

Three mechanisms for modeling articulation: selection, coordination, and intention

Sam Tilsen

Department of Linguistics,
Cornell University, Ithaca, NY
tilsen@cornell.edu

Abstract

A conceptual framework for modeling articulatory control is presented in a tutorial fashion. The framework incorporates three mechanisms: selection, coordination, and intention. Selection is a mechanism for governing the choice and ordering of articulatory movements, and operates through an activation code. Coordination is a mechanism for governing the control of movement timing in precise manner, and operates through a phase code regulated by coupled oscillators. Intention is a mechanism for determining the target state of the vocal tract, and operates through a spatial code derived from integrating over parameter fields. All three mechanisms are inherently dynamic, and their interactions provide a basis for understanding a wide variety of phonetic and phonological patterns in development and across languages. An integrated model of the mechanisms is described and applied to a range of empirical phenomena.

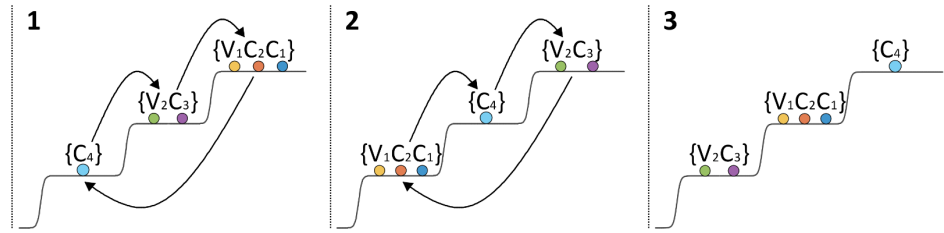
1. Introduction

This aim of this article is to describe a conceptual framework for modeling articulatory control. The scope of framework is broad in that it addresses both observations of phonetic variation and observations of developmental and phonological patterning within and across languages. The main premise is that approaches to modeling articulation must distinguish three fundamental aspects of control: [i] choice of articulatory movements (selection), [ii] control of movement timing (coordination), and [iii] control of movement targets (intention). Hence three control mechanisms are necessary: a selection mechanism for choosing sets of movements and ordering them in time, a coordination mechanism for precisely controlling the relative timing of selected movements, and an intention mechanism for determining the targets of movements. What makes the empirical study of articulatory control challenging is that these three mechanisms can interact in complicated ways: the interactions inevitably vary over the course of development, vary according to contextual and task-specific factors, and vary across languages, individuals, and utterances.

The three mechanisms are schematized in Fig. 1. Selection organizes articulatory gestures into competitively selected sets which obtain a pattern of relative activation. The pattern of relative activation, in combination with a selection threshold and feedback-driven reorganization processes, determines an order in which sets of gestures are selected for execution. Coordination involves a system of coupled oscillators which determines the relative timing of the initiation of selected gestures. Intention involves multiple parameter-planning fields which encode the targets of movements in motoric coordinates.

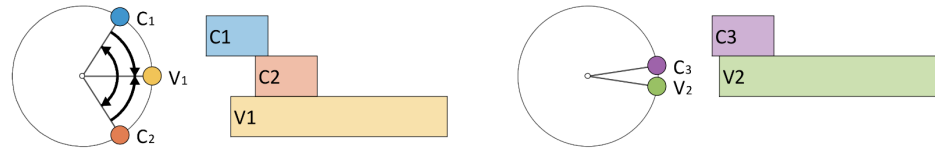
selection

organizes articulatory gestures into competitively selected sets



coordination

phase-coupled oscillators govern relative timing of gestural activation



intention

target vocal tract state determined by integrating population codes

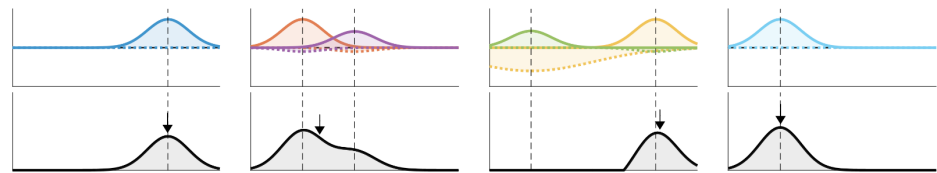


Fig. 1. Three mechanisms of articulatory control. Top: example of a series of relative activation patterns (i.e. activation potentials) for articulatory gestures (labeled as C_1 , V_1 , etc.). Gestures which occupy the same level of the potential are co-selected; gestures which occupy different levels are competitively selected. Middle: example of relative phases of coupled oscillators, shown on unit circles; relative phases determine the relative timing of initiation of selected gestures. Bottom: examples of spatial distributions of activation in parameter fields; the distributions arise from integrating gestural input (upper panels) and the centroid of field activation (arrows in lower panels) determines a current target state.

All three mechanisms can be understood to operate on time-varying codes: selection operates on an activation code for organizing choice and ordering of movements; coordination operates on a phase code for organizing the timing of movement execution; intention operates on a spatial code for organizing the muscular forces which drive movement and for mapping them to sensory expectations.

Sections 2-4 provide empirical motivation for coordination, selection, and intention mechanisms, and provide conceptual and mathematical background. Although the logical organization of the model suggests that selection be discussed first, we discuss coordination first for two reasons. One is that is that selection is held to operate on articulatory gestures, which should be understood as a form of coordinative organization; thus it is helpful to establish the concept of a gesture prior to addressing selection. The other reason is that there are important parallels between how gestures are modelled mathematically and certain aspects of selection—these parallels are more readily understood if the reader commands an understanding of articulatory gestures. Section 5 develops an integrated model of the mechanisms and examines how those interactions change in the course of development and in adult speech. Section 6 reviews applications of the model to various developmental, phonological, and experimental patterns.

2. Coordination

The term *coordination* is commonly used to describe mechanisms of organization on a variety of scales. To avoid conflating these mechanisms, three scales of coordination are distinguished: subgestural, gestural, and metrical. The main focus of this article is on gestural coordination, which involves control of the timing of articulatory gestures. Before examining gestural scale coordination, we examine the subgestural scale, which provides a basis for the concept of an articulatory *gesture*. Gestures have a

privileged status in the current approach because of their correspondence with empirical observations of movements in speech.

2.1 The degrees-of-freedom problem and task dynamics

Movement goals are typically accomplished by controlling more than one effector. For example, closing the lips involves a combination of raising the jaw, raising the lower lip, and lowering the upper lip. To model the behavior, it is common to identify effector positions and velocities (or angle and angular velocity for the jaw) as *degrees of freedom*, i.e. as variables which might be controlled. Each of these variables corresponds to a dimension in which the state of the system can be located, and these dimensions are combined orthogonally into a high-dimensional *state space*. The entire set of effectors is often referred to as *a system*. The values of the variables at a given time define the state of the system (*a state vector*), and over time the state vector traces out a trajectory in the state space.

