# A task-dynamic toolkit for modeling the effects of prosodic structure on articulation

*Elliot Saltzman*[1,2]*, Hosung Nam*[1]*, Jelena Krivokapic*[1,3]*, & Louis Goldstein*[1,4]

[1]Haskins Laboratories, USA; [2]Department of Physical Therapy and Athletic Training, Boston University, USA; [3]Department of Linguistics, Yale University, USA; [4]Department of Linguistics, University of Southern California, USA
esaltz@bu.edu; hosung.nam@haskins.yale.edu; jelena.krivokapic@yale.edu; goldstein@haskins.yale.edu

## Abstract

The original task-dynamic model of speech production incorporated the theoretical tenets of Articulatory Phonology and provided a dynamics of inter-articulator coordination for single and co-produced constriction gestures, given a gestural score that specifies a time-dependent vector of gestural activations for a given utterance. More recently, the model has been significantly extended to provide a framework for investigating the higher order dynamics of prosodic phrasing, syllable structure, lexical stress, and the prominence (accentual) properties associated with higher level prosodic constituents (e.g., foot, word, phrase, sentence). There are two new components in the model. The first is an ensemble of *gestural planning oscillators* that defines a dynamics of gestural score formation in that, once the ensemble reaches an entrained steady-state of relative phasing, the waveform of each oscillator is used to specify the activation function of that oscillator's associated constriction gesture and to trigger, thereby, the onset of the gesture. The second component is a set of *modulation gestures* ($\mu$-gestures) that, rather than activating constriction formation and release gestures in the vocal tract, serve to modulate the temporal and spatial properties of all concurrently active constriction gestures. Modulation gestures are of two types: temporal modulation gestures ($\mu_T$-gestures) that alter the rate of utterance timeflow by smoothly changing all frequency parameters of the planning oscillator ensemble; and spatial modulation gestures ($\mu_S$ -gestures) that spatially strengthen or reduce the motions of constriction gestures by smoothly changing the spatial target parameters of these constriction gestures. Key to the representation of prosodic phrasing has been use of clock-slowing temporal modulation gestures (called prosodic gestures [$\pi$-gestures] in previous work) that are locally active in the region of phrasal boundaries, and that slow the rate of utterance timeflow in direct proportion to the strength of the associated boundary. Central to the representation of syllable structure is the use of a *coupling graph* that defines the existence and strength of coupling in the network of gestural planning oscillators, and shapes the manner in which gestures are coordinated. Concepts from graph theory have been crucial to understanding how hypothesized differences among coupling graphs have correctly predicted empirically demonstrated intra-syllabic differences between onsets and codas in both the mean values and variabilities of C-C, C-V, and V-C timing patterns. In this paper, we describe a set of recent developments to our task-dynamic 'toolkit' (planning oscillator ensemble and temporal modulation gestures) and how they have been used to interpret and simulate experimental data on the interactions of stress and prominence in shaping the "prosodically driven phonetic detail" [14] of speech.

## 1. Introduction

To communicate a spoken a message to a listener, speakers produce coordinated articulatory movements that structure patterns in the acoustic medium through which the message is transmitted. The surface realizations of these articulatory and acoustic patterns are variable, and one of the main tasks of linguistic theory and speech science is to account for the relationship between the invariant linguistic units that are hypothesized to underlie spoken language and their variable surface manifestations. The *task-dynamic* model of speech production [52], [9], [42] has been able to capture this relationship by postulating articulatory gestures as dynamic units of speech production, that account for the invariant and variant properties of speech without a mediating level. In this paper, current research within the task-dynamic model of speech production will be presented. We start with a brief review of the main properties of the model and continue with recent developments in syllable, gestural, phrasal and foot modeling.

In the task-dynamic model, the spatiotemporal patterns of articulatory motion emerge as behaviors implicit in a dynamical system with two functionally distinct but interacting levels. The *interarticulator* coordination level is defined according to both *model articulator* (e.g. lips & jaw) variables and goal space or *tract-variables* (which are constriction-based, e.g. lip aperture [LA] & protrusion [LP]; Table 1). The *intergestural* level is defined according to a set of *planning oscillator* variables and *activation* variables. The activation trajectories shaped by the intergestural level define a *gestural score* (see Figure 1 for a schematic example using the word "spot") that provides driving input to the interarticulator level. The constriction gestures in an utterance's gestural score are defined by sets of invariant, context-independent dynamical parameters (e.g., target, stiffness, & damping coefficients) that characterize the gestures' point attractor dynamics, and are associated with corresponding subsets of model articulator, tract-variable, and activation variables. Each activation variable reflects the strength with which the associated gesture (e.g., lip closure) "attempts" to shape vocal tract movements at any given point in time. The tract-variables and model articulator variables associated with each gesture specify, respectively, the particular vocal-tract constriction (e.g. lips) and articulatory synergy (e.g., upper lip, lower lip, & jaw) whose behaviors are affected directly by the associated gesture's activation.

| Tract Variables | | Model Articulators |
|---|---|---|
| **LP** | lip protrusion | upper and lower lips |
| **LA** | lip aperture | upper and lower lips, jaw |
| **TDCL** | tongue dorsum constriction location | tongue body, jaw |
| **TDCD** | tongue dorsum constriction degree | tongue body, jaw |
| **LTH** | lower tooth height | jaw |
| **TTCL** | tongue tip constriction location | tongue tip, body, jaw |
| **TTCD** | tongue tip constriction degree | tongue tip, body, jaw |
| **TTCO** | tongue tip constriction orientation | tongue tip, body, jaw |
| **VEL** | velic aperture | velum |
| **GLO** | glottal aperture | glottal width |

Table 1. *Tract-variables and model articulators*

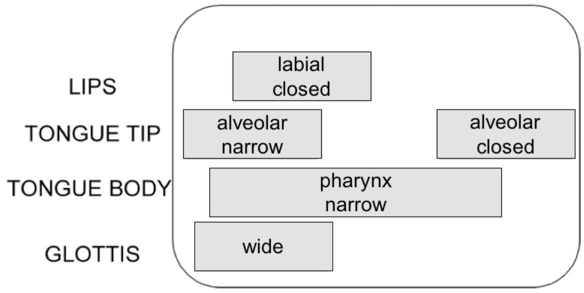

Figure 1. *Schematic gestural score for "spot", indicating the time intervals of gestural activation for the onset consonants (two oral gestures and a single laryngeal abduction gesture), the vowel, and the coda consonant.*

At the interarticulator level, each constriction gesture is modeled with invariant point-attractor dynamics, and the concurrent activation of multiple gestures results in correspondingly context-dependent patterns of coordinated articulator motion. These activation patterns are specified at the intergestural level of the model, and can be thought of as implementing a dynamics of *planning*—it determines the patterns of relative timing among the activation waves of gestures participating in an utterance as well as the shapes and durations of the individual gesture activation waves. Each gesture's activation wave acts to insert the gesture's parameter set into the interarticulator dynamical system defined by the set of tract-variable and model articulator coordinates (see [52], for further details). In the original version of the model (e.g. [8]), the activation variables in gestural scores were determined by a set of rules that specified the relative phasing of the gestures and calculated activation trajectories based on those phases and the time constants associated with the individual gestures. The gestural score then unidirectionally drove articulatory motion at the interarticulator level. Thus, intergestural timing was not part of the dynamical system, per se, and such a model was not capable of exhibiting dynamical coherence, such as can be seen, for example, in the temporal adjustment to external perturbation [50]. In the current model, however, intergestural timing is determined by the ensemble of nonlinear limit-cycle *planning oscillators* associated with the set of gestures in a given utterance, with one oscillator being associated with each gesture. As described in [54], an advantage of such a network architecture is that it can exhibit the hallmark nonlinear behaviors of coupled limit-cycle systems—entrainment, multiple stable modes, and changes in relative phasing (both gradual and abrupt)—all of which are relevant to speech timing.

