6. THE DISTINCTIONS BETWEEN STATE, PARAMETER AND GRAPH DYNAMICS IN SENSORIMOTOR CONTROL AND COORDINATION

Elliot Saltzman
Department of Rehabilitation Science, Boston University, Boston MA, USA; Haskins Laboratories, 300 George Street, New Haven, CT USA

Hosung Nam
Department of Linguistics, Yale University, New Haven, CT USA

Louis Goldstein
Haskins Laboratories, 300 George Street, New Haven, CT USA; Department of Linguistics, Yale University, New Haven, CT USA

Dani Byrd
Department of Linguistics, University of Southern California, Los Angeles, CA USA

Abstract
The dynamical systems underlying the performance and learning of skilled behaviors can be analyzed in terms of state-, parameter-, and graph-dynamics. We review these concepts and then focus on the manner in which variation in dynamical graph structure can be used to explicate the temporal patterning of speech. Simulations are presented of speech gestural sequences using the task-dynamic model of speech production, and the importance of system graphs in shaping intergestural relative phasing patterns (both their mean values and their variability) within and between syllables is highlighted.

I. Introduction
What is being learned when we learn a skilled behavior? In our opinion, and in those of many other proponents of the dynamical systems approach to sensorimotor control and coordination, what is learned is the underlying dynamical system or coordinative structure that shapes functional, task-specific coordinated activity across actor and environment. But this begs the question of just what sort of a beast a dynamical system is. Fortunately, it is not too difficult to define one. Roughly, a dynamical system is a system of interacting variables or components whose individual behaviors and whose modes of interaction are shaped by laws or rules of motion. The focus of this chapter is on the types of variables that comprise a dynamical system, the types of laws or rules that govern changes of these variables over time, and the manner in which such changes can be related to processes of coordination and control in skilled behavior, with particular emphasis on temporal patterning in the production of speech.

State-, Parameter- and Graph Dynamics. Any dynamical system can be completely characterized...
II. CONTROL OF RHYTHMIC ACTION

According to three sets of variables—state, parameter, and graph-variables (Farmer, 1990; Saltzman & Munhall, 1992)—and the laws or rules that govern their respective dynamical changes over time. State-variables can be viewed as the system's active degrees of freedom, and are represented as the dependent or output variables of the set of autonomous differential or difference equations of motion that are used to describe the system. More specifically, a given nth-order dynamical system has n state variables and can be described, equivalently, by a single nth-order equation of motion or by a set of nth-order equations of motion, with one nth-order equation of motion for each state variable. For example, a 2nd-order mechanical systems such as damped mass-spring systems or limit cycle pendulum clocks have two state variables, position and velocity; typical nth-order computational (connectionist) neural networks have n state variables that are defined by the activation levels of each of the network's n processing nodes. State dynamics refers to the manner in which changes over time of the state variables are shaped by the "forces" (more technically, the state-velocity vector field) inherent to the system that are described by the system's equation(s) of motion. A system's parameters are typically defined by the coefficients of constant terms in the equation of motion. For example, these could be the mass, damping, and stiffness coefficients or the constant target parameter in a damped, mass-spring equation, the length and mass of a pendulum, or the inter-node synaptic coupling strengths in a computational neural network. Parameter dynamics refers to the manner in which changes in parameter values are governed over time. In general, a system's parameters change more slowly than its state-variables, although this is not always the case. For example, a child's limb lengths and masses change at an ontogenetic timescale while their skilled limb movements unfold in real-time. Similarly, a network's synapses weights change over the timescale defined by the learning algorithm used to train the network to solve a given computational task; this learning timescale is typically much slower than the state-dynamic performance timescale of the activation state variables. It is possible, however, for system parameters to change on a timescale comparable to, or even faster than, the corresponding state-variables' timescale. For example, we can intentionally change the rate at which we reach toward a target, or even switch from one target to another, during the reaching motion itself.