Because movement goals often involve several or more degrees of freedom, this creates a problem: how does the nervous system determine the state space trajectory which should be followed to move the system to a given target position? There are always many different combinations that can achieve the same end result. Moreover, there are normally many muscles involved: production of even a simple syllable like [pa] involves approximately 70 muscles (Abbs & Connor, 1989). At the scale of muscular forces, there are even more degrees of freedom. Each of these muscles is comprised of many muscle fibers and there are many motor neurons whose action potentials control the tension in a given muscle fiber. The state space is profoundly high-dimensional at this scale. Given the need for context-dependent flexibility in how we accomplish movement goals, and the spontaneous variability that we observe in movement patterns, it is unlikely that agents use a detailed, predefined plan that specifies values for all of these degrees of freedom.

The Russian physiologist Nikolai Bernstein was one of the first scientists to emphasize the so-called *many-degrees-of-freedom problem*, i.e. the problem of how to control the many subcomponents of a system to achieve a desired goal (see Bernstein (1967) for English translations of some of his major papers). Bernstein and others proposed a number of constraints on how such control might be simplified by the nervous system. Thorough reviews of the relevant issues can be found in Turvey (1990), who makes the following observations: (i) The nervous system organizes muscles into groups called *synergies*, which should be defined functionally, i.e. relative to a behavioral context, rather than anatomically. (ii) A larger-scale description of synergistic organization involves effectors (limbs, joints, tissues, etc.), which form *coordinative structures* and can cooperate to achieve tasks. (iii) Within a coordinative structure, the system rapidly compensates for perturbation of one subcomponent by altering the behavior of other components. (iv) The phenomenon of rapid compensation for perturbations indicates that coordinative structures make use of feedback from haptic perceptual systems, including muscle stretch receptors.

From the current perspective, *subgestural coordination* refers to mechanisms which govern the control of systems that accomplish some particular task, such forming a labial closure for a bilabial stop or adducting the vocal folds for voicing. Because compensation for perturbations of coordinative structures can be very rapid, subgestural coordination is most likely accomplished via subcortical systems in the brainstem which integrate feedback from muscle fibers. For example, compensatory action of the lower and upper lip occurs as quickly as 30 ms in response to jaw perturbation during production of *bab* (Abbs & Gracco, 1984; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984). Rapid compensation for perturbation is one of the main forms of evidence for coordinative structures, and there are several theoretical frameworks for describing this level of organization: coordinative structures/coordination modes (Balasubramaniam & Turvey, 2004; Fitch, Tuller, & Turvey, 1982); the uncontrolled manifold

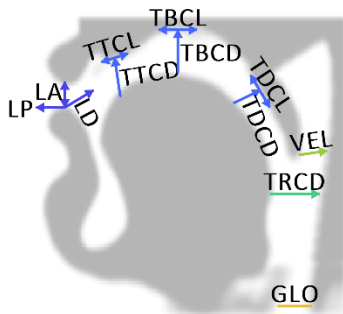
hypothesis (Latash, Scholz, & Schöner, 2002, 2007); the equilibrium point hypothesis (Feldman, 1986; Feldman & Levin, 2009).

Articulatory gestures should be interpreted as a coordinative structures of subgestural scale systems, and models of gestural scale coordination necessarily build upon this notion. The reader might be concerned that, not having definitively resolved the nature of subgestural control, any model of gestural control may be misguided. However, one can proceed with the assumption that understanding gestural coordination is possible because the details of organization mechanisms on higher levels do not depend strongly on details of lower-level organization mechanisms. As Turvey (1977) suggests “...while a representation of an intention in a higher domain is mapped into an immediately lower domain, the particular form that the representation will actually take in the lower domain cannot be known in advance, for the procedures operating in the lower domain have access to knowledge that is immaterial, in principle, to the procedures in the higher domain” (1977: 14).

Task dynamics (TD) is a mathematical framework for modeling control of movement. A key postulate of TD is that complex movement patterns can be usefully decomposed into combinations of simple, relatively abstract tasks, i.e. *gestures*. For example, production of the syllable [pa] can be decomposed into a number of gestures: closing the lips for [p], forming a narrow pharyngeal constriction with the tongue root for [a], abducting the vocal folds to achieve the voicelessness of [p], and adducting the vocal folds to achieve the voicing of [a]. The coordinative mechanisms of the TD model are primarily subgestural, but the model also describes how simultaneously active gestures interact to influence the state of the vocal tract and how multiple gestures are composed into specifications for gestural scale control.

The TD model distinguishes between a task space, which is a state space in which vocal tract states and targets can be defined, and an effector space, which is a space of articulator positions and velocities. Saltzman & Munhall (1989) (henceforth SM89) proposed a comprehensive set of task variables for speech, shown in Fig. 2A below. These are called *vocal tract variables* because they are defined in coordinates which index the anatomical geometry of the vocal tract. The tongue tip (TT), tongue body (TB), and tongue dorsum (TD) are employed with paired constriction degree (CD) and constriction location (CL) variables. There are also lip aperture (LA), lip protrusion (LP), labiodental constriction (LD), velum opening (VEL), tongue root constriction degree (TRCD), and glottal opening (GLO) variables.

A vocal tract variables



B non-zero partial derivatives

	LL x	UL x	LL y	UL y	JAW θ	TT x	TT y	TB x	TB y	TR x	TD x	TD y	VEL x	GLO z
LP	•													
LA		•												
LD			•											
TTCD						•	•	•	•	•				
TTCL						•	•	•	•	•				
TBCD								•	•	•				
TBCL								•	•	•				
TD CD										•	•	•		
TDCL										•	•	•		
VEL													•	
TRCD										•				
GLO														•

C model articulators

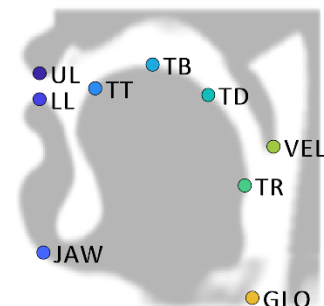


Fig. 2. Task dynamic state spaces and mapping. (A) The task space in which targets are defined consists of vocal tract variables. (B) A matrix of partial derivatives specifies a mapping from changes in vocal tract variables to changes in model articulator positions. (C) The effector space of model articulators.