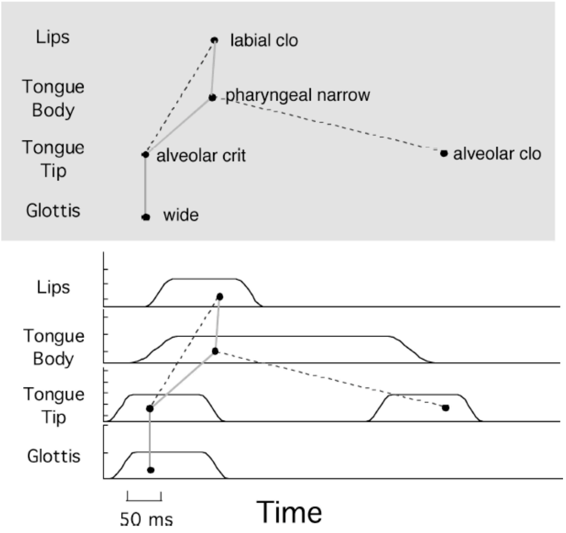

Figure 2. *Coupling graph for "spot" (top), also superposed on corresponding gestural score (bottom). Solid and dashed lines denote in-phase and anti-phase coupling relationships, respectively.*

We consider gestures to be the 'atomic' functional units of speech production that are combined with one another to form larger 'molecular' structures such as segments, syllables, and lexical items. Our focus to date has been on syllable- and word-sized molecules, ignoring the lexical stress and accentual characteristics displayed by such molecules. In our model, we create gestural molecules by coupling gestural planning oscillators to one another in a pairwise, bidirectional manner that is specific to the planned molecule. This coupling process creates a structure that can be represented as a (phonological) *coupling graph*, which is part of the lexical specification of the molecule and determines the coordination between gestures (Figure 2). In this graph, nodes represent gestures and internode links represent the intergestural coupling functions; once the coupling graph is specified, it is used to parameterize the equations of motion for the planning oscillators, which are then numerically integrated until the system reaches a steady-state pattern of interoscillator relative phasing (e.g., [42], [40]). This relative phasing pattern is then mapped into a corresponding pattern of gestural activations, creating a gestural score that is used to trigger the associated constriction gestures. The time taken by the planning oscillator ensemble to converge to a stable, steady-state pattern of relative phasing will differ as a function of the properties (e.g., graph topology, interoscillator coupling strengths and target relative phases) of the particular coupling graph that is implemented. We refer to this settling or stabilization time as the system's *planning time*. Nam [40] has found that the model's settling time correlates well with speakers' behavioral reaction times to begin to produce utterances that vary in phonological structure; in other words, the model's planning time appears to reflect the time taken by a speaker's real-time cognitive planning process.

Central to understanding syllable structure using a coupling model is the hypothesis that there are two basic types of constriction gestures—consonant and vowel—and that the internal structure of syllables results from different ways of coordinating gestures of these basic types. Recently, we have developed a theoretical account of why certain structural properties (such as being an onset vs. a coda consonant) can be considered as relatively less stable than others [10], [54], [53]. According to this account, gestures in different structural positions enter into a different number and different types (in-phase [0°], anti-phase [180°]) of coupling relations as specified by an utterance's coupling graph, and these coupling relations are assumed to exhibit different degrees of stability and planning stabilization time. This hypothesis of differential stability for in-phase vs. anti-phase is based, in part, on the fact that in human inter-limb coordination, certain rhythmic phase-locked modes are spontaneously available—in-phase and anti-phase (e.g. [60])—with the in-phase mode being the more stable of the two (e.g., [21], [56]). Other phase-locks can be learned, but only with difficulty, and we hypothesize that phonological systems make use of the more intrinsically stable modes where possible [20].

Thus, we specify constriction gestures to be pairwise coordinated in either in-phase or anti-phase modes, with syllable-initial consonants and their following vowels being coordinated in-phase with one another in what we call the *onset relation*. When *multiple* consonants occur in an onset, such as in the consonant cluster at the beginning of the word "spot," we assume that *each* of the consonants is coupled in-phase with the vowel (the syllable nucleus)—this is what makes them part of the onset. However, the consonant gestures must be at least partially sequential in order for the resulting form to be perceptually recoverable. Therefore, we specify anti-phase couplings between all consonants, with the result that *multiple, competing* coupling relations are specified in the coupling graphs for onsets [10]. In contrast, we hypothesize that a vowel is only coordinated directly with its first coda consonant, and that this *coda relation* is an anti-phase coordination. This hypothesis finds support in developmental data indicating an early preference for CV (in-phase) over VC (anti-phase) syllable production (e.g. [57]). Figure 2 (top) displays the coupling graph for "spot", in which both the tongue tip (fricative) gesture for /s/ and the lip closure gesture for /p/ are coupled in-phase to the tongue body (vowel) gesture, while they are also coupled anti-phase to one another, and in which the vowel gesture is coupled anti-phase to the tongue tip gesture for /t/; Figure 2 (bottom) shows the pattern of gestural activations (gestural score) that results when this graph is input to the planning model.

The planning oscillator model has provided a promising account of intergestural phasing patterns within and between syllables, capturing both the mean relative phase values (e.g., [11]) and the variability of these phase values observed in actual speech data (e.g., [12]). The emergence of cohesive intergestural relative phasing patterns in the model is a consequence of the magnet-like properties of entrainment (frequency & phase locking) that characterize nonlinear ensembles of coupled oscillators, and the different patterns displayed for different structural elements reflect corresponding differences in the topologies of the coupling graphs used to represent those elements.

While the planning oscillator model has offered a promising account of the lexical relation between syllabic and gestural structure, we have only recently begun to be apply it to modeling the influence of prosody on articulation. In previous work [13], we introduced a prosodic gestural component, the *prosodic gesture* ($\pi$-gesture) into the task-dynamic model that has been useful in understanding the temporal lengthening of gestures in the vicinity of phrasal boundaries. Unlike constriction gestures, $\pi$-gestures operate *transgesturally* during a relatively localized portion of an utterance (e.g., near phrasal boundaries). While they are active, $\pi$-gestures slow the articulation rate of all constriction gestures that are themselves active during this time. In this earlier work, $\pi$-gestures were used to nonlinearly and locally time-warp the gestural score once it had been specified by the intergestural level of our model. In recent work, we have extended the time modulation properties of $\pi$-gestures to a more general class of *temporal modulation gestures* ($\mu_T$-gestures), and have begun to incorporate these $\mu_T$-gestures into the planning oscillator model itself.

In section 2.1 below, we review the manner in which planning oscillator dynamics have been specified to model intergestural timing patterns both within and between syllables. In section 2.2, we describe recent work in which this model has been generalized to provide a dynamics of temporal patterning at levels of the prosodic hierarchy above the syllable (e.g., foot and phrase). Finally, in section 3, we describe how the planning oscillator model has been extended to the levels of foot-syllable dynamics (section 3.1), and how temporal modulation gestures have been incorporated into, and serve to modulate, these dynamics to model the within-foot production of stressed and unstressed syllables (section 3.2).

## 2. Planning oscillators: Gestural, foot, and phrase

### 2.1. Coupling graphs and intergestural phasing

We extended Saltzman & Byrd's [49] task dynamic model of intergestural phasing, which specified an interoscillator coupling function between two gestural *planning oscillators*, to the case in which multiple (more than two) gestural oscillators are allowed to interact (and potentially compete) in shaping the steady-state pattern of intergestural phase differences [42], [54], [53]. This multi-oscillator ensemble defines a dynamics of gestural score formation in that, once the ensemble reaches an entrained steady-state of relative phasing, the waveform of each oscillator is used to trigger the activation function of that oscillator's associated constriction gesture.

