evidence from inter-utterance rest position. *Phonetica* 61:220–33

Goldstein L, Byrd D, Saltzman E. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In *From Action to Language: The Mirror Neuron System*, ed. M Arbib, pp. 215–49. Cambridge, UK: Cambridge Univ. Press

Goldstein L, Chitoran I, Selkirk E. 2007a. Syllable structure as coupled oscillator modes: evidence from Georgian versus Tashlhiyt Berber. In *Proceedings of the 16th International Congress of Phonetic Sciences* (*ICPhS XVI*), ed. J Trouvain, pp. 241–44. Saarbrücken, Ger.: Univ. Saarlandes

Goldstein L, Nam H, Saltzman E, Chitoran I. 2009. Coupled oscillator planning model of speech timing and syllable structure. In *Frontiers in Phonetics and Speech Science*, ed. G Fant, H Fujisaki, J Shen, pp. 239–50. Beijing: Commercial

Goldstein L, Pouplier M, Chen L, Saltzman E, Byrd D. 2007b. Dynamic action units slip in speech production errors. *Cognition* 103:386–412

Grice M, Ritter S, Niemann H, Roettger TB. 2017. Integrating the discreteness and continuity of intonational categories. *J. Phon.* 64:90–107

Hagedorn C, Sorensen T, Lammert A, Toutios A, Goldstein L, et al. 2019. Engineering innovation in speech science: data and technologies. *Perspect. ASHA Spec. Interest Groups* 4(2):411–20

Hardcastle WJ. 1985. Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Commun.* 4:247–63

Harrington J, Fletcher J, Roberts C. 1995. An analysis of truncation and linear rescaling in the production of accented and unaccented vowels. *J. Phon.* 23:305–22

Hermes A, Mücke D, Auris B. 2017. The variability of syllable patterns in Tashlhiyt Berber and Polish. *J. Phon.* 64:127–44

Hermes A, Mücke D, Grice M. 2013. Gestural coordination of Italian word-initial clusters: the case of 'impure s.' *Phonology* 30:1–25

Holst T, Nolan F. 1995. The influence of syntactic structure on [s] and [ʃ] assimilation. In *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, ed. B Connell, A Arvanti, pp. 315–33. Cambridge, UK: Cambridge Univ. Press

Honorof DN, Browman CP. 1995. The center or edge: How are consonant clusters organized with respect to the vowel? In *Proceedings of the 13th International Congress of Phonetic Sciences* (*ICPhS 95*), Vol. 3, ed. K Elenius, P Branderud, pp. 552–55. Stockholm: KTH/Stockholm Univ.

Hsieh FY. 2017. *A gestural approach to the phonological representation of English diphthongs*. PhD Thesis, Univ. South. Calif., Los Angeles

Hu F. 2016. Tones are not abstract autosegmentals. In *Proceedings of the 8th International Conference on Speech Prosody*, ed. J Barnes, A Brugos, S Shattuck-Hufnagel, N Veilleux, pp. 302–6. Boston: ICSA