Each articulatory gesture is associated with a target value for exactly one dimension of the task space of vocal tract variables. For example, a bilabial closure gesture is associated with a lip aperture (LA) target of 0 mm. For variables involving TT, TB, and TD, constriction degree and location gestures are commonly

paired and the constriction location gesture is often omitted from representations. To avoid confusion, it is helpful distinguish between three different uses of the word *gesture*:

Broad sense: *gesture* is used to refer to an articulatory movement or set of movements conducted to accomplish a task. In this sense, a gesture is understood *as a movement*.

Theoretical sense: *gesture* refers to a basic element of phonological representation—an action—and each such action is associated with a set of parameters, one of which is a target, i.e. a specific value of vocal tract variable.

Technical sense: *gesture* refers to a *gestural activation* interval, i.e. period of time in which a set of forces, described by various parameters associated with a gesture, act on a tract variable. These forces drive the tract variable toward an equilibrium value.

The literature commonly presents descriptions in which portions of movement trajectories—measured for example via electromagnetic articulography, X-ray microbeam, or real-time MRI—are referred to as “gestures”, in the broad sense. For current purposes this broad usage of *gesture* is avoided, because empirical observations of tract variables in speech are very often the consequence of multiple gestures in the theoretical/technical senses. Distinguishing between theoretical and technical uses of *gesture* is also worthwhile, because the theoretical sense is associated with a number of additional hypotheses which impose constraints on the nature of gestural parameters.

Saltzman & Munhall (1989) emphasized that the importance of the task dynamic (TD) framework does not lie in the details of the mathematical instantiation of the model, but rather, in the applicability of various concepts in the framework. It is nonetheless instructive to consider how SM89 instantiated these concepts mathematically. Each tract variable was modeled with a second order differential equation of the form in Eq. 1, which describes three forces on a tract variable: a damping force, a restoring force, and a driving force. To more clearly see the decomposition of forces on the system, Eq. 2 recasts the second order equation as two coupled first order differential equations, accomplished by defining the variable $y = \dot{x}$.

$$\text{Eq. 1} \quad m\ddot{x} + \beta\dot{x} + k(x - T(t)) = 0$$

$$\text{Eq. 2} \quad \dot{x} = y$$
$$\dot{y} = -\beta y - kx + kT(t)$$

The equations analogize tract variable dynamics to the dynamics of a driven, damped mass-spring system. The variable x is the displacement of the mass from 0, and the variable y is the first derivative of x . The mass m (an inertial parameter) is assumed to have value of 1 and thus can be omitted from the equations. The term $kT(t)$ is a driving force, and the value of $T(t)$ can be conceptualized as an equilibrium length of the spring, i.e. a target value for x . Changing T corresponds to altering the equilibrium length of the spring. Setting both of the first order equations to zero shows that equilibrium is achieved when $y = 0$ and $x = T$.

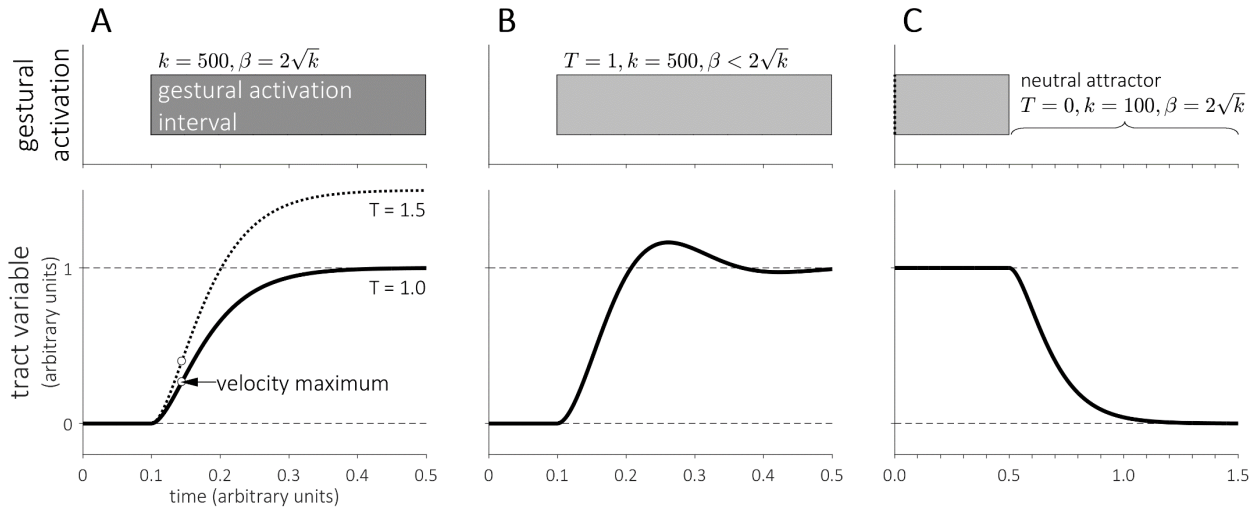


Fig. 3. Influence of gestural activation on a tract variable. (A) Critically damped gestures with different target parameters, $T=1.0$ and $T=1.5$. (B) An underdamped gesture. (C) Influence of the neutral attractor in the absence of gestural activation.

In the TD model, each articulatory gesture is associated with a specific value of the target parameter T . The key idea is that gestures become active, i.e. “turn on”, and this changes the parameter T in the associated tract variable equation. A gesture can thus be conceptualized as the presence of a force that acts to drive a tract variable to a new target position. Each gesture is also associated with *stiffness* parameter, k , which determines the strength of the restoring force, and a damping parameter, β , which determines the strength of the damping force. The stiffness parameter influences how quickly the tract variable changes toward the new target, and the damping parameter is hypothesized to be $2k^{(1/2)}$, i.e. *critical damping*, which results in the tract variable approaching the target as quickly as possible without overshoot. Fig. 3A compares the evolution of a generic tract variable when gestures with different targets ($T=1$ and $T=1.5$) are active. Fig. 3B shows an example of a underdamped gesture, which exhibits a decaying oscillation around the target. The hypothesis of critical damping is motivated by the empirical observation that tract variables exhibit the critically damped pattern, not the underdamped one.

Both stiffness and movement range influence the evolution of a tract variable. Movement range is the distance between an initial condition and target at the time when a gesture becomes active. Stiffness can only be determined from an empirically observed maximal velocity when (i) the movement range is known, (ii) critical damping is assumed, and (iii) the target is assumed to change instantaneously to a constant value at the onset of gestural activation. In contrast to the value of maximal velocity, the period of time from the onset of gestural activation to the timepoint of maximal velocity is fully determined by stiffness and damping, independent of movement range. It is often assumed that the target, stiffness, and damping parameters associated with a gesture are invariant for a given speaker, but these hypotheses generally have not been verified by empirical work. The reader should note that parametric invariance is not essential in a gestural model.