The notion of a system's graph is a less familiar one, at least in the domain of sensorimotor control and coordination, than that of the system's set of state-variables or parameters. The graph of a system represents the "architecture" of the system's equation of motion, and denotes the parameterized set of relationships defined by the equation among the state-variables. For connectionist systems, this graph is simply the standard node-linkage diagram used to represent such systems (see Figure 1). For non-connectionist systems, the conceptual connection between a system's graph and its equation of motion is less straightforward, but becomes clearer when one realizes that a symbolically written differential or difference equation can be represented equivalently in pictorial form as a circuit diagram. The latter type of representation can be used to construct an electronic circuit to simulate the system on an analog computer, or to graphically define the equation in an application such as MathLab's Simulink for simulating the system on a digital machine. Figure 2 shows a circuit diagram used to simulate a 2nd-order damped mass-spring system using Simulink.

Graph dynamics refers to the manner in which the system graph changes over time. This can include changes in the system's dimensions, i.e., the number of active state variables; or changes in the structure of the state variable functions included in the system's equation of motion. In a behavioral context, the number of active state variables might change due to a decision or instruction to switch from unimanual to bimanual lifting of a given object, or to the recruitment of the trunk in addition to the arm when reaching for a distant target. Relatively, in connectionist systems trained by "constructive" learning algorithms, nodes and linkages can be added or deleted to create a network whose structural complexity is adequate for instantiating a "grammar" that is sufficient for learning particular classes of functions (e.g., Hsuung, Saratchandran, & Sundararajan, 2002; Quartz & Sejnowski, 1997). Additionally, the damping functions implemented during discrete point attractor tasks such as reaching or pointing will be qualitatively different from those required to perform sustained rhythmic, limit cycle polishing or stirring tasks; similarly, interlimb coupling functions will be different for skilled performances of bimanual polyrhythms with correspondingly different mn frequency ratios.

In the following sections, we will review some recent work of ours that highlights the role of system graphs in shaping the temporal patterning of speech. Our work is presented within the framework of the Task-Dynamic model of speech production (e.g., Saltzman 1986; Saltzman & Munhall 1989; Saltzman & Byrd, 2001), and we focus on the manner in which the relative timing of speech gestures, i.e., intergestural timing, within and between syllables (with respect to both mean intergestural time intervals and their variability) can be understood with reference to the system's underlying intergestural coupling structures.

II. The Task-Dynamic Model of Speech Production: An Overview

The temporal patterns of speech production can be described according to four types of timing properties: intragestural, intergestural, inter- and global. Intragestural timing refers to the temporal properties of a given gesture, e.g., the time from gestural onset to peak velocity or to target attainment; intragestural timing refers to modulations of the timing properties of all gestures active during a relatively localized portion of an utterance; intergestural timing refers to the relative phasing among gestures (e.g., between bilabial closing and laryngeal devoicing gestures for /p/); between coronal/bilabial closing and vocalic tongue dorsum shaping gestures for /p/); and global timing refers to the temporal properties of an entire utterance, e.g., overall speaking rate or style.
In the task-dynamic model of speech production, the spatiotemporal patterns of articulatory motion emerge as behaviors implicit in the dynamical system with two functionally distinct but interacting levels. As shown in Figure 3, the inarticulator level is defined according to both model articulator (e.g., lips & jaw) and motor-kinematic (e.g., tongue) coordinates. The interarticulator level is defined according to a set of activation coordinates. Invariant gestural units are posed in the form of context-independent sets of dynamical parameters (e.g., target, stiffness, and damping coefficients) and are associated with corresponding subnets of model articulator, tract-variable, and activation coordinates. Each unit's activation coordinate reflects the strength of the associated gestural units (e.g., bilateral closing) "attempts" to shape vocal tract movements at any given point in time. The tract-variable and model articulator coordinates of each unit specify, respectively, the particular vocal-tract construction (e.g., bilateral) and articulatory synergy (e.g., lips and jaw) whose behaviors are affected directly by the associated unit's activation. The inarticulator level accounts for the coordination among articulators at a given point in time due to the currently active gestural set. The interarticulator level governs the patterns of relative timing among the gestural units participating in an utterance and the temporal evolution of the activation trajectories of individual gestures in the utterance. The trajectory of each unit's activation coordinate defines a forcing function specific to the gesture and acts to insert the gesture's parameter set into the interarticulator dynamical system defined by the set of tract-variable and model articulator coordinates. Additionally, the activation function gates the components of the forward kinematic model (from model articulators to tract variables) associated with the gesture into the overall forward and inverse kinematic computations (see Selman & Munhall, 1989, for further details).