As in the earlier model of Saltzman and Byrd [49], each oscillator in the *N*-oscillator ensemble is specified by the following intrinsic, second-order, limit-cycle dynamics:

$$\ddot{\boldsymbol{x}} = \ddot{\boldsymbol{x}}_I(\boldsymbol{x},\dot{\boldsymbol{x}}) \tag{1a}$$

where $\boldsymbol{x}, \dot{\boldsymbol{x}},$ and $\ddot{\boldsymbol{x}}$ are *Nx1* vectors of position, velocity, and acceleration, respectively; and $\ddot{\boldsymbol{x}}_I(\boldsymbol{x},\dot{\boldsymbol{x}})$ is the *Nx1* vector of uncoupled, oscillator-state-dependent intrinsic accelerations, for which:

$$\ddot{x}_{I,i} = -\alpha_i \dot{x}_i - \beta_i x_i^2 \dot{x}_i - \gamma_i \dot{x}_i^3 - \omega_{0i}^2 x_i \tag{1b}$$

($i$ = 1, 2, ..., $N$), where $\alpha_i$, $\beta_i$, and $\gamma_i$ are linear, nonlinear (van der Pol), and nonlinear (Rayleigh) damping coefficients, respectively; and $\omega_{0i}$ is the oscillator's linear natural frequency. In these simulations, in which all oscillators are entrained in a *1:1* frequency relationship, $\omega_{0i} = 1$, $-\alpha_i = \beta_i = \omega_{0i}$, and $\gamma_i = 1/\omega_{0i}$. Unless noted otherwise, these parameter values are used for all simulations reported in this paper.

The oscillator ensemble's ongoing Cartesian-coordinate state vector $(\boldsymbol{x}, \dot{\boldsymbol{x}})$ is transformed into a corresponding *Nx1* ongoing radial-coordinate state vector $(\boldsymbol{\phi}, \boldsymbol{A})$ that defines the sets of oscillator phases and amplitudes, respectively, where:

$$\phi_i = -\tan^{-1}\left(\frac{\dot{x}_i/\omega_{0i}}{x_i}\right) \tag{2a}$$

$$A_i = \sqrt{x_i^2 + \left(x_i/\omega_{0i}\right)^2} \tag{2b}$$

for $i$ = 1, 2, ..., $N$. The choice of which planning oscillators to couple to one another, along with the strengths of the corresponding interoscillator coupling functions, creates a "wiring diagram" that can be represented by a *system graph*. In this formulation, the task space of the oscillator ensemble is represented as a graph in which the *node* variables represent oscillator phases, $\phi_i$, and the internode *edges* represent the associated interoscillator relative phases, $\psi_k$ (Fig. 3). More specifically, the relative phase, $\psi_k$, associated with the edge between a given pair of oscillator phase nodes, $\phi_i$ and $\phi_j$, is defined for the case of *1:1* frequency-locked oscillators according to a convention in which an oriented line (arrow) is drawn from node-*i* (the origin node) to node-*j* (the insertion node), and relative phase is specified as:

$$\psi_k = \phi_j - \phi_i \tag{3}$$

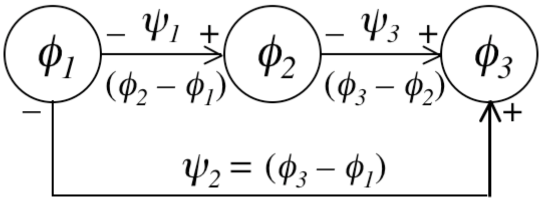

Figure 3. *Coupling graph for three 1:1 frequency-locked planning oscillators. Arrows are definitional only and do not indicate direction of coupling force.*

The *M* relative phases of a graph, and their relation to the corresponding set of *N* oscillator phases, can be expressed compactly using the graph's *MxN* edge-node *incidence matrix*, *C*, in which rows and columns correspond to edges and nodes, respectively (e.g., [6], [22], [58]):

$$\boldsymbol{\psi} = C\boldsymbol{\phi} \tag{4}$$

where $\boldsymbol{\psi}$ is the *Mx1* vector of relative phases, and $\boldsymbol{\phi}$ is the *Nx1* vector of individual oscillator phases. In particular, the $i^{th}$ row of *C* corresponds to the $i^{th}$ edge of the graph and contains the elements, $c_{ij}$, which equal +1 when node-*j* is the end-node for edge-*i*, equal −1 when node-*j* is the initial-node for edge-*i*, and equal 0 otherwise. The incidence matrix defines what is called in the movement science and robotics literature the *forward kinematic model* from component variables to task variables. In the present context, *C* represents the mapping from ongoing component-level oscillator phases to ongoing task-level interoscillator relative phases.

The task-space dynamics are defined by assigning a coupling function to each edge that is a function of that edge's relative phase. These relative phase coupling forces are defined in first-order point attractor terms:

$$\dot{\boldsymbol{\psi}}_{\boldsymbol{T}} = \boldsymbol{f}_{\boldsymbol{T}}(\boldsymbol{\phi}) \tag{5}$$

where $\dot{\boldsymbol{\psi}}_{\boldsymbol{T}}$ is the *M x 1* vector of task-specific state (relative phase) velocities; $\boldsymbol{\phi}$ is the *N x 1* vector of oscillator phases; and $\boldsymbol{f}_{\boldsymbol{T}}(\boldsymbol{\phi}) = \eta\Lambda\sin(\boldsymbol{\psi} - \boldsymbol{\psi}_{\boldsymbol{0}})$, the *M x 1* vector of task-space, phase-dependent forcing functions, for which: $\Lambda$ = a diagonal *MxM* matrix of normalized edge coupling strengths, with edge-*i's* value expressed as $0 \leq -\lambda_i \leq 1$; $\eta$ = global scaling value for coupling strengths; $\sin(\boldsymbol{\psi} - \boldsymbol{\psi}_{\boldsymbol{0}})$ = an *Mx1* vector of coupling forces; and $\boldsymbol{\psi}_{\boldsymbol{0}}$ = an *Mx1* vector of *target* relative phases. This task-space equation of motion (Eq. 5) defines bidirectional, symmetric coupling forces between the members of each oscillator pair. Unless otherwise noted, for all simulations reported in this paper, $\Lambda$ = I (the identity matrix) and $\eta$ =1.0.

These relative phase forces are used to derive acceleration coupling forces for the system of component oscillators (Eq. 1) in several steps. Differentiating Equation 4 with respect to time, we first derive the forward kinematic relation between phase-velocities and phase-difference-velocities:

$$\dot{\boldsymbol{\psi}} = C\dot{\boldsymbol{\phi}} = J_{\psi}\dot{\boldsymbol{\phi}} \tag{6}$$

where $J_{\psi}$ denotes the MxN Jacobian matrix of partial derivatives, $[\partial\psi_i/\partial\phi_j]$, of the forward model from oscillator phase to relative phase; because the elements of C are constants, $J_{\psi}$ is simply equal to C. We then use this expression to derive the vector of phase coupling forces, $\dot{\boldsymbol{\phi}}_{\boldsymbol{T}}$, from the task-space vector of relative phase forces specified by Equation 5:

$$\dot{\boldsymbol{\phi}}_{\boldsymbol{T}} = J_{\psi}^{+}\dot{\boldsymbol{\psi}}_{\boldsymbol{T}} = J_{\psi}^{+}\boldsymbol{f}_{\boldsymbol{T}} \tag{7}$$

where $J_{\psi}^{+}$ is the NxM pseudoinverse of $J_{\psi}$. Next, this phase coupling vector is transformed into a corresponding task-specific acceleration coupling vector for the planning oscillators:

$$\ddot{\boldsymbol{x}}_{\boldsymbol{T}} = J_x\dot{\boldsymbol{\phi}}_{\boldsymbol{T}} \tag{8}$$

where $J_x$ is the NxN diagonal Jacobian matrix whose elements $J_{x,ii} = (\partial\dot{x}_i/\partial\phi_i)$, for which: $\dot{x}_i = -\omega_{0i}A_i\sin\phi_i$; and hence $\partial\dot{x}_i/\partial\phi_i = -\omega_{0i}A_i\cos\phi_i$.