An important feature of the TD model is that it distinguishes the dynamics of one-dimensional task variables from control of the articulators which are used to accomplish tasks. For example, the dynamics of the tract variable lip aperture (LA) are distinct from the dynamics of the jaw (JAW), lower lip (LL), and upper lip (UL), the combined states of which determine lip aperture. The TD model generates articulator movement patterns by mapping changes in task variables to changes in model articulators. This is accomplished via a matrix of partial derivatives (called an inverse Jacobian matrix). As shown in Fig. 2B, each row of the matrix corresponds to a task variable, and each column to a model articulator. Each cell

is a partial derivative; for example, the rate of change in the model articulator JAW per unit of change in the task variable LA. The inverse Jacobian is a subgestural manifestation of the many-degrees-of-freedom problem: how are the values of the partial derivatives determined? SM89 used equal weightings of model articulators for a given tract variable; subsequent work has attempted to estimate the weights from empirical observations of the forward Jacobian, i.e. observed changes in task variables which arise from measured changes in articulator positions (Lammert, Goldstein, & Iskarous, 2010; Sorensen, Toutios, Goldstein, & Narayanan, 2016). Some studies suggest that these weightings can vary within a given speaker (Tilsen, 2017).

When multiple gestures are active and those gestures exert forces on the same tract variable, or when multiple gestures are active and their associated tract variables are coupled to the same model articulator(s), additional mechanisms of blending and parameter tuning are necessary to generate empirically adequate movement profiles. Implementations of these blending and tuning mechanisms are detailed in SM89, but they are challenging to test empirically because they require assumptions about the invariance of gestural parameters and inverse Jacobian weightings.

The TD model also holds that in the absence of influences from active gestures, all tract variables evolve according to a default/neutral attractor, as shown in Fig. 3C. The neutral attractor can be understood as a *speech-ready posture* of the vocal tract (or *articulatory setting*) and differs from resting posture (Ramanarayanan, Goldstein, Byrd, & Narayanan, 2013; Ramanarayanan et al., 2013; Ramanarayanan, Lammert, Goldstein, & Narayanan, 2014). One finding that may be problematic for the notion of a default speech ready posture is that prior to utterance initiation, the vocal tract can be anticipatorily postured with subtle assimilation to the targets of upcoming gestures (Tilsen et al., 2016).

The TD model is very useful because of its explicitness, conceptual simplicity, and ability to generalize to a wide range of contexts. However, TD provides a somewhat ad-hoc account of gestural timing, i.e. when gestures become active and when they cease to be active. Early work in the framework held that gestural activation intervals were lexically specified in *gestural scores* (Browman, Goldstein, Saltzman, & Smith, 1986; Saltzman, Goldstein, Browman, & Rubin, 1988). Systematic accounts of gestural timing were subsequently developed in the theory of Articulatory Phonology.

2.2 Articulatory Phonology

Mainstream phonological theories analyze speech as a structure of abstract categories or features. In contrast, Articulatory Phonology (AP) is based on the premise that speech is an organization of actions (cf. Goldstein & Fowler, 2003).. In AP, gestures are taken to be the fundamental combinatorial elements of speech. It is important to recognize that the concept of a gesture and the traditional alternatives—segments or features—are constructed from entirely different conceptual metaphors. Segments and features are based on the *units-are-objects* metaphor, which promotes reasoning about speech with spatial relations and patterns of connection; such reasoning often adopts a mapping from spatial arrangement to temporal order—the so-called “beads on a string” picture (Saltzman & Munhall, 1989). In contrast, gestures are based on the metaphor that *actions-are-trajectories in a state space*, and this promotes conceptualizing the production of speech as a temporal phenomenon in which various forces are responsible for the organization and control of movement.

As a consequence of viewing gestures (actions) as the fundamental elements of speech, AP naturally invites agnosticism regarding more complex patterns of organization, e.g.

“While it is clear that gestures must themselves be coordinated in larger scale stable patterns, it is not yet clear what these patterns are. Nor is it clear to what phonological

units (if any) such larger scale stability might correspond: consonant clusters, onsets and rimes, syllables, words, phrases, etc.” (Browman & Goldstein, 1988).

There is currently no consensus on what role traditional units have in relation to gestures. One view, espoused in Fowler (1980), identifies phonological segments with sets of coordinative structures, with different parameter values being nearly equivalent to distinctive features. A more flexible view is expressed by Goldstein & Fowler (2003), who state that “bonding strengths” in “molecules” of gestures “can be used to define a hierarchy of unit types, including segments, onset and rimes, syllables, feet, and words”. Hence gestural parameters are not necessarily identified with traditional phonological features. Note that terms such as “bonding,” “molecule,” and “constellation” are not ideal for describing interactions between gestures and their temporal organization, since these terms tend to evoke the units-are-objects metaphor. *Phase coupling forces* and *network of interactions* are preferable terms because, as we see below, conceptual and mathematical models of gestural interactions make use of forces that relate to oscillation phases rather than object positions.

An important innovation of the AP/TD framework is a conceptual blend of linear time and cyclic time. This blend provides a basis for reasoning about gestural timing which is not available in segmental/featural approaches. As shown in Fig. 4A, different absolute durations of linear time can be mapped to a common normalized interval of time, where the temporal coordinate is referred to as *phase*. This can be blended with a cyclic conception of time (Fig. 4B) by mapping the start of a movement to an arbitrary phase angle (e.g. 0° or 0 radians) and mapping the achievement of a movement target to another phase, such as 270° or 3π/2 radians (Browman & Goldstein, 1995: 183). The temporal normalization parameter is $\omega = 2\pi f$, an angular frequency.

