Modeling the durational difference of stressed vs. unstressed syllables

Hosung Nam[1], Elliot Saltzman[1,2], 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

hosung.nam@haskins.yale.edu; esaltz@bu.edu; jelena.krivokapic@yale.edu; goldstein@haskins.yale.edu

Abstract

Speech production exhibits temporal coherence among speech gestures, and also systematic modulation of durational patterns as a function of the hierarchical level of prosodic structure, e.g., the foot. Intergestural coherence has been understood with reference to dynamic coupling within an ensemble of *planning oscillators*, and a coupled oscillator model of intergestural timing has been employed to simulate relative timing patterns observed inter-gesturally within and between syllables. In this paper, we integrate *temporal modulation gestures* into the oscillator model and modulate coupling parameters to simulate the durational asymmetry between stressed and unstressed syllables.

1. Introduction

Speech can be decomposed into a set of vocal tract constriction gestures ([4], [5], [6]). Each gesture can be viewed as an invariant, temporally extended unit of action, whose activation creates task-specific patterns of articulator motion that achieve linguistic gestural goals ([5], [22], [25]). The task-dynamic model of speech production ([22]) incorporated the theoretical tenets of articulatory phonology and provides a mathematical implementation of gesture-to-articulator mapping, in which articulators are appropriately coordinated for a given gesture. Input to the task-dynamic model is provided by a *gestural score*, which specifies the intervals of time during which particular constriction gestures are *active* in the vocal tract.

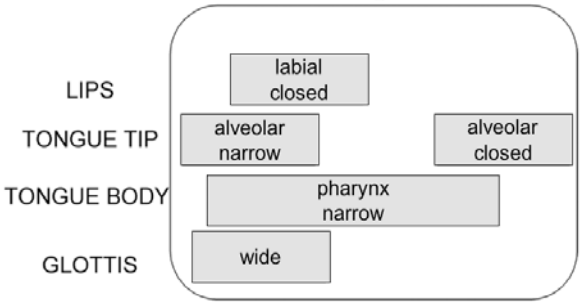

Figure 1: *Schematic gestural score for "spot", where the time intervals of gestural activation (gestural boxes) are specified at different organs (lips, tongue tip, tongue body, glottis).*

Each gesture that is activated by the gestural score is a dynamical system that controls a constricting device (lips, tongue tip, tongue body, velum, glottis) in the vocal tract. The goals of the constriction gestures are specified as targets in *tract variable* space (Figure 2) that represent the degrees and locations of the controlled constrictions. Lips, tongue tip, tongue body gestures are defined in terms of constriction location and degree, while VEL and GLO gestures are defined only by constriction degree. Each tract variable involves different articulators (e.g. bilabial gestures, LA and LP, use upper/lower lips and jaw motions to achieve their gestural goals). Some articulators are shared by different gestures (e.g. the jaw is shared by LP, LA, TTCL, TTCD, TBCL, and TBCD).

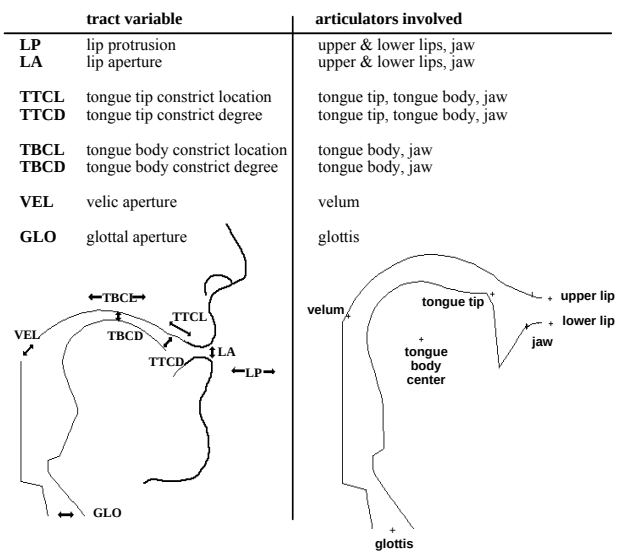

Figure 2: *Mapping between tract variables and their involving articulators. Tract variables are defined in constriction degree (LA, TTCD, TBCD, VEL, GLO), and location (LP, TTCL, TBCL) in some cases.*