Finally, these task-specific acceleration coupling terms are added to the right-hand side of the equation of motion for the planning oscillators' intrinsic dynamics (Eq. 1a) to define the overall system dynamics for the planning oscillators:

$$\ddot{\boldsymbol{x}} = \ddot{\boldsymbol{x}}_{\boldsymbol{I}} + \ddot{\boldsymbol{x}}_{\boldsymbol{T}} \tag{9}$$

When we implemented the system graph (Figure 4) proposed by Browman & Goldstein [10] that was based on their macroscopic observations of the temporal organization of gestures in a syllable, we found that the model automatically produced the systematic differences found empirically between intergestural timing behavior in onsets and codas, both in their mean values (i.e., the "C-center" behavior displayed by onset clusters but not coda clusters [7], [24], [11]) and their variability (less variability in syllable onsets than in codas, and even more variability when the consonants are heterosyllabic [12] [heterosyllabic elements are not shown in Figure 4]) [42], [54]. We have recently shown that these differences between intergestural timing behavior in onsets and codas can be related rigorously to topological properties of the system graph: the C-center behavior can be explained by the *loop constraint equations* (e.g., [6], [22], [55], [58]) that are derivable in graph theory from the geometric structure of a system's incidence matrix (see Eq. 4 above), and the variability differences can be captured by a quantitative index of internode graph *connectivity* [53].

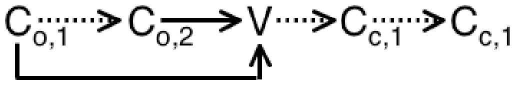

Figure 4. *System graph for CCVCC syllable. $C_{o,i}$ and $C_{c,i}$ denote the $i^{th}$ consonant in the syllable's onset and coda, respectively; solid and dashed edges denote in-phase ($0^o$) and anti-phase ($180^o$) target relative phases, respectively (see also Fig.2 for a related example)*

### 2.2. Harmonic entrainment (*1:n*) between nested phrase and foot oscillators.

In a now-classic experiment, Cummins and Port [15] investigated a speech production task in which a phrase such as "big for a duck" (underlining indicated the phrase's stressed syllables) was repeated rhythmically in time. An auditory metronome (high pitch tone) specified a succession of global cycles that began with each onset of the phrase-initial stressed syllable, and within each trial a low-pitch tone was presented at a given phase of the global cycle. In essence, the auditory pattern was generated by two metronomes that were *1:1* frequency-locked, with a phase-locking value (targeted relative phase) specified by the experimenter for each trial. Using a synchronization-continuation paradigm, subjects were asked on each trial to synchronize the phrase-initial syllable with the high-tone and the phrase-final syllable with the intervening low-tone (synchronization task), and then attempted to continue producing the same temporal pattern after the tones had ceased (continuation task). Although a continuous range of target phases were probed over the course of the experiment, speakers showed clear biases toward producing relative phasings (defined as fractions of the global cycle) of 1/2, 1/3, and 2/3, despite the goal of reproducing the task-demanded phasing patterns. Cummins and Port [15] hypothesized that these constrained phasing relationships were the result of frequency-locking between a pair of metrical oscillators, one at the foot level (feet are defined by the intervals between stressed syllable) and one at the phrase level. Two foot-cycles per phrase-cycle results in the phrase's second stressed syllable being produced with a relative phase of 1/2; three feet per phrase-cycle results in a relative phasing of either 1/3 or 2/3.

The authors did not verify their hypothesis using a coupled oscillator model, however. We have recently done this [41], and have not only replicated their results successfully but, in doing so, have demonstrated that the planning oscillator model can generalize to constituent levels higher than constriction-gestural, where frequency ratios between successively higher levels are *n:1* ($n \neq 1$). Doing so required us to generalize to the multifrequency case the manner in which internode link variables, forward models, and incidence matrices are defined (see section 2.1). In our simulations of the Cummins and Port [15] data, we assumed that the externally imposed rhythmic structure provided by the dual metronomes in the experimental paradigm (*1:1* frequency-locking, with a variable interoscillator phasing) acted to perceptually induce a *n:1* frequency-locked pattern in the speaker/listener's "internal" set of foot and phrase planning oscillators, respectively. We thus assumed the ability of the subject to optimally induce the temporal structure of perceived prominences, whether in an experimentally imposed metronome pattern or in a series of prominences produced by an interlocutor that are cued by $F_0$, gestural duration and spatial extent, sonority, vocal effort, etc. We did not model this induction process explicitly; rather, our assumption is based on the results of others (e.g., [2], [3], [33], [32], [37], [39]) who have designed "adaptive oscillator" models that, given a complex temporal structure of input beats, can induce these structures and represent them in the dynamics of a multifrequency oscillator ensemble.

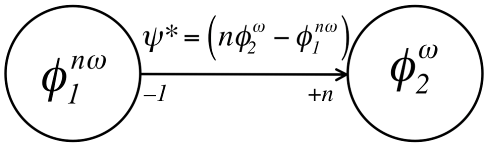

Figure 5. *Coupling graph convention for defining phases, $\phi$, and generalized relative phase, $\psi$*, in oscillator pair with n:1 frequency-locking. Phase superscripts denote oscillation frequency, with $\omega$ denoting baseline intrinsic frequency; subscripts denote oscillator identity.*

Because this is a multifrequency system, a *generalized* relative phase, $\psi^*$, (e.g., [49]) is used to specify phase-locking. For a task in which two oscillators are entrained with *n:1* frequency-locking, $\psi^*$ is specified according to the coupling graph conventions illustrated in Figure 5. The relationships between node-based oscillator phases and edge-based generalized relative phases can be represented succinctly by the coupling graph's *M*x*N generalized incidence matrix*, $C^*$, a generalization of the standard incidence matrix:

$$\boldsymbol{\psi}^* = C^* \boldsymbol{\phi} \qquad (10)$$

where $\boldsymbol{\psi}^*$ is an $M$x1 vector of edge-based generalized relative phases, and $\boldsymbol{\phi}$ is an $N$x1 vector of node-based oscillator phases. The elements of $C^*$ are defined by: $c^*_{ij} = +n$ when node-*j* is the terminal node; $= -1$ when node-*j* is the initial node; and = 0, otherwise.