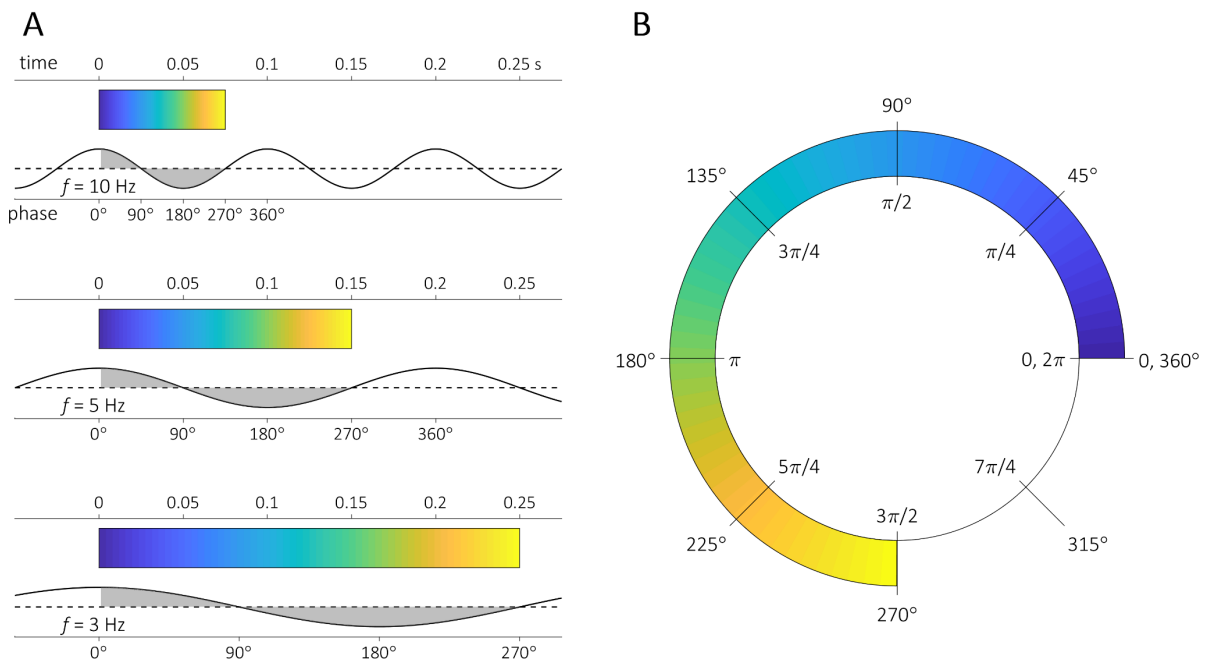


Fig. 4. Linear-cyclic time blend. (A) Three gestures with different durations mapped to the same phase interval. (B) The mapping of a gesture in phase coordinates to a unit circle.

The linear-cyclic time blend forms the basis for the concept of gestural *phasing* in AP/TD, where each gestural activation interval has a normalized, intrinsic time frame. Control of the timing of the initiation

of the activation interval of gesture B relative to the activation interval of gesture A derives from a specification of the target phase of A at which B is initiated. One advantage of this is that timing is not specified in absolute units of seconds, which is desirable given the vast range of factors which can systematically influence absolute durations in movement timing. A deeper advantage is that a common coordinate system for timing control can be applied to a set of local interactions between gestures.

The linear-cyclic time blend for conceptualizing movement timing appears at least as early as Greene (1972) and Turvey (1977), and provides a basis for the notion of *intrinsic* time in Fowler (1980), which is understood in opposition to *extrinsic* time. In extrinsic timing models, timing is excluded from the mental representation of the articulatory plan. Fowler argued that the phenomenon of coarticulation, i.e. temporal overlap of articulatory movements, cannot be adequately understood with such models; instead, an adequate model must incorporate timing information directly into the representation of the articulatory plan. Fowler further emphasized that co-production of coordinative structures (gestures) involves temporal overlap. She used the phrase “patterned in time” (1980: 127) in discussing the results of studies showing invariance of normalized linear timing patterns (i.e. invariant relative phasing). This conception of phasing plays a fundamental role in subsequent empirical and theoretical work (e.g. Browman & Goldstein, 1988, 1990, 1992, 1995; Byrd, 1995).

2.3 The coupled oscillators model and syllable structure

More recently, the AP/TD model has been extended by incorporating systems of coupled oscillators to generate phasing relations between gestures. This extension is based on the notions (i) that the cycles associated with gestures may be iterated for multiple periods, (ii) that there are systems associated with those cycles—i.e. gestural planning oscillators (henceforth GPOs), and (iii) that these oscillators interact through phase coupling forces. One influential model which motivates this extension is the Haken, Kelso, & Bunz (1985) coupled oscillators model (henceforth HKB85), in which the relative timing of finger wagging movements is organized through relative phase coupling. HKB85 developed a phenomenological model of the system of fingers and their relative phase. The model describes the empirical observation that anti-phase timing transitions to in-phase timing as wagging frequency increases. The model of the transition is schematized in Fig. 5. Fig. 5A shows the positions of the fingers, and Fig. 5B shows their relative phase, defined as the difference of individual phases. The model imposes a loss of stability of the anti-phase mode of coordination by constructing a relative phase potential function that changes as movement frequency (ω) increases. Stable modes of coordination correspond to valleys in the relative phase potential functions in Fig. 5C, which is comprised of an in-phase component, $a \cos(\omega\phi)$ and an anti-phase component, $b \cos(2\omega\phi)$. HKB85 asserted that the ratio of the amplitudes of these components, b/a , decreases as ω increases, which eventually results in the anti-phase mode becoming unstable.