Patterns of intergestural timing in gestural scores were originally specified explicitly by rule or by hand in the model's original formulation and, hence, could not account for the real-time temporal coherence among gestures that has been reported in the literature ([20], [21]). Recently, however, we extended the model to incorporate a dynamics of intergestural coordination from which these relative timing patterns and their coherence could emerge as implicit consequences of these dynamics. Specifically, we incorporated an ensemble of coupled, nonlinear limit cycle *planning oscillators* into the intergestural level of the model ([7], [19], [17], [24], [15], [23]). Each planning oscillator is associated with a given gesture in a specified lexical item, and the pair-wise coupling links are used to define a *coupling graph* for the lexical item.

In addition to intergestural coherence, speech production exhibits systematic modulation of gestural durations as a function of the hierarchical level of prosodic structure, e.g., the foot. Models of timing in speech ([1], [2], [17], [18]) have addressed one or another of these phenomena, but not their integration. In this paper, we report on a modeling effort that has begun to explore one approach to this integration. This effort is based on the coupled planning oscillator model of inter-unit timing in speech production that we have used previously to simulate relative timing patterns observed inter-gesturally within and between syllables ([19], [17]), and between feet within phrases ([16]).

Additionally, prosodic gestures (π-gestures) [8] have been introduced recently into the task-dynamic framework to capture the temporal lengthening of gestures in the vicinity of phrasal boundaries. Like constriction gestures, π-gestures are also active over finite time intervals. However, π-gestures act transgesturally to slow down the utterance 'clock' by locally warping the time-axis of the gestural score, and lengthening all concurrently active constriction gestures within the π-gesture's activation interval. The degree of such clock-slowing is determined by the magnitude of π-gesture. In this paper, we describe how we have generalized the clock-slowing properties of π-gestures and defined a more general class of temporal modulation gestures (μ_T-*gestures*) and applied them at the syllabic level of the prosodic hierarchy to model lexical stress. Furthermore, we describe how these μ_T-*gestures* have been integrated into the dynamics of our planning oscillator ensemble to provide a dynamics that can shape the durational differences between stressed and unstressed syllables. Thus, unlike π-gestures, which warp the time axis of gestural scores after they have been generated explicitly by rule or by hand, μ_T-*gestures* are incorporated into the dynamics of the planning oscillator ensemble and contribute implicitly to the shaping of the gestural score itself

In section 2.1 , we review our coupled oscillator model into which μ_T-*gestures* will be incorporated. In section 2.2, we describe how coupling graphs are specified for utterances and how intergestural timing is derived from the coupling network. In section 2.3, we describe how the model has been generalized to account for temporal patterns of prosodic units at different levels in hierarchy. In section 3, we detail how syllable and foot oscillators are entrained and mutually influence each other's temporal patterns (section 3.1); additionally, we detail how stress is incorporated into the ensemble of syllable and foot oscillators by employing μ_T-*gestures* (section 3.2). Finally, we describe how our model has been extended to account for more complicated temporal patterns within a foot (section 3.3).

2. Intergestural timing model

2.1. Coupled oscillator model

Our coupled oscillator model of intergestural timing is based on Saltzman & Byrd's [19] model, which implements a targeted coupling relation between a pair of planning oscillators. In the model, each oscillator is a limit-cycle system as shown in the formula below:

$$\ddot{x}_i = -\alpha_i \dot{x}_i - \beta_i x_i^2 \dot{x}_i - \gamma_i \dot{x}_i^3 - \omega_{0i}^2 x_i \quad (1)$$

where α_i, β_i, and γ_i are linear, nonlinear (van der Pol), and nonlinear (Rayleigh) damping coefficients, respectively, for the ith oscillator, and ω_{0i} is the oscillator's linear natural frequency. Each planning oscillator is associated with a constriction gesture and represents a node in the coupling graph. Pairwise inter-oscillator couplings are represented by internode edges in the graph, and are defined according to the relative phase variables between oscillators:

$$\psi_{ij} = \phi_j - \phi_i \quad (2)$$

where, ϕ_i and ϕ_j are the i^{th} and j^{th} oscillator phases, respectively; and ψ_{ji} is the relative phase between the two oscillators. When two oscillators are coupled to each other, a task-specific potential energy function (V) [11][19], centered at a target relative phase (ψ_0) is used to define a coupling force function. The potential function is defined as $V(\psi) = -a \cos(\psi - \psi_0)$, where a is a global magnitude of coupling strength, which determines how fast the ensemble of oscillators stabilizes. The derived coupling forces in the system evolve over time as functions of the interoscillator relative phases, and are added to the accelerations of the individual oscillators until steady-state relative phasing is attained. Interoscillator coupling forces are bidirectionally defined, but can be specified asymmetrically using the local coupling strength parameters, λ_{ij} and λ_{ji}, between oscillators i and j.

2.2. Coupling graph

There are two types of components in each coupling graph [23]: nodes, which represent planning oscillator phases, and edges, which represent relative-phase-dependent coupling relations, respectively. Lexical items are distinguished by corresponding differences in their associated coupling graphs that both index the items' phonological structures and can account for their patterns of intergestural relative phasing (mean values and variabilities) during speech production. For each edge, either in-phase (0º) or anti-phase (180º) coupling targets are specified for each link in the graph ([7], [17], [10], [15], [16]). These coupling modes are known to be voluntarily available during skilled movements, with the in-phase mode being the more stable [11]. The in-phase mode defines the syllable *onset relation* (CV) for which the constriction gestures in onset are all coupled in-phase to the vowel. The anti-phase mode defines the syllable *coda relation* (VC), for which anti-phase (sequential) coupling is defined between vowel and the first coda consonant. Within all consonant clusters, sequential inter-consonant coupling is defined to ensure their perceptual recoverability. The gestural coupling graph for "spat" is illustrated in Figure 3 (top), where dots and lines connecting the dots are oscillator nodes and edges, respectively; solid links denote synchronous coupling for the CV relation, and dotted links denote sequential couplings for either CC or VC relations.

Note that the graph involves multiple, potentially competitive, coupling specifications. For example, the TT (alveolar crit) gesture (for /s/) and the LIPS (clo) gesture (for /p/) are both synchronized with the vowel, but are sequentially coordinated with one another. Use of such multiple, competitive structures have been shown to capture some of the gross temporal regularities of speech and at the same time some of its more fine-grained variation; e.g., multiply-linked gestures exhibit less temporal variability than singly-linked ones ([7], [17], [15]).

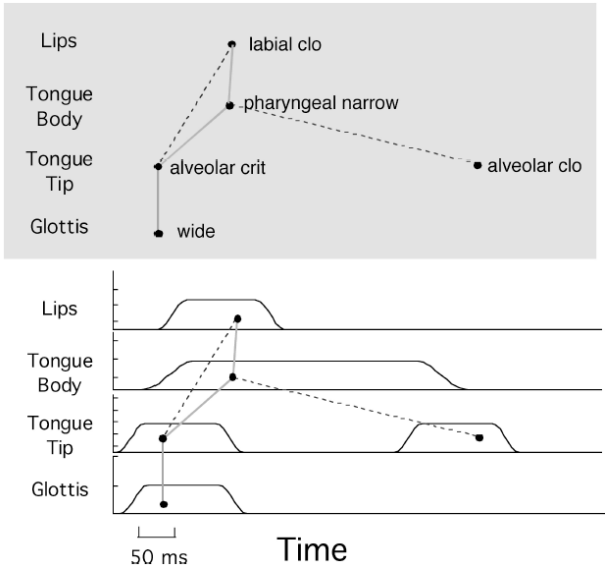

Figure 3: *Top panel is a gestural coupling graph for "spat". Solid lines are synchronous and dotted lines are sequential. Bottom panel is a gestural score generated from the coupling graph, where its underlying coupling graph is superimposed.*

Once a coupling graph for a given utterance is specified, the graph is used to parameterize the motion equations for the planning oscillators. A set of equations of motion for the coupled oscillator system is implemented using the task-dynamical coupled oscillator model of Saltzman & Byrd ([19]) for controlling pairwise relative phasing between oscillators. The system equations are numerically integrated until the ensemble of oscillators reaches a steady-state relative phasing pattern. Gestural activation onset and offset times are then specified as a function of this steady-state pattern to define the utterance's gestural score.