In our simulation, we assume that four internal oscillators are induced during the experimental speech task (see figure 6). Of these four oscillators, the two *metronome* oscillators—the phrase metronome ($M_P$; $\omega_{MP}$ = 1 Hz) and the foot metronome ($M_F$; $\omega_{MF}$ = 1 Hz)—are bidirectionally coupled, and *1:1* frequency-locked with a target relative phase that varies from trial to trial. The two internal *planning* oscillators—the foot planning oscillator ($P_F$; $\omega_{PF}$ = 2 Hz or 3 Hz, depending on the trial) and phrase planning oscillator ($P_P$; $\omega_{PP}$ = 1 Hz)—are bidirectionally coupled, and *2:1* or *3:1* frequency-locked with generalized relative phases $\psi^* = (\phi_F - 2\phi_P)$ or $(\phi_F - 3\phi_P)$, and target generalized relative phase, $\psi_0^*$, = 0°. Finally, the phrase metronome unidirectionally drives the phrase planning oscillator with *1:1* frequency-locking and a target generalized relative phase = 0°; and the foot metronome unidirectionally drives the foot planning oscillator with 1:2 or 1:3 frequency-locking, also with a target generalized relative phase = 0°. Since bidirectional coupling has default status in our model, we specify such unidirectional coupling by setting to zero the coupling forces from the planning oscillators to the metronomes.

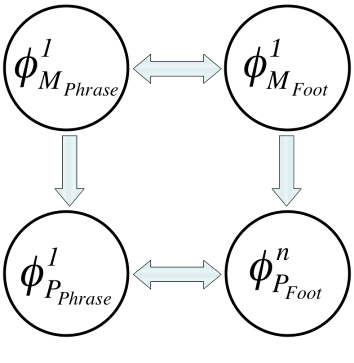

Figure 6. *System of four 'internal' oscillators used to model the data of Cummins and Port [15]. Subscripts denote oscillator identity; superscripts denote intrinsic oscillator frequency. Double- and single- headed arrows denote bidirectional and unidirectional interoscillator coupling, respectively. (see text for further details).*

Each simulation trial was run with a randomly chosen target relative phase between the two metronome oscillators, and small amounts of additive noise were included to stochastically perturb all oscillators on each time step. The results replicated the patterns found by Cummins and Port [15]: there was a systematic bias away from the "target" phases specified by the metronome and toward the phases 1/2, 1/3, or 2/3. This was due to the fact that mutual entrainment of the planning oscillators "pulled" the system away from the "target" phases specified by the metronome and toward the phases 1/2 (for the 1:2 frequency-locked planning oscillators), or toward 1/3 or 2/3 (for the 1:3 frequency-locked planning oscillators).

## 3. Polysyllabic shortening: Foot and syllable oscillators; Temporal modulation gestures

Durational processes operating on the level of the word have been extensively investigated (see overview in [59]). One of the best known temporal phenomena is *polysyllabic shortening* (also called *stress-timed shortening* by Beckman and Edwards, 1990)—the shortening of the stressed syllable as the number of syllables in a word increases, e.g, the successive shortening of "speed" in "speedy" and "speedily" (e.g., [35], [29], [25], [36] for Swedish; [17] for Estonian, as reported in [19]). Recent work by Kim & Cole [27], [28] has further demonstrated that while with the increase of the number of syllables in a foot the stressed syllable shortens, the duration of the unstressed syllables remains unchanged. This leads to the duration of the foot increasing with the number of syllables, while the duration of the stressed syllable decreases.

### 3.1. Harmonic entrainment (*1:n*) between nested foot and syllable oscillators.

O'Dell and colleagues [43], [44] have described a dynamical model with two oscillators—a syllable and foot oscillator—that provided a crucial first step in simulating the dynamics of polysyllabic shortening. Their model is based on the equation of motion for the generalized relative phase between foot and syllable *phase-oscillators*, and was derived using the *Averaged Phase Difference* technique of Kopell [30]. When he applied this model to Eriksson's [18] cross-language regression analysis of foot duration vs. number of syllables-per-foot, O'Dell was able to simulate the manner in which foot duration increased with the number of syllables or, equivalently, the manner in which *average* syllable duration decreased with increasing numbers of syllables in the foot. The key theoretical result was that the behavior of foot duration as a function of number of syllables depended on the degree of *asymmetry* of the coupling forces between the syllable and foot oscillators. Roughly speaking, in O'Dell and Nieminen's [44] model, 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; the ratio of inter-level coupling strengths determines the degree of relative inter-oscillator dominance in this competition. For English, the coupling from foot to syllable dominated the coupling from syllable to foot, and the ratio of coupling strengths could be specified as a function of the regression parameters in Eriksson's analyses.

We have reproduced O'Dell and Nieminen's results using our planning oscillator model for feet with 2 or 3 syllables. In all simulations, the foot oscillator's intrinsic angular frequency parameter, $\omega_F$, = 1 rad/s, and the syllable oscillator's intrinsic frequency, $\omega_\sigma$, = 2 rad/s. This specification is based on the assumption, following Hayes [23], that the default foot in English has two syllables. Cross-linguistically phonologists have also postulated the binary foot as the default (e.g., [47], [26]; see also [34] for phonetic evidence of disyllabic sequences as basic rhythmic units in Czech, Finnish, Estonian, Serbo-Croatian). Additionally, the directionally symmetric interoscillator coupling strengths specified by equation 5 are replaced by asymmetric ones—the coupling strength from foot to syllable oscillator, $\lambda_{F\sigma}$, = 5, and the strength of coupling from syllable to oscillator, $\lambda_{\sigma F}$, = 1. For the 2 syllable case, generalized relative phase, $\psi^*$, = $(\phi_\sigma - 2\phi_F)$; for the 3 syllable case, $\psi^* = (\phi_\sigma - 3\phi_F)$; in both cases, target generalized relative phase, $\psi_0^*$, = 0°. The results are shown in Figure 7. The durations of all syllables in the 2-syllable and 3-syllable feet, respectively, = 3.1s, and 2.5s. Adding syllables temporally expands foot durations and, in turn, feet also provide a temporally compressive "frame" that reduce syllable durations as syllables are added.

As O'Dell and Nieminen [44] point out, however, one drawback to these results is that all syllables in a given foot have equal durations. Such within-foot temporal symmetry is typically broken in many languages by the longer duration of the stressed syllable relative to the durations of the remaining unstressed syllables. These authors hypothesized that such broken symmetry could be provided by a ‘stress function’ that would depend on the ongoing phase of the foot oscillator and that would “slow the syllable down in the vicinity of some particular phase representing stress.” (p. 1077). In previous work, we have used related “clock-slowing” *prosodic gestures* ($\pi$-gestures) to model the temporally local slowing of articulation rate that occurs in the vicinity of phrasal boundaries [12]. In that work, however, the $\pi$-gestures did not directly alter the dynamics of intergestural timing; rather, they were used to locally time-warp the gestural score once it had been specified by the intergestural level of the model.

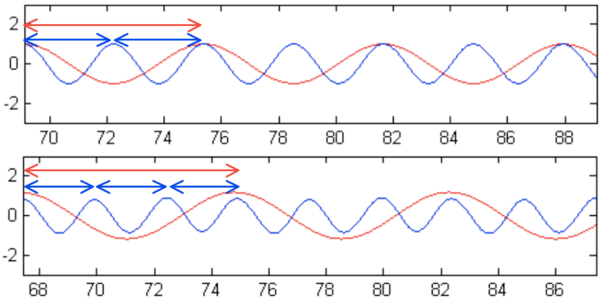

Figure 7. *Steady-state patterns of (slow) foot and (fast) syllable oscillators, with asymmetrical (foot-dominant) coupling between foot and syllable oscillators. Top panel: 2 syllables per foot, with both syllable durations = 1/2 foot duration; Bottom panel: 3 syllables per foot, with all syllable durations = 1/3 foot duration. Horizontal axis = time (s); vertical axis = oscillator position (arbitrary units). Each panel starts at* $\phi_F = 0$ *rad.*

In the following section, we introduce a more general class of *temporal modulation gestures* ($\mu_T$-gestures; of which $\pi$-gestures are a special case), and describe how these $\mu_T$ -gestures have been incorporated into the dynamics of our model’s planning oscillator ensemble to create appropriate durational differences between stressed and unstressed syllables, thereby breaking the temporal symmetry of syllables nested within feet.