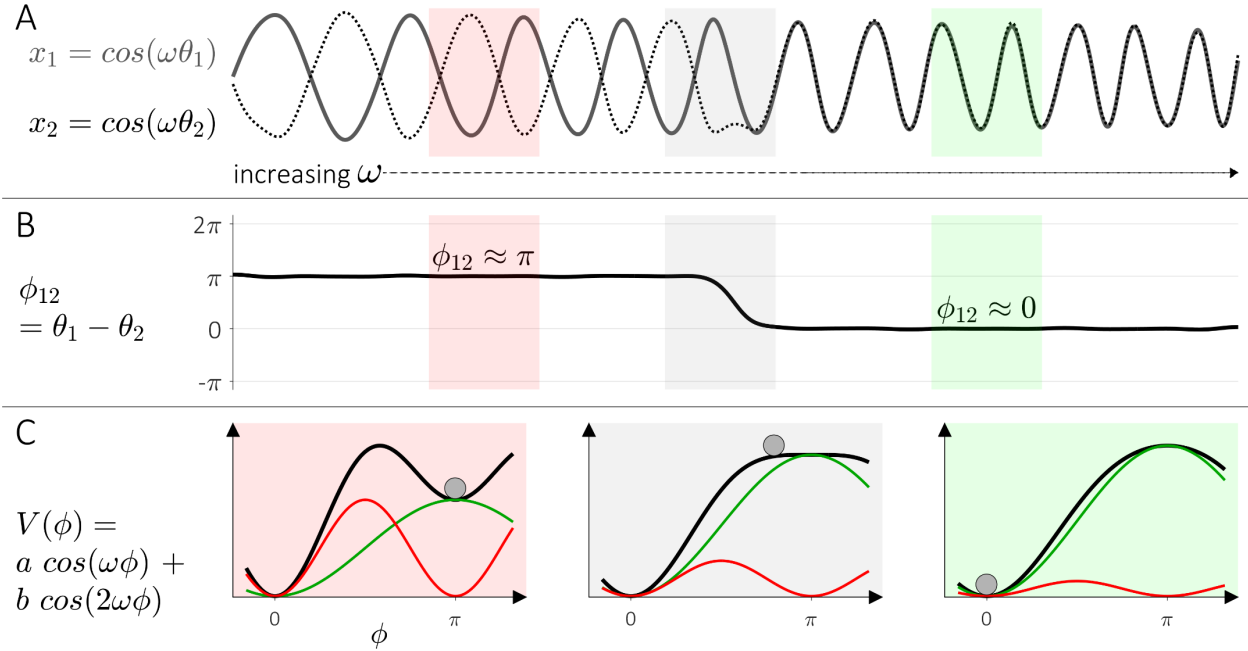


Fig. 5. Haken-Kelso-Bunz model of finger wagging phase transition. (A) Finger positions over time as wagging frequency increases. (B) Relative phase of fingers. (C) Relative phase potential functions and potential function components before, during, and after the phase transition.

The HKB85 coupled oscillators model of finger wagging is compelling, but articulatory gestures are not rhythmic movements. Hence a conceptual innovation is necessary to apply a mechanism of coupled oscillations to non-oscillatory movements. One early appearance of this innovation may be in Kelso & Tuller (1987), where the authors explicitly associate each gesture with a “virtual cycle”, but this can alternatively be interpreted as description of the linear/cyclic time blend. A more explicit connection is made in Browman & Goldstein (1991), who suggested that patterns of gestural overlap could be understood by extending research on bimanual rhythmic movements which demonstrates the existence of stable in-phase and anti-phase coordinative modes. The idea is subsequently elaborated in (Browman & Goldstein, 2000; Goldstein, Byrd, & Saltzman, 2006; Saltzman & Byrd, 2000).

In the coupled oscillators model, each gesture is associated with a *gestural planning oscillator* (GPO). The state of a GPO is described by a phase angle, θ , which is 2π -periodic. The equation of motion for GPOs is shown in Eq. 3. The first term is an intrinsic frequency (f_i) for GPO i . A common simplifying hypothesis is that f_i is the same for all GPOs. The second term is a sum over relative phase coupling forces. The force is defined as the opposite of the derivative of a relative phase potential function, $V(\phi)$, defined in Eq. 4. Relative phase ϕ is a difference of phase angles, i.e. $\phi_{ij} = \theta_i - \theta_j$, and C_{ij} is a matrix of coupling force strengths.

$$\text{Eq. 3} \quad \dot{\theta}_i = 2\pi f_i + \sum_j C_{ij} \frac{-dV(\phi_{ij})}{d\phi_{ij}}$$

$$\text{Eq. 4} \quad V(\phi_{ij}) = -C_{ij} \cos(\phi_{ij})$$

An in-phase coupling force corresponds to the case when $C_{ij} > 0$. The potential function for in-phase coupling between an onset consonantal GPO and vocalic GPO is shown in Fig. 6A, and the corresponding

coupling force is shown in Fig. 6B. Note that the potential is periodic and has minima at integer multiples of 2π radians. These minima correspond to ϕ values where the force on ϕ is zero. The potential can be interpreted schematically as follows: the variable ϕ experiences a force that results in a decrease of $V(\phi)$, as if ϕ were a ball rolling down a hill in the $V(\phi)$ landscape. The ball will slow to a halt at a local minimum of $V(\phi)$, as if it were submerged in a viscous fluid that perfectly compensates for inertial forces.

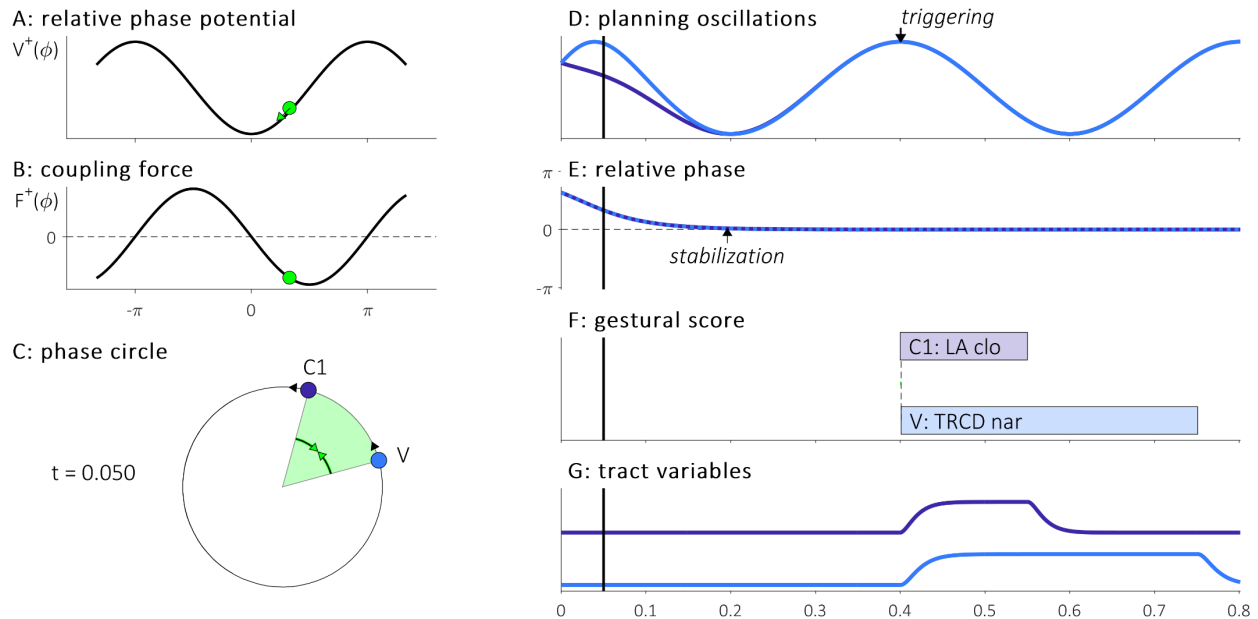


Fig. 6. In-phase coupling in a CV syllable. (A) in-phase relative phase potential. (B) in-phase coupling force. (C) C_1 and V GPO phases at time $t=0.050$ s. (D) planning oscillations, defined as $\cos(\theta)$. (E) relative phase, $\phi_{C1V} = \theta_{C1} - \theta_V$. (F) gestural score. (G) tract variables.

