

Inter-unit Timing in Speech Production: Pattern, Stability, and Hierarchy

Elliot Saltzman
Boston University and Haskins Laboratories

The original task-dynamic model of speech production at Haskins Laboratories incorporated the theoretical tenets of Browman & Goldstein's 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 hierarchical 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 trigger the activation function of that oscillator's associated constriction gestures. 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 (mT-gestures) that alter the rate of utterance timeflow by smoothly changing all frequency parameters of the planning oscillator ensemble; and spatial modulation gestures (mS-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. 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 talk, I will describe recent developments to the task-dynamic toolkit (original task-dynamic model, planning oscillator ensemble, and modulation gestures) and how they have been used to interpret and simulate experimental data on the interactions of stress and prominence in shaping the kinematic details of speech production.