### 3.2. Incorporating stress into the foot-syllable oscillator ensemble

#### *3.2.1. Temporal modulation ($\mu_T$) gestures*

We used a “clock”-slowing $\mu_T$ -gesture to slow the rate of phase-flow in the foot-syllable oscillator ensemble during the first, stressed syllable of 2- and 3-syllable feet. More specifically, we applied the $\mu_T$ -gesture during the interval of the foot’s phase cycle $0 \le \phi_F < \kappa$, where $\kappa = (1/n)2\pi$ denotes the fraction of the foot cycle taken by each syllable, and $n = 2$ or 3, respectively (Figure 8 displays the 3-syllable case). A half-cosine ramping function was used to allow the $\mu_T$ -gesture‘s activation value, $a_{\mu_T}$, to smoothly increase from 0 ($\mu_T$ is “off”) to 1 ($\mu_T$ is “on”) at the stressed syllable’s onset and to decrease from 1 to 0 before the onset of the following unstressed syllable; the duration of these ramping functions, expressed in units of $\phi_F$, is $\rho\kappa$, where $\rho = .2$. After returning to zero at the onset of the first unstressed syllable (where $\phi_F = \kappa$), $\mu_T$ remains off during the unstressed syllables until the beginning of the next foot cycle (i.e., $\mu_T$ remains off over the interval $\kappa \le \phi_F < 2\pi$) . The activation function for the $\mu_T$ -gesture, $a_{\mu_T}(\phi_F)$, is expressed as follows:

$$a_{\mu_T} = \begin{cases} -.5\cos\left(\phi_F\left(1/\rho\kappa\right)\pi\right), & \text{if } 0 \le \phi_F < \rho\kappa \\ 1, \ \text{if } \rho\kappa \le \phi_F < (1-\rho)\kappa \\ .5\cos\left(\left[\phi_F - (1-\rho)\kappa\right]\left(1/\rho\kappa\right)\pi\right), & \text{if } (1-\rho)\kappa \le \phi_F < \kappa \\ 0, \ \text{if } \kappa \le \phi_F < 2\pi \end{cases} \tag{11}$$

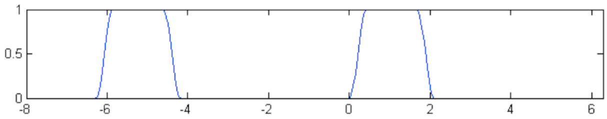

Figure 8. *Successive* $\mu_T$*-gestures specified as function of foot oscillator phase,* $\phi_F$*, for 3-syllable foot. The gesture’s activation,* $a_{\mu_T}$ *= 1 during the stressed syllable, and = 0 during the unstressed syllables. Since* $\omega_F$ *= 1 rad/s, the period,* $T_F$*, = 2*$\pi$ *s. Horizontal axis =* $\phi_F$ *(rad); vertical axis =* $a_{\mu_T}$ *(arbitrary units)*

The $\mu_T$–gesture’s ongoing activation value, $a_{\mu_T}$, was used to slow the “clock-rate” of the foot-syllable planning oscillator ensemble by modulating the ongoing values of the natural frequency parameters of both the foot ($\omega_{o,F}$) and syllable ($\omega_{o,\sigma}$) oscillators according to:

$$\omega^*_{o,i} = (1 - \delta a_{\mu_T})\omega_{o,i} \tag{12}$$

where $i = F$ or $\sigma$; and $\delta = .5$ denotes the *strength* of the $\mu_T$ –gesture and is proportional to the degree of slowing induced in the oscillator ensemble. As in the previous section, asymmetric interoscillator coupling strengths were used, with $\lambda_{F\sigma}, = 5$, $\lambda_{\sigma F}, = 1$.

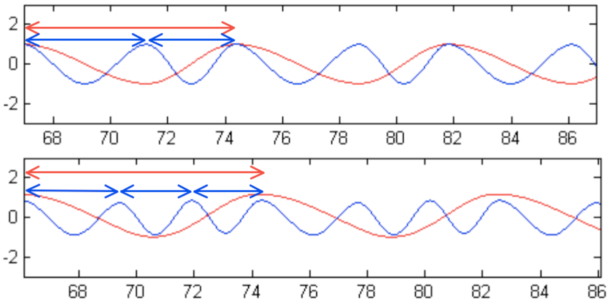

Figure 9. *Steady-state patterns of (slow) foot and (fast) syllable oscillators, with asymmetrical (foot-dominant) coupling between foot and syllable oscillators, and a* $\mu_T$ *-gesture added to the stressed syllable. Top panel: 2 syllables per foot; Bottom panel: 3 syllables per foot. Horizontal axis = time (s); vertical axis = oscillator position (arbitrary units). Each panel starts at* $\phi_F = 0$ *rad. (See text for further details)*

The results of these simulations can be seen in Figure 9. Syllable durations in the 2-syllable foot are 4.3s (stressed) and 3.1s (unstressed); syllable durations in the 3-syllable foot are 3.4 (stressed) and 2.5s (unstressed). As was the case without the $\mu_T$ –gesture, adding syllables temporally expands foot durations; and feet provide a temporally compressive "frame" that reduces syllable durations as syllables are added. Additionally, and crucially, the addition of the $\mu_T$–gesture in the present simulation breaks the temporal symmetry of the syllables nested within the foot cycles—stressed syllables are longer than unstressed.

### 3.2.2. *Modulation of inter-level coupling strength*

In the previous section, we described simulations in which the temporal properties of stressed and unstressed syllables within a foot emerged from the interaction between the asymmetrically coupled foot and syllable planning oscillators (foot dominates syllable), and a temporal modulation ($\mu_T$) gesture that is active during the stressed syllable. Such results were encouraging in that they provided an account of: a) foot lengthening with increasing number of syllables; b) longer duration of the stressed syllable compared to the unstressed syllables; and c) shortened duration of the stressed syllable with increasing number of syllables in the foot. However, it also resulted in shortened durations of the *unstressed* syllables with increasing number of syllables per foot, which is contrary to the data patterns reported by Kim & Cole [27], [28]. As was mentioned earlier, these researchers showed that, although stressed syllables increasingly shorten within feet as syllables are added, the durations of the remaining unstressed syllables do *not* change with increasing numbers of syllables per foot.

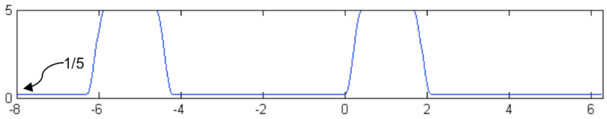

Figure. 10. *Trajectory for the foot-syllable oscillator coupling strength ratio, $\varepsilon(\phi_F) = \lambda_{F\sigma}(\phi_F)/\lambda_{\sigma F}(\phi_F)$, specified as a function of foot oscillator phase, $\phi_F$, for the 3-syllable foot. $\varepsilon(\phi_F) = 5$ during the stressed syllable, and = 1/5 during the unstressed syllables. Since $\omega_F = 1$ rad/s, the period, $T_F$, = $2\pi$ s. Horizontal axis = $\phi_F$ (rad); vertical axis = $\varepsilon(\phi_F)$ (dimensionless units).*