In the original version of the model, the interarticulator level used general scores (e.g., Brown & Goldstein, 1990) that explicitly specified the activation of gestural units over time and that unidirectionally drove articulatory motion at the interarticulator level. In these gestural scores, the shapes of activation waves were restricted to step functions for simplicity's sake, and the relative timing and durations of gestural activations were determined, until relatively recently, either with reference to the explicit rules of Brown and Goldstein's Articulatory Phonology (e.g., Brown & Goldstein, 1990) or by "hand." Thus, activation trajectories were modeled as switching discretely between values of zero (a gesture has no influence on tract shape) and one (the gesture has maximal influence on tract shape). However, it had not been noted for some time that a simple step-function activation waveguide is an over-simplification (e.g., Bullock & Grossberg, 1988; Covington, 1976; Kroger, Schröder, & Oppe-Rhein 1995; Ostry, Gribble, & Guoico, 1996), and we have now explored some of the consequences of non-step-function activation waveguides on articulator kinematics. In particular, we explicitly specify activation functions by hard- or half-coin-shaped rises and falls of varying durations, which we have been able to create articulatory trajectories whose kinematics capture individual differences in gestural velocity profiles found in experimental data (Floyd & Saltzman, 1998).

We have also explored two types of dynamical system for shaping activation trajectories in a relatively self-organizing manner. Both types can be related to a class of rather generic recurrent connectionist network architectures (e.g., Jordan, 1986, 1990; see also Bally, Laboissière, & Schwartz, 1991; Latour, 1989). In the former networks (see Figure 4), outputs correspond to gestural activations with one output node per gesture. The temporal patterning of gestural activation trajectories can, to a large extent, be viewed as the result of the network's state unit activity. This activity can be conceived as defining a dynamical flow with a time scale that is intrinsic to the intended speech sequence and that creates a temporal context within which gestural events can be located. The patterning of activation trajectories is a consequence of the trained nonlinear mapping from the state unit flow to the output units' gestural activations. (For purposes of the present discussion we will ignore the plan units, which provide a unique identifying label for each sequence that the network is trained to perform, and that remain constant during the learning and performance of their associated sequences.)

There are two types of state unit structures that have been used in such networks (e.g., Jordan, 1986, 1992). In the first, each state unit is a self-recurrent, first-order filter that provides a decaying exponential representation of time, and that receives recurrent input from a given output unit. Thus, state units and output units are defined in a one-to-one manner, and the state units effectively provide a sequence-specific, exponentially weighted average of the activities of their associated output units. This is the type of state unit structure used in our previous work on the dynamics that give rise to a gesture's anticipatory interval of articulation, defined operationally as the time from gestural onset to the time of required target attainment (e.g., Saltzman & Mitra, 1998; Saltzman, Lefevre, & Mitra, 2000). Using such a model we were able to capture individual differences demonstrated experimentally among speakers in the temporal elasticity of anticipation intervals (e.g., Aby & Laucocke, 1995), according to which an interval lengthens (i.e., begins earlier), but only fractionally, with increasing numbers of preceding non-conflicting gestures.

The second type of state unit structure is a linear second-order filter composed of an antisymmetrically coupled pair of first-order units. Each such structure provides an oscillatory representation of time, generating a circular trajectory in the cartesian activity space of its two component units. The rate of rotation of this circle is a function of the weights of the units' self-recurrent and cross-coupling connections. In some cases, only one such oscillator is defined for a network with multiple output units (e.g., Bally, Laboissière, & Schwartz, 1991; Laboissière, Schwartz, & Bally, 1991; Jordan, 1990), and the state unit oscillator defines a relatively simple "clock," whereby the output units are activated whenever the clock passes through a corresponding set of time or phase values that are acquired during network training. For example, if two outputs correspond, respectively, to the activation of bilabial and laryngeal gestures for /p/, the relative phasing of these gestures would be determined by the relative values of their associated phase values in the state clock. Additionally, if the clock were gradually sped up in order to increase the flow rate of the state clock, but intergestural relative phasing would not change.