The effect of the in-phase coupling force on the phases of GPOs is to pull their phases closer together. As pictured on a phase circle in Fig. 6C, the phases of the GPOs rotate counterclockwise, due to their intrinsic frequency; the relative phase coupling force (green arrows) speeds up or slows down these rotations until relative phase stabilizes. Viewed over time, the effect of the force is to synchronize GPO oscillations, which are shown as $\cos(\theta_i)$ in Fig. 6D. This corresponds to an evolution of relative phase to 0, shown in Fig. 6E. Note that an arbitrary initial condition was chosen for Fig. 6.

A key link between the coupled oscillators model and TD/AP is the *stabilization-triggering hypothesis* (cf. Nam, 2007; Nam, Goldstein, & Saltzman, 2009). This hypothesis holds that once the entire system of GPOs associated with an utterance has stabilized, the gestures associated with those GPOs are triggered (i.e. become active) at some particular phase of the GPO cycle (the choice of triggering phase is arbitrary but must be the same for all gestures). The moment of relative phase stabilization is indicated with an arrow in Fig. 6E, and the moment when both gestures are triggered is shown in Fig. 6D. Notice that this corresponds to the onset of activation of the corresponding gestures in the gestural score (Fig. 6F), and accordingly a change in tract variables (Fig. 6G).

An important assumption of the stabilization-triggering hypothesis is that no gestures become active until all GPOs have stabilized. Thus there is an expected “settling time” between the moment when GPOs begin oscillating and the initiation of gestural activation (Nam, 2007; Nam et al., 2009). Predictions of this hypothesis have been explored empirically in Mooshammer et al. (2012) and Tilsen (2014), but currently the stabilization hypothesis is primarily conjectural. An alternative interpretation is that GPOs describe the net effects of forces acting on supra-utterance scales, which serve to balance perceptual recoverability

of gestures and rapidity of information transmission. In this view, the c-center effect and the empirical manifestation of the basic coupling hypotheses are epiphenomena of communicative forces which do not directly influence timing on utterance scales.

The in-phase coupling force between onset consonantal and vocalic GPOs contrasts with an anti-phase coupling force between coda consonantal and vocalic GPOs. An anti-phase potential and coupling force is shown in Fig. 7A and Fig. 7B. The anti-phase coupling force corresponds to the condition in which $C_{ij} < 0$ in Eq. 4. In this case, the minima of the potential are at relative phases $\pi(2n + 1)$, the coupling force has stable equilibria at $\pm\pi$, and GPO phases are driven to a separation of π radians. The stabilization-triggering hypothesis predicts a temporal difference of movement initiation $(2f)^{-1}$.

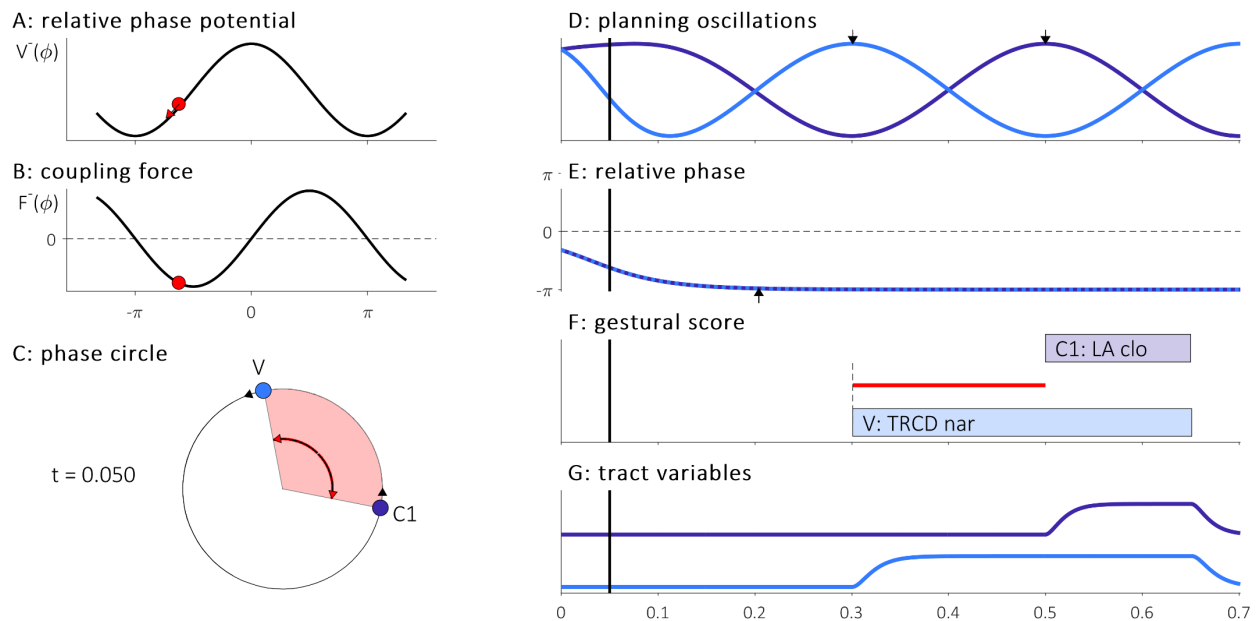


Fig. 7. Anti-phase coupling for VC syllable. See Fig. 6 for description of panels.

One thing to emphasize about the coupled oscillators model is that, in its most basic form, it is a model of only the *initiation* of gestural activation (and in effect, of movement initiation). It does not specify how long a given gesture remains active, and it does not specify control over an initiation relative to any other type of event which is not initiational, such as a target achievement.

The coupled oscillators model of gestural coordination is appealing because it is able to derive patterns of intergestural timing from a small set of hypotheses regarding how GPOs are coupled, according to their syllable positions and associations with consonants and vowels. Patterns of relative phasing in gestural scores can be understood in a systematic way, rather than being specified in an ad hoc manner. Relatedly, the model provides a mechanism for organizing control of timing on-line, as opposed to requiring long-term memory (i.e. lexical) storage of relative phase relations.

The basic coupling hypotheses of the AP coupled oscillators model are that (i) CV coordination is governed by in-phase coupling, and (ii) all other forms of coordination such as VC, VC, and CC are governed by anti-phase coupling (cf. Goldstein, Byrd, & Saltzman, 2006). One of the important successes of the model is that these basic hypotheses can be combined to generate a specific pattern of timing observed in complex onsets, known as the *c-center effect* (Browman & Goldstein, 1988), of which neither extrinsic timing models nor ad-hoc intrinsic phasing models can provide a satisfactory account.