Karlin R, Tilsen S. 2015. The articulatory tone-bearing unit: gestural coordination of lexical tone in Thai. *Proc. Meet. Acoust.* 22:060006
Katsika A. 2012. *Coordination of prosodic gestures at boundaries in Greek*. PhD Thesis, Yale Univ., New Haven, CT
Katsika A. 2016. The role of prominence in determining the scope of boundary-related lengthening in Greek. *J. Phon.* 55:149–81
Katsika A. 2018. The kinematic profile of prominence in Greek. In *Proceedings of the 9th International Conference on Speech Prosody*, ed. K Klessa, J Bachan, A Wagner, M Karpiński, D Śledziński, pp. 764–68. Boston: ICSA
Katsika A, Krivokapić J, Mooshammer C, Tiede M, Goldstein L. 2014. The coordination of boundary tones and its interaction with prominence. *J. Phon.* 44:62–82
Keating P, Cho T, Fougeron C, Hsu C. 2004. Domain-initial articulatory strengthening in four languages. In *Phonetic Interpretation: Papers in Laboratory Phonology VI*, ed. J Local, R Ogden, R Temple, pp. 143–61. Cambridge, UK: Cambridge Univ. Press
Keating P, Shattuck-Hufnagel S. 2002. A prosodic view of word form encoding for speech production. *UCLA Work. Pap. Phon.* 101:112–56
Kelso S, Saltzman E, Tuller B. 1986. The dynamical perspective on speech production: data and theory. *J. Phon.* 14:29–59
Kim H, Cole J. 2005. The stress foot as a unit of planned timing: evidence from shortening in the prosodic phrase. In *Proceedings of the 9th European Conference on Speech Communication and Technology* (*INTERSPEECH 2005*), pp. 2365–68. Boston: ICSA
Kim J. 2020. *Individual differences in the production and perception of prosodic boundaries in American English*. PhD Thesis, Univ. Mich., Ann Arbor
Kim S, Jang J, Cho T. 2017. Articulatory characteristics of preboundary lengthening in interaction with prominence on tri-syllabic words in American English. *J. Acoust. Soc. Am.* 142:EL362–68
Klatt DH. 1973. Interaction between two factors that influence vowel duration. *J. Acoust. Soc. Am.* 54:1102–4
Kochetov A, So CK. 2007. Place assimilation and phonetic grounding: a cross-linguistic perceptual study. *Phonology* 24:397–432
Krivokapić J. 2007a. Prosodic planning: effects of phrasal length and complexity on pause duration. *J. Phon.* 35:162–79
Krivokapić J. 2007b. *The planning, production, and perception of prosodic structure*. PhD Thesis, Univ. South. Calif., Los Angeles
Krivokapić J. 2012. Prosodic planning in speech production. In *Speech Planning and Dynamics*, ed. S Fuchs, M Weirich, D Pape, P Perrier, pp. 157–90. Frankfurt, Ger.: Lang
Krivokapić J. 2013. Rhythm and convergence between speakers of American and Indian English. *J. Lab. Phonol.* 4(1):39–65
Krivokapić J. 2020. Prosody in articulatory phonology. In *Prosodic Theory and Practice*, ed. S Shattuck-Hufnagel, J Barnes. Cambridge, MA: MIT Press. In press
Krivokapić J, Byrd D. 2012. Prosodic boundary strength: an articulatory and perceptual study. *J. Phon.* 40(3):430–42
Krivokapić J, Styler W, Parrell B. 2020. Pause postures: the relationship between articulation and cognitive processes during pauses. *J. Phon.* 79:100953
Krivokapić J, Tiede M, Tyrone M. 2017. A kinematic study of prosodic structure in articulatory and manual gestures: results from a novel method of data collection. *J. Lab. Phonol.* 8(1):3
Krivokapić J, Tiede M, Tyrone M, Goldenberg D. 2016. Speech and manual gesture coordination in a pointing task. In *Proceedings of the 8th International Conference on Speech Prosody*, ed. J Barnes, A Brugos, S Shattuck-Hufnagel, N Veilleux, pp. 1240–44. Boston: ICSA
Kühnert B, Hoole P, Mooshammer C. 2006. Gestural overlap and C-center in selected French consonant clusters. In *Proceedings of the 7th International Seminar on Speech Production*, ed. HC Yehia, D Demolin, R Laboissière, pp. 327–34. Pampulha, Braz.: CEFALA
Ladd DR. 2006. Segmental anchoring of pitch movements: autosegmental association or gestural coordination? *Ital. J. Linguist.* 18(1):19–38
Ladd DR, Faulkner D, Faulkner H, Schepman A. 1999. Constant 'segmental' anchoring of f_0 movements under changes in speech rate. *J. Acoust. Soc. Am.* 106:1543–54

Large EW. 2008. Resonating to musical rhythm: theory and experiment. In *The Psychology of Time*, ed. S Grondin, pp. 189–232. Bingley, UK: Emerald

Large EW, Fink P, Kelso JAS. 2002. Tracking simple and complex sequences. *Psychol. Res.* 66:3–17

Lee S, Byrd D, Krivokapić J. 2006. Functional data analysis of prosodic effects on articulatory timing. *J. Acoust. Soc. Am.* 119:1666–71

Lee Y. 2018. *The prosodic substrate of consonant and tone dynamics*. PhD Thesis, Univ. South. Calif., Los Angeles