One problem with this simple state clock model is that stability is lost in systematic ways when speaking rate is increased, such that the state unit and intergestural relative phasing are observed at critical rates. Specifically, intergestural phase transitions are observed during rate-scaling experiments, showing discontinuous transitions of intergestural phasing with continuous increases in speaking rate (Kelsay, Saltzman, & Tuller, 1986a, 1986b; Tuller & Kebo, 1991). In these studies, when subjects speak the syllable /sip/ repetitively at increasing rates, the relative phasing of the bilabial and laryngeal gesture associated with the /p/ does not change from the pattern observed at a self-selected, comfortable rate. However, when the repeated syllable /sip/ is similarly increased in rate, its relative phasing pattern switches relatively abruptly at a critical speed—provided that the correct rate to the pattern observed for the /p/ sequence. This phase transition in the intergestural timing of bilabial and laryngeal gestures implies that at least two separate state-unit oscillators are required, possibly one oscillator for each gestural unit, and that during the performance of a given sequence these state-oscillators behave as functionally coupled, nonlinear, limit-cycle oscillators. In unperturbed cases, the observed pattern of gestural activations would correspond to an associated pattern of synchronization (entrainment) and relative phasing among the state-oscillators that was acquired during training. Similarly, the intergestural phase transitions may be viewed as behaviors of a system of nonlinearly coupled, limit-cycle oscillators that bifurcate from a modal pattern.
that becomes unstable with increasing rate to another
modal pattern that retains its stability (e.g., Halen,
Kelso, & Bunz, 1985). The implications for models of
activation dynamics are that the state unit clock
should contain at least one oscillator per get-
ture, where the oscillators are governed by (nondissipative)
limit cycle dynamics, and that these oscillators should be
mutually coupled with one another.

The hypothesis of an ensemble of oscillators, one
oscillator per gesture, that comprises a "clock" gov-
erning the timing of each gesture in a speech se-
quence echoes an earlier hypothesis by Brownman
and Goldstein (1990). According to their hypothesis,
there is an abstract (linear) "timing oscillator" associated
with each gestural unit, and that these oscillators are
coupled in a manner that is responsible for the relative
timing of gestural onsets and offsets in the sequence.
Further, they hypothesized that different types of
intergestural coupling structures existed within different
parts of syllables and between syllables, and that tim-
ing patterns observed experimentally, both intra- and
inter-syntactically, could be understood with reference
to these coupling patterns. In the following section,
we describe our recent work in which state unit limit
cycle oscillators ("plotting oscillators") are defined in
a 1:1 manner for each gestural activation node. In this
work, a system graph is used to specify the coupling
structure among the oscillators (i.e., the presence or
absence of inter-oscillator linkages and, if present,
the strength and target relative phase associated with each
linkage) for a given gestural sequence, and the steady-
state output of this oscillatory ensemble is used to
specify the onset and offsets of gestural activations for
use in a gestural score for the sequence. Remarkably,
the resultant relative timing patterns reflect both
the mean values and variability observed experimentally
for intra- and inter-syntactic gestural timing patterns.