We interpret this phenomenon in the context of our model as resulting from a weakening of the foot oscillator's temporal compression on the syllable oscillator for unstressed syllables relative to the stressed syllable. In this section, we report the results of recent simulations in which such within-foot asymmetry in foot-to-syllable temporal compression is implemented using a corresponding within-foot modulation of the ratio of coupling strengths between the foot and syllable oscillators. This modulation is viewed as a parameter-dynamic process in which the coupling strength ratio, $\varepsilon = \lambda_{F\sigma}/\lambda_{\sigma F}$, and the coupling strengths themselves, are modulated as functions of the ongoing phase of the foot oscillator, $\phi_F$. We define the "target" coupling ratio for the stressed syllable as $\varepsilon_{stress}$. To minimize the number of free system parameters, we constrain the target coupling ratio for the unstressed syllables, $\varepsilon_{unstress}$, to equal ($1/\varepsilon_{stress}$); in addition, we yoke the phasing of $\varepsilon(\phi_F)$ to that of $a_{\mu_T}(\phi_F)$ (equation 11 above) as follows:

$$\varepsilon(\phi_F) = \left(\varepsilon_{stress} - \left(1/\varepsilon_{stress}\right)\right) a_{\mu_T}(\phi_F) + \left(1/\varepsilon_{stress}\right) \qquad (13)$$

In the previous section, we used coupling strength values ($\lambda_{F\sigma} = 5$, $\lambda_{\sigma F} = 1$) and a coupling ratio ($\varepsilon = [\lambda_{F\sigma}/\lambda_{\sigma F}] = 5$) that remained constant throughout the simulation. In this section, we adopt these coupling strength values as the target coupling strengths for the stressed syllable, i.e., $\lambda_{F\sigma,stress} = 5$, $\lambda_{\sigma F,stress} = 1$, giving $\varepsilon_{stress} = 5$; and $\lambda_{F\sigma,unstress} = 1$, $\lambda_{\sigma F,unstress} = 5$, giving $\varepsilon_{unstress} = 1/5$ for the unstressed syllables. The trajectory of $\varepsilon(\phi_F)$ in these simulations is shown in Figure 10 for the 3 syllable per foot case.

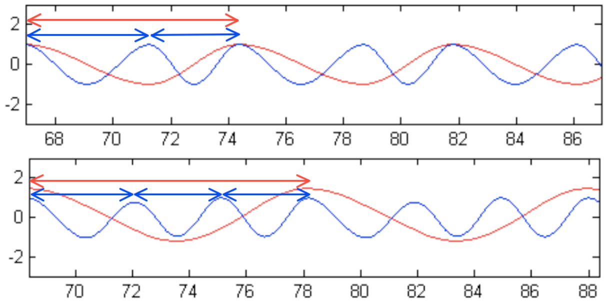

Figure 11. *Steady-state patterns of (slow) foot and (fast) syllable oscillators, with coupling strength ratio, $\varepsilon(\phi_F)$ varied between the stressed syllable and the remaining, unstressed syllables, and with a $\mu_T$-gesture added to the stressed syllable. Top panel: 2 syllables per foot; Bottom panel: 3 syllables per foot. Horizontal axis = time (s); vertical axis = oscillator position (arbitrary units). Each panel starts at $\phi_F = 0$ rad. (See text for further details)*

The results of these simulations are shown in Figure 11. Syllable durations in the 2-syllable foot are 4.3s (stressed) and 3.1s (unstressed); syllable durations in the 3-syllable foot are 3.4 (stressed) and 3.1s (unstressed). It can be seen that the within-foot modulation of coupling strength ratio, combined with the effects of the $\mu_T$–gesture, had the desired effect of producing invariant durations of unstressed syllables regardless of the number of syllables in a foot, consistent with the data reported by Kim & Cole [27], [28]. Importantly, this invariance is obtained while maintaining the previously modeled patterns of foot-syllable temporal elasticity (i.e., increased foot duration, decreased average syllable duration, and decreased stressed syllable duration with increasing number of syllables), and of within-foot temporal asymmetry between (longer) stressed and (shorter) unstressed syllables.

## 4. Concluding Remarks

In the preceding pages, we have reviewed recent developments of the task-dynamic model of speech production that have allowed us to model several aspects of prosodic structure within a unified dynamical framework. Key aspects of these developments have been to specify patterns of intergestural relative phasing according to the dynamics of an ensemble of nonlinear, limit-cycle planning oscillators, and to incorporate parameter-dynamic processes that modulate the ongoing values of a subset of the ensemble's parameters (intrinsic oscillator frequencies, and interoscillator coupling strength ratios). At the level of modeling intergestural patterning within and between syllables—the level at which we have done most of our work to date—we have applied the concepts and tools of dynamics and graph theory to the behavior of the planning oscillator ensemble to provide a remarkably parsimonious account of empirically demonstrated phenomena in both the mean values and variabilities of C-C, C-V, and V-C timing patterns ([42], [53]). Our recent work described above on nested foot and phrase oscillators, and on nested syllable and foot oscillators, provides additional support for the hypothesis that a hierarchically and harmonically nested rhythmic system can account for articulatory patterns shaped by the relations between constituents in neighboring levels of the prosodic hierarchy [41]. Taken together with the work of others (e.g., [2], [3], [15], [38], [43], [44]), these results demonstrate the value of representing an utterance's central "clock" as a mutually entrained oscillatory ensemble (cf., alternative views in which a *single* oscillator serves as a clock that triggers activity of lower level units at certain phases of the clock, e.g., the activation of constriction gestures according to the triggering phases of a superordinate syllable oscillator; [1], [31], [61]).

In addition, we have reviewed recent progress in modeling the influence of prosodic factors such as stress by modulating a subset of the ensemble's parameters (oscillator frequency, inter-level coupling strength ratios) as autonomous functions of the ensemble's state. Such an approach provides a promising account of the durational properties of both stressed and unstressed syllables in polysyllabic shortening. Our use of temporal modulation gestures ($\mu_T$ –gestures) to implement the durational distinction between stressed and unstressed syllables is similar in spirit to the work of others (e.g., [1], [4], [31], [44], [46], [51], [61]) who posited that an utterance's local speaking rate is modulated according to its segmental, syllabic and/or foot structure. Again, however, we would insist that the 'clock' being modulated in such instances is that defined by a multilevel oscillator ensemble containing oscillators whose rhythmic periods harmonically subsume and constrain those of their immediately subordinate tiers.

## 5. References

[1] Bailly, G., Laboissière, R., & Schwartz, J. L. (1991). Formant trajectories as audible gestures: An alternative for speech synthesis. *Journal of Phonetics*, *19*, 9-23.

[2] Barbosa, P. A. (2002). Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production. In *Proc. 1st International Conference on Speech Prosody*, Aix-en-Provence, France.

[3] Barbosa, P. A. (2007). From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication*, *49*, 725-742.

[4] Barbosa, P., & Bailly, G. (1994) Characterisation of rhythmic patterns for text-to-speech synthesis. *Speech Communication*, *15*, 127-137.

[5] Beckman, M. E., & Edwards, J. R. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston, & M. E. Beckman, (Eds.). *Papers in laboratory phonology I: Between the grammar and the physics of speech*. Cambridge: Cambridge University Press. Pp. 152-178.

[6] Bollobás, B. (1998). *Modern graph theory*. New York, NY: Springer.

[7] Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica*, 45, 140-155.

[8] Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology: I. Between the grammar and the physics of speech*. (pp. 341-338) Cambridge, England: Cambridge University Press.

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

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

[11] Byrd, D. (1995). C-centers revisited. *Phonetica,* 52, 285-306.

[12] Byrd, D. (1996). A phase window framework for articulatory timing. *Phonology*, *13*, 139-169.

[13] Byrd, D., & Saltzman, E. (2003). The elastic phrase: Dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*, 149-180.

[14] Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, *35*, 210-243.

[15] Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*, 145-171.

[16] Di Cristo, A., & Hirst, D.J., (1996). Modelling French micromelody. *Phonetica* 43(1), 11-30.