2.3. Entrainment of n:1 oscillators

We have generalized the use of planning oscillators to shape the temporal patterning of units at levels higher than constriction gestures, e.g., units at the levels of syllables, feet, and phrases. In these simulations, a given level's oscillators are harmonically entrained, not only with each other, but also with those at both the immediately "higher" (slower, lower frequency) and "lower" (faster, higher frequency) levels. Consequently, multi-level oscillatory ensembles are modeled with n:1 frequency ratio entrainment existing between adjacent levels, providing a temporally nested hierarchical structure for the overall ensemble. To do this, one must generalize the definition of relative phase used in equation (2), and define a *generalized relative phase* for each internode link:

$$\psi_{ij} = n\phi_j - \phi_i \tag{3}$$

where the frequency ratio of i^{th} and j^{th} oscillators are n:1 and n multiplies the slow oscillator phase (ϕ_j). Overall patterns of inter-unit timing (both within and between levels) are specified by the steady-state relative phase patterns of the ensemble's oscillatory units. Using generalized relative phase in multi-frequency oscillator ensembles leads to the entrainment of oscillators with n:1 frequency locking. In section 3, such frequency-locking will be used to model differing numbers of syllables nested within a foot.

3. Temporal patterns of unstressed & stressed syllables

A dynamical model of syllable and foot timing should account for the well known phenomenon of *polysyllabic shortening* (also called *stress-timed shortening* by [1]). As the number of syllables within a foot increases, the duration of each syllable is shortened even though the duration of the foot is lengthened (e.g. shortening of /i/: 'speed' > 'speedy' > 'speedily'). Further, recent work by Kim & Cole ([13], [14]) has shown that the compression of syllable duration as a function of the number of syllables within a foot occurs only for the stressed syllable, and not the unstressed syllables.

3.1. Polysyllabic shortening effect

O'Dell & Nieminen [18] have provided a way of modeling polysyllabic shortening in an oscillator framework, using the generalized relative phase between two oscillators with different frequencies for the syllable and foot. Languages show different degrees of syllable duration shortening as the number of syllables per foot increases. For example, stress-timed languages such as English exhibit greater shortening of syllables than syllable-timed ones such as Spanish. Key to O'Dell & Nieminen's model is that the language-specific differences in syllable duration compression with increasing number of syllables per foot can be simulated by applying *asymmetrical* coupling forces between syllable and foot oscillators. We reproduced their results using our planning oscillator model. In all our simulations, the natural frequencies of the foot and syllable oscillators were fixed at 1 and 2 (radian/sec), respectively. We chose this 1:2 frequency ratio between foot and syllable based on the cross-linguistic phonological observation that disyllabic structure appears to be the default/basic rhythmic unit ([12]).

The number of syllables per foot is specified by varying n in the generalized relative phase equation (3): $\psi = 2\phi_F - \phi_\sigma$ for two syllables per foot, and $\psi = 3\phi_F - \phi_\sigma$ for three syllables per foot. As in O'Dell & Nieminen's work [18], we created polysyllabic compression using foot-dominant asymmetric coupling strengths such that the strength from foot to syllable ($\lambda_{F\sigma}$=5) was greater than that from syllable to foot ($\lambda_{\sigma F}$ = 1). The target relative phase was specified as 0° in all simulations. Results from our simulations of two and three syllables per foot are shown in Fig 4. Syllable durations in the 2-syllable and 3-syllable feet are 3.1s and 2.5s, respectively. The entire foot duration with three syllables is greater than that with two syllables, but syllable duration decreases as the number of syllables increases.