Lee Y, Danner SG, Parrell B, Lee S, Goldstein L, Byrd D. 2018a. Articulatory, acoustic, and prosodic accommodation in a cooperative maze navigation task. *PLOS ONE* 13(8):e0201444

Lee Y, Goldstein L, Byrd D. 2018b. *Dynamical systems model of prosodic asymmetries in the co-expression of phrasal and segmental tone*. Paper presented at the Linguistic Society of America Annual Meeting, New York, NY, Jan. 3–6

Lehiste I. 1972. The timing of utterances and linguistic boundaries. *J. Acoust. Soc. Am.* 51(6, Part 2):2018–24

Lim Y, Zhu Y, Lingala SG, Byrd D, Narayanan S, Nayak K. 2019. 3D real-time MRI of the vocal tract during natural speech. *Magn. Reson. Med.* 81(3):1511–20

Marin S, Pouplier M. 2010. Temporal organization of complex onsets and codas in American English: testing the predictions of a gestural coupling model. *Motor Control* 14(3):380–407

McClean M. 1973. Forward coarticulation of velar movement at marked junctural boundaries. *J. Speech Hear. Res.* 16:286–96

Mücke D, Hermes A, Tilsen S. 2019. Strength and structure: coupling tones with oral constriction gestures. In *Proceedings of the 20th European Conference on Speech Communication and Technology* (*INTERSPEECH 2019*), ed. G Kubin, Z Kačič, pp. 914–18. Boston: ICSA

Mücke D, Nam H, Hermes A, Goldstein LM. 2012. Coupling of tone and constriction gestures in pitch accents. In *Consonant Clusters and Structural Complexity*, ed. P Hoole, L Bombien, M Pouplier, C Mooshammer, B Kühnert, pp. 205–30. Berlin: Mouton de Gruyter

Nam H. 2007. A competitive, coupled oscillator model of moraic structure: split-gesture dynamics focusing on positional asymmetry. In *Laboratory Phonology 9*, ed. J Cole, JI Hualde, pp. 483–506. Berlin: Mouton de Gruyter

Nam H, Goldstein L, Saltzman E. 2009. Self-organization of syllable structure: a coupled oscillator model. In *Approaches to Phonological Complexity*, ed. F Pellegrino, E Marisco, I Chitoran, C Coupé, pp. 299–328. Berlin/New York: Mouton de Gruyter

Nam H, Saltzman E. 2003. A competitive, coupled oscillator of syllable structure. In *Proceedings of the 12th International Congress of Phonetic Sciences* (*ICPhS XII*), Vol. 3, ed. P Romeas, pp. 2253–56. Aix-en-Provence, Fr.: Inst. Phon.

Nava E, Tepperman J. 2010. *Modeling second language prosody acquisition: a dynamic systems theory approach*. Paper presented at the 35th Annual Boston University Conference on Language Development (BUCLD 35), Boston, Novemb. 5–7

Niemann HH. 2016. *The coordination of pitch accents with articulatory gestures: a dynamical approach*. PhD Thesis, Univ. Köln, Köln, Ger.

Niemann HH, Mücke D, Nam H, Goldstein L, Grice M. 2011. Tones as gestures: the case of Italian and German. In *Proceedings of the 17th International Congress of Phonetic Sciences* (*ICPhS XVII*), pp. 1486–89. London: IPA

O'Dell M, Nieminen T. 1999. Coupled oscillator model of speech rhythm. In *Proceedings of the 14th International Congress of Phonetic Sciences* (*ICPhS 14*), Vol. 2, ed. J Ohala, Y Hasegawa, M Ohala, D Granville, A Bailey, pp. 1075–78. Berkeley: Univ. Calif.