[17] Eek, A., & Remmel. M. 1974. Context, contacts and duration. Preprints of speech communication seminar; vol. 1-2, pp. 187-192. Speech Transmission Laboratory, Stockholm.

[18] Eriksson, A. (1991). *Aspects of Swedish speech rhythm*. University of Göteborg, Göteborg.

[19] Fowler, C. A. 1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research.* 24: 127-139.

[20] Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. Arbib, (Ed.). *Action to Language via the Mirror Neuron System*. New York: Cambridge University Press, Pp. 215-249.

[21] Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*, 347-356.

[22] Harary, F. (1969). *Graph theory*. Reading, MA: Addison-Wesley.

[23] Hayes, B. (1980). *A metrical theory of stress rules*. Ph.D. dissertation. Department of Linguistics, Massachusetts Institute of Technology, Cambridge, MA.

[24] Honorof, D., & Browman, C. (1995). The center or edge: How are consonant clusters organized with respect to the vowel? In K. Elenius and P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences, vol. 3*. Stockholm: KTH and Stockholm University. Pp. 552-555.

[25] Huggins, A. W. F. (1975) On isochrony and syntax, In *Auditory Analysis and Perception of Speech*, Fant, G. & Tatham, M. A. A. (eds.), Academic Press, 455-464.

[26] Kager, R. (1989). *A metrical theory of stress and destressing in English and Dutch*. Dordrecht: Foris.

[27] Kim, H., & Cole, J. (2005). The stress foot as a unit of planned timing: Evidence from shortening in the prosodic

phrase. *Proceedings of Interspeech 2005*, Lisbon, Portugal.

[28] Kim, H., & Cole, J. (2006). "Evidence for rhythm shortening in American English as conditioned by prosodic phrase structure," to appear in the Proceeding of 42nd annual Meeting of the Chicago Linguistic Society

[29] Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America*, 54(4), 1102-1104.

[30] Kopell, N. (1988). Toward a theory of modeling central pattern generators. In A. H. Cohen, S. Rossignol, & S. Grillner, (Eds.). *Neural control of rhythmic movement in vertebrates*. New York: John Wiley & Son. Pp. 369-413.

[31] Laboissière, R., Schwartz, J.-L., & Bailly, G. (1991). Motor control for speech skills: A connectionist approach. In D. S. Touretzky, J. L. Elman, T. J. Sejnowski, & G. E. Hinton, (Eds.), Connectionist models. *Proceedings of the 1990 Summer School*. San Mateo, CA: Morgan Kaufmann. (pp. 319-327).

[32] Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106,* 119–159.

[33] Large, E. W., & Kolen, J. K. (1994). Resonance and the perception of musical meter. *Connection Science*, *6*, 177-208.

[34] Lehiste, I. (1970). *Suprasegmentals.* Cambridge, MA: M.I.T. Press.

[35] Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, *51*, 2018–2024.

[36] Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish, PILUS (Papers from the Institute of Linguistics, University of Stockholm) **21**,1-59.

[37] McAuley, J. D. (1994). Finding metrical structure in time. In M. C. Mozer, P. Smolensky, D. S. Touretsky, J. L. Elman, & A. S. Weigend, (Eds.). *Proceedings of the 1993 Connectionist Models Summer School*. Hillsdale, NJ: Erlbaum. (pp. 219-227).

[38] McCauley, J. D., & Dilley, L. C. (2006). Perceptual organization in intonational phonology: A test of parallelism. In *Proceedings of 10th Laboratory Phonology Conference*, Paris, France.

[39] McAuley, J. D., & Kidd, G. R. (1998) Effect of deviations from temporal expectations on tempo discrimination of isochronous tone sequences. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1786-1800.

[40] Nam, H. (in press). A competitive, coupled oscillator model of moraic structure: Split-gesture dynamics focusing on positional asymmetry. In J. Cole & J. Hualde (Eds.). *Papers in Laboratory Phonology IX.*

[41] Nam, H., Goldstein, L., & Saltzman, E. (2006). Dynamical modeling of supragestural timing. In *Proceedings of the 10th Laboratory Phonology Conference. Paris, France.*

[42] Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator of syllable structure. *Proc. XVth International Congress of Phonetic Sciences, Barcelona, 3-9 Aug 2003.*

[43] O'Dell, M. L. (1995). *Intrinsic timing in a quantity language*. Licentiate Dissertation, Department of Linguistics, University of Jyväskylä, Finland.

[44] O'Dell, M. L., & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey, (Eds.). *Proceedings of the XIVth International Congress of Phonetic Sciences, Vol. 2*. New York: American Institute of Physics. (pp. 1075-1078).

[45] Port, R. (1986). *Translating linguistic symbols into time.* (Research in Phonetics and Computational Linguistics, Rep. No. 5), Dept. Linguistics & Computational Linguistics, Indiana University. (pp. 156-173).

[46] Port, R. F., & Cummins, F. (1992). The English voicing contrast as velocity perturbation. In *Proceedings of the 1992 International Conference on Spoken Language Processing (ICSLP '92), vol. 2, Banff, Alberta, Canada* Edmonton, Canada: Priority Printing. (pp. 1311-1314).

[47] Prince, A. (1980). A metrical theory for Estonian quantity. *Linguistic Inquiry*, *11*, 511-562.

[48] Rossi, M., 1999. *L'intonation. Le système du français : description et modélisation*. Gap: Ophrys.

[49] Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, *19*, 499-526.

[50] Saltzman, E., Löfqvist, A., Kay, B., Kinsella-Shaw, J., & Rubin, P. (1998). Dynamics of intergestural timing: A perturbation study of lip-larynx coordination. *Experimental Brain Research*, *123*, 412-424.

[51] Saltzman, E., Löfqvist, A., & Mitra, S. (2000). "Clocks" and "glue"—Global timing and intergestural cohesion. In *Papers in Laboratory Phonology V*. (M. B. Broe & J. B. Pierrehumbert, editors). pp. 88-101. Cambridge, UK: Cambridge University Press.

[52] Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*, 333-382.

[53] Saltzman, E., Nam, H., & Goldstein, L. (submitted). Intergestural timing in speech production: The role of graph structure. *Human Movement Science.*

[54] Saltzman, E., Nam, H., Goldstein, L., & Byrd, D. (2006) The distinctions between state, parameter and graph dynamics in sensorimotor control and coordination. In M. Latash, & F. Lestienne, (Eds.). *Motor Control and Learning*. New York: Springer. Pp. 63-73

[55] Shearer, J. L., Murphy, A. T., & Richardson, H. H. (1971). *Introduction to system dynamics*. Reading, MA: Addison-Wesley.

[56] Sternad, D., Amazeen, E. L., & Turvey, M. T. (1996). Diffusive, synaptic, and synergetic coupling: An evaluation through inphase and antiphase rhythmic movements. *Journal of Motor Behavior*, *28*, 255-269.

[57] Stoel-Gammon, C. (1985). Phonetic inventories, 15-24 months: A longitudinal study. *Journal of Speech and Hearing Research, 18,* 505-512.

[58] Strang, G. (1986). *Introduction to applied mathematics*. Wellesley, MA: Wellesley-Cambridge Press.

[59] Turk, A. E. & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics, 28*, 397-440.

[60] Turvey, M. T. (1990). Coordination. *American Psychologist, 45*(8), 938-953.

[61] Vatikiotis-Bateson, E., Hirayama, M., Honda, K., & Kawato, M. (1992). The articulatory dynamics of running speech: Gestures from phonemes. In *Proc. 1992 Int. Conf. on Spoken Language Processing, vol. 2*. Edmonton, Canada: Priority Printing. (pp. 887-890).