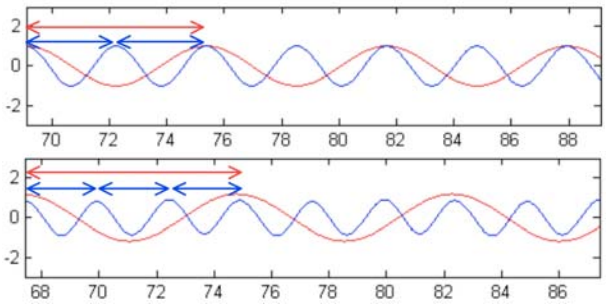

Figure 4: *Steady-state waves of foot (slow) and syllable (fast) oscillators (top panel: two syllables per foot; bottom panel: three syllables per foot). Coupling strengths to syllable and foot are 5 and 1, respectively. Vertical axis is oscillator position and horizontal axis is time.*

3.2. Asymmetrical duration between stressed and unstressed syllables

In the simulations in Figure 4, the higher-level (slower foot oscillator) period is partitioned symmetrically into N ($N = 2$ or 3 in our syllable-foot example) subintervals of equal duration by the oscillator(s) at the immediately lower (faster syllable oscillator) level. This durational symmetry poses a problem for the modeling of stress, however, where a given foot is partitioned asymmetrically into a longer duration stressed syllable and a series of shorter unstressed syllables.

We have generalized our previous use of clock-slowing π-gestures in the vicinity of prosodic boundaries, and have integrated these generalized *temporal modulation gestures* (μ_T*_gesture*) into our planning oscillator model in a manner that has allowed us to simulate the within-foot durational asymmetry between stressed and unstressed syllables. More specifically, we incorporated a μ_T*_gesture* into the foot-syllable oscillator ensemble, and defined the μ_T*_gesture's* activation interval to be a function of the syllable oscillator's ongoing phase. Since the stressed syllable corresponds to the first cycle of the syllable oscillator within the foot cycle, the μ_T*_gesture* was defined over the following phase interval of the foot oscillator: $0 \le \phi_F < \kappa$, (κ is $2\pi/n$, n is the number of syllable per foot). The corresponding activation function ($a_{\mu T}$) of the μ_T*_gesture* is illustrated in Figure 5, and is defined in detail in equation (4):

$$a_{\mu_T} = \begin{cases} -.5\cos(\phi_F (1/\rho\kappa)\pi), & \text{if } 0 \le \phi_F < \rho\kappa \\ 1, & \text{if } \rho\kappa \le \phi_F < (1-\rho)\kappa \\ .5\cos([\phi_F - (1-\rho)\kappa](1/\rho\kappa)\pi), & \text{if } (1-\rho)\kappa \le \phi_F < \kappa \\ 0, & \text{if } \kappa \le \phi_F < 2\pi \end{cases} \quad (4)$$

In this equation, the rise and fall intervals of the activation function are half-cosine ramp functions that ensure smooth on/off transitions, and occupy a phase interval of $\rho\kappa$ in the foot cycle. As shown in Figure 5, during the interval for stressed syllable (when $0 \le \phi_F < \kappa$), the μ_T*_gesture's* activation gradually increases to 1 from 0 and decreases to 0 from 1 by the onset (when $\phi_F = \kappa$) of the first unstressed syllable.

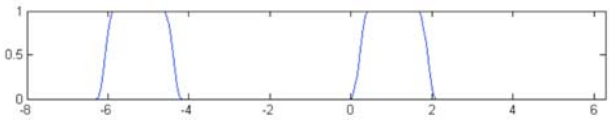

Figure 5: μ_T*_gesture coordinated to syllable-foot ensemble (three syllable per foot case).* $a_{\mu T} = 1$ *during the interval for stressed syllable and* $a_{\mu T} = 0$ *during the interval for unstressed syllables.* $\rho = .2$ *for half-cosine ramping function. Vertical axis is activation magnitude and horizontal axis is foot oscillator phase (radian).*

When active, the μ_T*_gesture* has the effect of transiently lowering the values of the frequency parameters of both oscillators in the foot-syllable ensemble, thereby generating the required durationally asymmetric pattern of stressed and unstressed syllables within feet. This frequency modulation is described by equation (5):

$$\omega^*_{o,i} = (1 - \delta\, a_{\mu_T})\omega_{o,i} \quad (5)$$

where δ denotes modulation strength and i is oscillator identity (i = syllable or foot oscillator). As μ_T*_gesture* activation increases, ω^* will decrease and consequently slow the rate of phase flow of the oscillator ensemble during the interval for the stressed syllable.