Oh M, Byrd D. 2019. Syllable-internal corrective focus in Korean. *J. Phon.* 77:100933

Oh M, Lee Y. 2018. ACT: an Automatic Centroid Tracking tool for analyzing vocal tract actions in real-time magnetic resonance imaging speech production data. *J. Acoust. Soc. Am.* 144(4):EL290–96

Parrell B. 2012. The role of gestural phasing in Western Andalusian Spanish aspiration. *J. Phon.* 40:37–45

Parrell B, Goldstein L, Lee S, Byrd D. 2014. Spatiotemporal coupling between speech and manual motor actions. *J. Phon.* 42:1–11

Parrell B, Narayanan S. 2018. Explaining coronal reduction: prosodic structure and articulatory posture. *Phonetica* 75(2):151–81

Port R. 1981. Linguistics timing factors in combination. *J. Acoust. Soc. Am.* 69:262–74

Port R. 2003. Meter and speech. *J. Phon.* 31:599–611

Port R, Cummins F, Gasser M. 1995. A dynamic approach to rhythm in language: toward a temporal phonology. arXiv:cmp-lg/9508007

Pouplier M. 2012. The gestural approach to syllable structure: universal, language- and cluster-specific aspects. In *Speech Planning and Dynamics*, ed. S Fuchs, M Weirich, D Pape, P Perrier, pp. 63–96. Frankfurt, Ger.: Lang

Pouplier M, Beňuš Š. 2011. On the phonetic status of syllabic consonants: evidence from Slovak. *J. Lab. Phonol.* 2(2):243–73

Prieto P, Torreira F. 2007. The segmental anchoring hypothesis revisited: syllable structure and speech rate effects on peak timing in Spanish. *J. Phon.* 35:473–500

Rakerd B, Sennett W, Fowler CA. 1987. Domain-final lengthening and foot level shortening in spoken English. *Phonetica* 44:147–55

Ramanarayanan V, Bresch E, Byrd D, Goldstein L, Narayanan SS. 2009. Analysis of pausing behavior in spontaneous speech using real-time magnetic resonance imaging of articulation. *J. Acoust. Soc. Am.* 126(5):EL160–65

Ramanarayanan V, Goldstein L, Byrd D, Narayanan SS. 2013. An investigation of articulatory setting using real-time magnetic resonance imaging. *J. Acoust. Soc. Am.* 134:510–19

Rasskazova O, Mooshammer C, Fuchs S. 2018. Articulatory settings during inter-speech pauses. In *Proceedings of the 13th Conference on Phonetics and Phonology in the German-Speaking Countries* (*P&P 13*), ed. M Belz, S Fuchs, S Jannedy, C Mooshammer, O Rasskazova, M Zygis, pp. 161–64. Berlin: ZAS/Humboldt Univ.

Rasskazova O, Mooshammer C, Fuchs S. 2019. Temporal coordination of articulatory and respiratory events prior to speech initiation. In *Proceedings of the 20th European Conference on Speech Communication and Technology* (*INTERSPEECH 2019*), ed. G Kubin, Z Kačič, pp. 884–88. Boston: ICSA

Reichel U, Cole J. 2016. Entrainment analysis of categorical intonation representations. In *Proceedings of the 12th Conference on Phonetics and Phonology in the German-Speaking Countries* (*P&P 12*), ed. O Jokisch. Munich, Ger.: Univ. Munich

Reichel U, Mady K, Cole J. 2018. Prosodic entrainment in dialog acts. arXiv:1810.12646 [cs]

Roessig S, Mücke D. 2019. Modeling dimensions of prosodic prominence. *Front. Commun.* 4:44

Roessig S, Mücke D, Grice M. 2019. The dynamics of intonation: categorical and continuous variation in an attractor-based model. *PLOS ONE* 14(5):e216859

Roy J, Cole J, Mahrt T. 2017. Individual differences and patterns of convergence in prosody perception. *J. Lab. Phonol.* 8(1):22

Saltzman E. 1995. Dynamics and coordinate systems in skilled sensorimotor activity. In *Mind as Motion: Dynamics, Behavior, and Cognition*, ed. RF Port, T van Gelder, pp. 149–73. Cambridge, MA: MIT Press

Saltzman E, Löfqvist A, Mitra S. 2000. "Glue" and "clocks": intergestural cohesion and global timing. In *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, ed. MB Broe, JB Pierrehumbert, pp. 88–101. Cambridge, UK: Cambridge Univ. Press

Saltzman E, Munhall KG. 1989. A dynamical approach to gestural patterning in speech production. *Ecol. Psychol.* 1:333–82

Saltzman E, Nam H, Krivokapić J, Goldstein L. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th International Conference on Speech Prosody*, ed. PA Barbosa, S Madureira, C Reis, pp. 175–84. Boston: ICSA