III. Coupling Graphs and Intergestural
Cohesion: Intra- and Inter-Syllabic Effects

In our present work, we have extended Saltzman &
Byrd's (2000) task dynamic model of intergestural
phasing in a coupled pair of oscillators to the case
in which multiple (more than two) oscillators are
allowed to interact in shaping the steady-state pattern
of intergestural phase differences (Nam & Saltzman,
2003). For a single pair of oscillators in the absence
of added perturbations or noise, the system always settles
or "relaxes" from its initial condition to a steady-state
amplitude and phase for each oscillator) to a steady-state
attractor characterized by a relative phasing pattern
that is identical to the target relative phase value that
is used to parameterize the bidirectional coupling function
between the oscillators. When the system contains
more than two oscillators, this is no longer necessarily
the case, since the pairwise target relative
phasings may be incompatible and in competition
with one another. For example, for a system of three
oscillators (A, B, C) with competing or incompatible
target parameters for relative phase, e.g., 20° for
the AB pair, 40° for the BC pair, and 30° for the
AC pair, and with relatively equal strengths for each
of the coupling functions (coupling strength is a second
parameter of the coupling functions), none of the
oscillator pairs will attain a steady-state relative phasing
that matches the corresponding targets. For systems
with unequal coupling strengths across the oscillator
pairs, however, those pairs with relatively larger
coupling strengths will attain steady-state relative phases
closer to their targets than pairs with lesser coupling
strengths. On the other hand, in a similar system with
equal coupling strengths but with no such phasing tar-
et competition, e.g., 20° for the AB pair, 40° for
the BC pair, and 60° for the AC pair, all oscillator
pairs will achieve their targets in the steady-state.

Thus, the choice of which gestures to couple to
each other (identified by nonzero-strength inter-
node links in the oscillatory ensemble's system graph,
including the relative strengths of the intergestural
coupling functions, strongly influences the re-
sultant steady-state patterns of intergestural timing.
When we implemented the system graphs proposed by
Brownman & Goldstein (2000), we found that the
model automatically displayed symmetries of intra-
syllabic gestural behavior that have been observed
empirically, namely that syllable-initial consonant se-
quences (onssets) behave differently from syllable-final
sequences (codas) in two ways 1. Onsets have been
displayed to play a characteristic pattern of mean in-
tergestural relative phasing values, labeled the C-center
effect by Brownman and Goldstein (2000), that codas
do not display. Additionally, onsets also exhibit less
variability (i.e., greater stability) of intergestural rela-
tive phasing compared to the variability found in codas
(Byn, 1996).

The C-center effect describes the fact that, as con-
sonants are added to onsets, the resultant timing of all consonant gestures changes with respect to the

1 The onset of a syllable denotes the consonants in the syllable that precede the vowel, syllable codas denote the consonants in a syllable following a vowel, syllable ripe (or offset) denotes the vowel or nucleus of a syllable together with the following consonants in that syllable. So, for example, in the word green, the onset is /e/ in the velar nucleus, /n/ is the coda, and /n/ is the ripe.

A: # C₁ ---- C₂ --- V
B: V — C₁ ---- C₂ #

FIGURE 5. Syllabic articulatory gestures derived from X-ray micro-beam data in Brownman and Goldstein (2000) for consonant-vowel sequences in 'sayed', 'spayed' and 'played'. Vertical dotted line denotes the temporal "center" of gravity (C-centers) for the onset consonant gestures. Horizontal arrows show invariant time from onset centers to the vowels (C-center effect).

A:  V — C₁ —— C₂ —— V
B: V — C₁ —— C₂ #

6. THE DISTINCTIONS BETWEEN STATE, PARAMETER AND GRAPH DYNAMICS

for onsets (Figure 6A) defines the C-C coupling as well as (identical) C-V couplings for each consonant to the vowel, and there is competitive interaction between the C-C and C-V couplings; for codas (Figure 6B), however, the graph defines a similar C-C coupling, but only the first consonant is coupled to the preceding vowel (V-C coupling) and there is no comparable competition.

We used these onset and coda coupling graphs to parameterize simulations based on our extended coupled oscillator model consisting of three pairwise-coupled oscillators (Nam & Saltzman, 2003). Implementing the graph in Figure 6A for onset cluster simulations, we set the target relative phase parameters of the coupling functions to 50°, 50° and 30°, respectively, for the C₁-V, C₂-V couplings (Figure 7, top row, left) and the C₁-C₂ couplings (Figure 7, top row, right). All coupling strength parameters were set to equal 1. When the system settled into its entrained steady-state, the resultant intergestural relative phases were 59.9° for C₁-V, 39.9° for C₂-V, and 19.9° for C₁-C₂ (Figure 7, bottom). Thus, implementing the graph in Figure 6A resulted in none of the intergestural relative phases achieving their target values due to the competitive interactions between the C-V and C-C couplings. Importantly, however, the