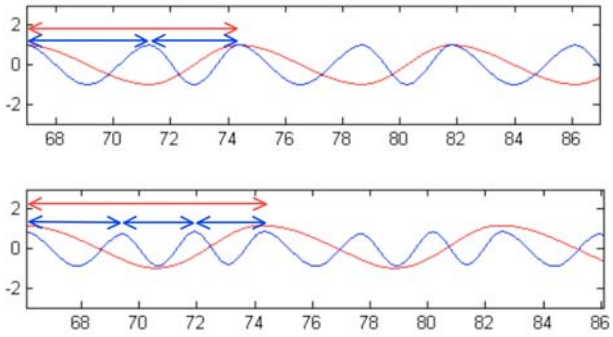

Figure 6: *Steady-state waves of foot (slow) and syllable (fast) oscillators (top panel: two syllables per foot; bottom panel: three syllables per foot). Coupling strengths to syllable and foot are 5 and 1, respectively.* μ_T*_gesture is active during the interval of stressed syllables. Vertical axis is oscillator position and horizontal axis is time.*

We conducted a simulation incorporating a 'clock'-slowing μ_T*_gesture* to model the within-foot durational asymmetry between stressed and unstressed syllables using the same parameter settings as used in section 3.1, and with $\delta = .5$. As illustrated in Figure 6, the interval of the stressed syllable is longer than that of the unstressed syllable due to the effect of the μ_T*_gesture* in both the two and three syllable cases. The durations of syllables in the 2-syllable case are 4.3s (stressed) and 3.1s (unstressed); in the 3-syllable case, 3.4s (stressed) and 2.5s (unstressed). Importantly, the polysyllabic compression effect is still maintained: the foot duration in the three syllable case is lengthened than that in the two syllable case while the duration of an individual syllable is shortened with the syllable number increased. However, the shortening effect in the unstressed syllables is contrary to the temporal pattern in Kim & Cole ([14], [15]).

3.3. Weakening of polysyllabic shortening in unstressed syllables

Kim & Cole ([13], [14]) showed that polysyllabic shortening is observed only in stressed syllables, and not in unstressed syllables. We attribute this phenomenon to a weakening of the foot oscillator's temporal compression on the unstressed syllables relative to its strength on the stressed syllable. Recall that a coupling strength asymmetry between syllable and foot oscillator determines the degree of the polysyllabic shortening. The weakening of this compression can be implemented by varying the degree of asymmetry in foot-syllable coupling strength across stressed and unstressed syllables. We used the same activation function ($a_{\mu T}$) for the μ_T_*gesture* as was used in the previous section (equation 4). Unlike the simulations described in sections 3.1 and 3.2, in which the coupling strength ratio, ($\lambda_{F\sigma}/\lambda_{\sigma F}$) remained constant throughout the simulations, we now varied the value of $\lambda_{F\sigma}/\lambda_{\sigma F}$ across the stressed and unstressed syllables according to:

$$\varepsilon(\phi_F) = \left(\varepsilon_{stress} - \left(1/\varepsilon_{stress}\right)\right) a_{\mu_T}(\phi_F) + \left(1/\varepsilon_{stress}\right) \quad (6)$$

In this equation, ε (ϕ_F) denotes the foot-syllable oscillator coupling strength ratio which varies as a function of foot oscillator phase. During the stressed syllable, its plateau value is $\varepsilon_{stress} = 5$, which is the same value as the constant value used in previous simulations. The plateau value of ε (ϕ_F) during the unstressed syllables is constrained to equal $1/\varepsilon_{stress} = 1/5$. Thus, as shown in Figure 7, $\lambda_{F\sigma} = 5$ and $\lambda_{\sigma F} = 1$ for the stressed syllable; $\lambda_{F\sigma} = 1$ and $\lambda_{\sigma F} = 5$ for the unstressed syllables,.