Shaw JA, Gafos AI. 2015. Stochastic time models of syllable structure. *PLOS ONE* 10(5):e0124714

Shaw JA, Gafos AI, Hoole P, Zeroual C. 2009. Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation. *Phonology* 26(1):187–215

Shaw JA, Gafos AI, Hoole P, Zeroual C. 2011. Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters. *Phonology* 28(3):455–90

Shih S, Inkelas S. 2019. Autosegmental aims in surface optimizing phonology. *Linguist. Inq.* 50(1):137–96

Swets B, Desmet T, Hambrick DZ, Ferreira F. 2007. The role of working memory in syntactic ambiguity resolution: a psychometric approach. *J. Exp. Psychol. Gen.* 136:64–81

Swets B, Petrone C, Fuchs S, Krivokapić J. 2017. *Individual differences in planning strategies among French, German and English speakers*. Poster presented at the 58th Annual Meeting of the Psychonomic Society, Vancouver, Can., Novemb. 9–12

Tabain M. 2003. Effects of prosodic boundary on /aC/ sequences: articulatory results. *J. Acoust. Soc. Am.* 113:2834–49

Tang C, Hamilton LS, Chang EF. 2017. Intonational speech prosody encoding in the human auditory cortex. *Science* 357(6353):797–801

Tiede M, Mooshammer C. 2013. Evidence for an articulatory component of phonetic convergence from dual electromagnetic articulometer observation of interacting talkers. *J. Acoust. Soc. Am.* 19:060138

Tilsen S. 2016. Selection and coordination: the articulatory basis for the emergence of phonological structure. *J. Phon.* 55:53–77

Tilsen S. 2019. Space and time in models of speech rhythm. *Ann. N. Y. Acad. Sci.* 1453:47–66

Tilsen S, Arvaniti A. 2013. Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *J. Acoust. Soc. Am.* 134:628–39

Tilsen S, Spincemaille P, Xu B, Doerschuk P, Luh W-M, et al. 2016. Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units. *PLOS ONE* 11(1):e0146813

Toutios A, Byrd D, Goldstein L, Narayanan S. 2019. Advances in vocal tract imaging and analysis. In *Routledge Handbook of Phonetics*, ed. WF Katz, PF Assmann, chapter 2. London: Routledge. **https://www.routledgehandbooks.com/doi/10.4324/9780429056253-3**

Toutios A, Lingala SG, Vaz C, Kim J, Esling J, et al. 2016. Illustrating the production of the International Phonetic Alphabet sounds using fast real-time magnetic resonance imaging. In *Proceedings of the 17th European Conference on Speech Communication and Technology* (*INTERSPEECH 2016*), ed. N Morgan, pp. 2428–32. Boston: ICSA

Tuller B, Case P, Ding M, Kelso JAS. 1994. The nonlinear dynamics of speech categorization. *J. Exp. Psychol.* 20:3–16

Vatikiotis-Bateson E, Barbosa AV, Best CT. 2014. Articulatory coordination of two vocal tracts. *J. Phon.* 44:167–81

Wagner M. 2005. *Prosody and recursion*. PhD Thesis, MIT, Cambridge, MA

Wagner M, Watson DG. 2010. Experimental and theoretical advances in prosody: a review. *Lang. Cogn. Proc.* 25(7–9):905–45

Wagner P, Malisz Z, Kopp S. 2014. Gesture and speech in interaction: an overview. *Speech Commun*. 57:209–32

Zhang M, Geissler C, Shaw JA. 2019. Gestural representations of tone in Mandarin: evidence from timing alternations. In *Proceedings of the 19th International Congress of Phonetic Sciences* (*ICPhS 2019*), ed. S Calhoun, P Escudero, M Tabain, P Warren, pp. 1803–7. Canberra: Australas. Speech Sci. Technol. Assoc.

Zsiga EC. 2000. Phonetic alignment constraints: consonant overlap and palatalization in English and Russian. *J. Phon.* 28:69–102

RELATED RESOURCES

Gestural model. **https://haskinslabs.org/research/features-and-demos/gestural-model**

Annual Review of Linguistics

Volume 7, 2021

Contents

Errata

An online log of corrections to *Annual Review of Linguistics* articles may be found at http://www.annualreviews.org/errata/linguistics