following vowel in a way that preserves the overall
timing of the center of the consonant sequence with
respect to the vowel (see Figure 5). In contrast, how-
ever, as consonants are added to coda sequences, the
temporal distance of the center of the cluster from the
preceding vowel simply increases with the number of
hyposthesized that these different behaviors originated in
different underlying coupling structures for the con-
sonant gestures in onsets and codas. As shown in
Figure 6, these different structures can be represented as correspondingly different system graphs. The graph

FIGURE 6. Coupling graphs proposed by Brownman & Goldstein (2000) for syllable onsets (6A) and codas (6B). # denote syllable boundaries.

The above simulations focused on the effects of system graph structure on patterns of intergestural relative phasing displayed within syllables. These intrasyllabic simulations were purely deterministic and contained no contributions of stochasticity or noise. Consequently, the steady-state patterns observed for a given parameterization and set of initial conditions behaved identically from one "trial" to the next, and may best be considered to reflect the mean values of intergestural phasing observed experimentally. There is, however, a behavioral asymmetry in the relative stability or variability found in onsets and codas—intergestural phasing is less variable (more stable) in onsets than in codas (Byrd, 1996). Bowman & Goldstein (2000) hypothesized that, similar to the asymmetries found between mean intrasyllabic phasing patterns, the asymmetries in stability could also be accounted for by the differences between coupling graphs for onsets (Figure 6A) and codas (Figure 6B). To test this hypothesis we conducted a set of simulations in which stochasticity was incorporated and the resultant variability of relative phasing was measured.

In these simulations, we used the same onset and coda graphs as above (Figure 6A and 6B), and incorporated variability by introducing trial-to-trial random variation in the detuning (i.e., the difference between the natural frequencies) of the component oscillator pairs. We ran groups of simulated trials for each utterance type, adding a random amount of detuning in each trial to each oscillator pair via the associated intersyllabic coupling function. The detuning parameter, $\theta$, in each coupling function was defined as a random variable with a mean and standard deviation equal to zero and $\sigma$, respectively. The value of $\sigma$ (the amount of detuning noise) was manipulated in a series of five noise conditions, increasing from 0.05 to 0.5 in 0.05 increments. 200 simulation trials were run for each utterance (onset, coda) x noise-level (5 levels) condition, and we measured the standard deviation of the final steady-state relative phase between $C_1$ and $C_2$ for each condition. Figure 9 displays the amount of variability shown in each condition. Not surprisingly, both onset and consonant clusters are greater resultant steady-state variability as the added noise level increases. More importantly, however, is the fact that at each noise level the variability in relative phasing is smaller for onset clusters than for coda clusters. This reflects the experimental finding that onset clusters are more stable than coda clusters in their relative timing (Byrd, 1996), and supports the hypothesis of Bowman and Goldstein (2000) that this asymmetry in variability/stability is the emergent consequence of differences between onsets and codas in their underlying intergestural coupling graphs.
utterance conditions. Additionally, the variability in onset clusters in smaller than in coda clusters, replicating our earlier results but now using larger gestural ensembles that include inter syllabic coupling. Finally, and crucially, the variability across syllabic boundaries was largest of all, in agreement with empirical data reported by Byrd (1996).

IV. Concluding Remarks

We have reviewed the manner in which dynamical systems for coordination and control can be analyzed in terms of their state variables, parameters, and graphs. For the most part, system graphs have been ignored in studies focusing on the spatiotemporal properties of skills actions. However, in at least the case of inter syllabic timing patterns in speech production, it appears that system graphs can be invoked not only as the source of the mean timing properties of an utterance, but of the particular structure of its variability as well. We are encouraged by the power of this approach and are both curious and eager to see how this focus can be brought to bear as well on the study of nonspeech skilled activities.

Acknowledgements

This work was supported by NIH grants DC-03663 and DC-03172.

References