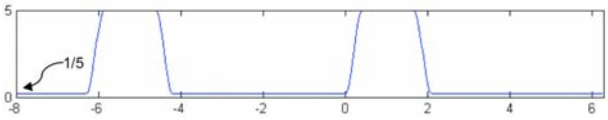

Figure 7: *Coupling strength ratio as a function of foot oscillator phase (three syllable per foot case). ε = 5 during the interval for stressed syllable and ε = 1/5 during the interval for unstressed syllables. Vertical axis is magnitude of the ratio and horizontal axis is foot oscillator phase (radian).*

Results of these simulations are shown in Figure 8. Syllable durations in the 2-syllable case are 4.3s (stressed) and 3.1s (unstressed); in the 3-syllable case, 3.4s (stressed) and 3.1s (unstressed). The two effects (polysyllabic shortening and durational difference between unstressed and stressed syllables) generated in the previous simulations (section 3.1 and 3.2) are maintained, but now, consistent with the data of Kim & Cole ([14], [15]), the durations of unstressed syllables remains constant regardless of the number of syllable per foot.

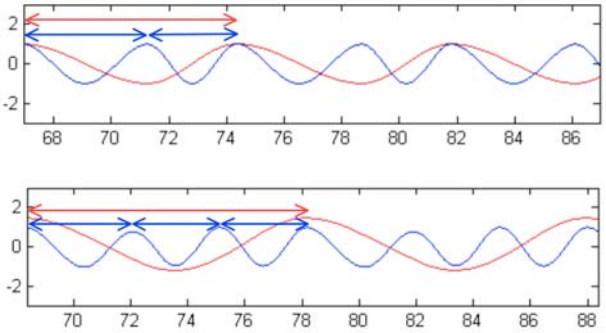

Figure 8: *Steady-state waves of foot (slow) and syllable (fast) oscillators (top panel: two syllables per foot; bottom panel: three syllables per foot). Coupling strengths to syllable and foot are 5 and 1, respectively. μ_T_gesture is active during the interval of stressed syllables. $\lambda_{F\sigma} = 1$ and $\lambda_{\sigma F} = 5$ for unstressed syllable, and $\lambda_{F\sigma} = 5$ and $\lambda_{\sigma F} = 1$ for stressed syllables. Vertical axis is oscillator position and horizontal axis is time.*

4. Summary and Conclusion

In section 3.1, we modeled the polysyllabic shortening effect with our planning oscillator model using O'Dell and Nieminen's ([18]) method. However, syllable durations within a foot were constant, and syllable period simply represented an average duration of within-foot syllable periods. These simulations were not able, therefore, to model the durational asymmetry between stressed and unstressed syllables within feet. In section 3.2, we incorporated a 'clock'-slowing μ_T_*gesture* to implement this durational asymmetry between unstressed and stressed syllables. In section 3.3, we showed how the additional modulation of interoscillator coupling strength ratio was able to generate the data pattern reported by Kim & Cole ([14], [15]), namely, polysyllabic shortening of the stressed syllable and durationally invariant unstressed syllables when syllables are added to feet. Our planning oscillator model, once combined with the operation of μ_T_*gestures*, captures all these temporal patterns in one integrated package and is indexed by a set of 3 parameters: interoscillator coupling strength ratio (section 3.1), μ_T_*gesture* strength (section 3.2), and the relative degree of foot dominance over stressed vs. unstressed syllables (section 3.3). In future work, we will investigate how language-specific settings of these parameters can account for corresponding cross-language differences in temporal patterning phenomena within and across levels of the prosodic hierarchy.

5. Acknowledgments

This was supported by NIH grant DC-02717.

6. References

[1] Barbosa, P. A. (2002). Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production. In B. Bel, & I. Marlien, (Eds.). *Proceedings of the 1st International Conference on Speech Prosody*, Aix-en-Provence, France. Pp. 163-166.

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

[3] 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.

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

[5] 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.

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

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

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

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

[10] 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.

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

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

[13] 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.

[14] 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

[15] 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.*

[16] Nam, H., Goldstein, L., & Saltzman, E. (2006). Self-organization of syllable structure: A coupled oscillator model. In I. Chitoran, C. Coupé, E. Marsico, & F. Pellegrino, (Eds.). *Approaches to Phonological Complexity*. New York: Mouton de Gruyter.

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

[18] 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).

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

[20] 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.

[21] 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.

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

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

[24] 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

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