The c-center effect is a phenomenon whereby the initiations of the consonantal gestures in a complex syllable onset are equally displaced in opposite directions in time from the initiation of the vocalic gesture (Browman & Goldstein, 1988). The pattern is shown in the gestural score in Fig. 8F. Prior to the development of the coupled oscillators model, the pattern was described as a global organization. Browman & Goldstein (1988) hypothesized that the c-center, defined as the temporal average of the onset consonantal gestures (and hence a “global” measure) is phased to a point somewhere early in the vocalic gesture. There are various reasons why it is not straightforward to detect the c-center pattern empirically, one of which is that gestural blending between onset consonantal gestures and the vocalic gesture can bias measurements of gestural initiation. Hence the pattern has often been identified indirectly by measuring the timing of consonantal target achievements relative to some acoustic or articulatory landmark associated with the vocalic target or a post-vocalic consonant (cf. Tilsen et al., 2012 for further detail on the methodological issues).

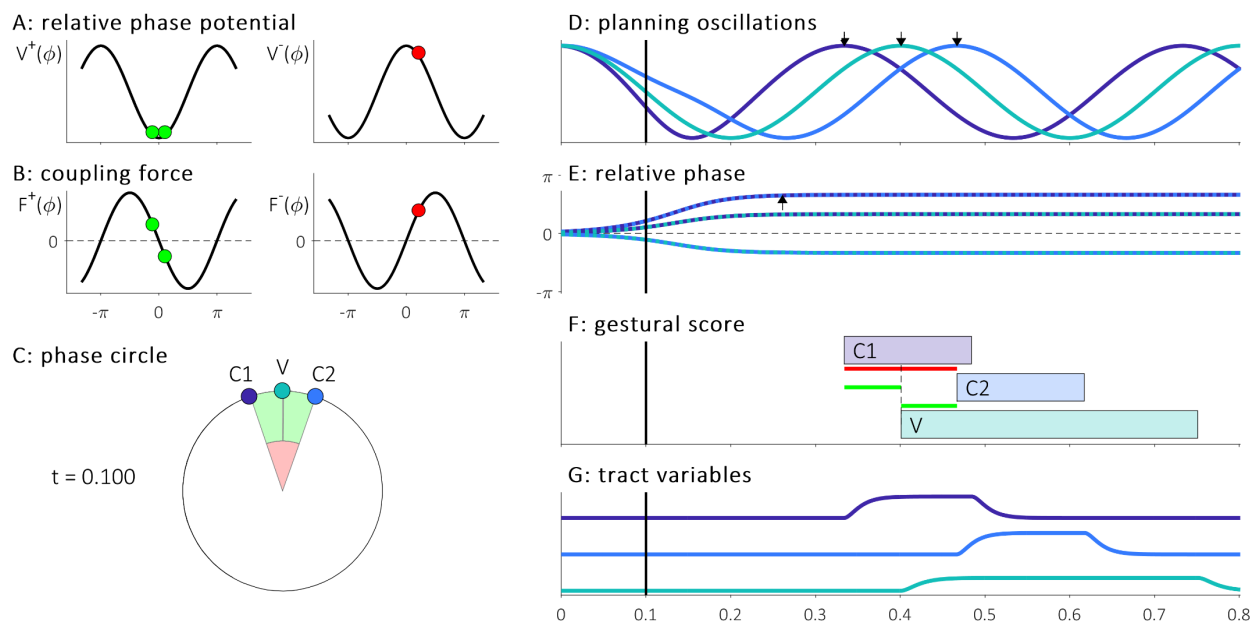


Fig. 8. The c-center effect as a balance of in-phase and anti-phase coupling forces. In-phase and anti-phase potential functions and forces are shown in (A) and (B), respectively. The stable relative phases of the system arise from a balance between in-phase and anti-phase forces.

A major appeal of the coupled oscillators model is the fact that combining just two types of forces, in-phase coupling and anti-phase coupling, generates the c-center effect. However, as we discuss below, these hypotheses incur subtle contradictions when applied simultaneously to post-vocalic and pre-vocalic coordination. These contradictions can be resolved in a number of ways, all of which compromise the simplicity of the model to some degree. One issue which underlies this discussion is whether there are only two stable modes of relative phase coupling, or whether any arbitrary relative phase target is possible. An alternative form of the relative phase potential is shown in Eq. 5, where the parameter Φ_{ij} specifies a relative phase target (Nam & Saltzman, 2003; Saltzman & Byrd, 2000). If Φ is constrained to be either 0 or $\pm\pi$ and $c_{ij} > 0$, the potential is equivalent to the one in Eq. 4. However, if Φ can take any arbitrary value as implied by Eq. 5, then the model is much less constrained.

$$\text{Eq. 5 } V(\phi) = c_{ij} \cos(\phi - \Phi_{ij})$$

The contradiction that arises between post-vocalic and pre-vocalic coordination can be understood by considering the frequency parameter f_i in Eq. 3, which maps from intergestural phasing to intergestural timing. Recall that f_i is assumed for simplicity to be the same for all GPOs associated with a given score. This assumption of uniform intrinsic frequency is desirable for a number of reasons: (i) the model is more constrained, (ii) allowing oscillator frequencies to differ introduces an ambiguity in the mapping from relative phase to time differences, and (iii) anti-phase coupling no longer exhibits a phase difference of $\pm\pi$ when $f_i \neq f_j$.

The problem that arises under the uniform intrinsic frequency assumption is evident when considering a $C_1C_2V_3C_4$ syllable. The frequencies $f_1, f_2,$ and f_3 which are necessary to generate empirically plausible intergestural timing intervals for a c-center effect are too small to be consistent with the frequencies f_3 and f_4 which are necessary to generate the full range of empirically plausible timing intervals between vocalic and coda gestural initiation. To see this, take Δt_{ons} as the time from C_1 to C_2 movement initiation and Δt_{cod} as the time from V_3 to C_4 movement initiation. Typical values of Δt_{ons} and Δt_{cod} could be 0.125 s and 0.250 s, respectively (cf. Tilsen, 2014). The relative phase (ϕ_{12}) of C_1 and C_2 is determined by GPO frequency and the strength of anti-phase coupling relative to in-phase coupling (Tilsen, 2017). For a constant Δt_{ons} , as anti-phase coupling strength increases, ϕ expands and hence GPO frequency (f) must increase. Because $\Delta t = \phi/(2\pi f)$ and the maximal ϕ is π , the maximal f that can generate $\Delta t_{\text{ons}} = 0.125$ is 4.0 Hz. The relative phase ϕ_{34} of V_3 and C_4 necessarily stabilizes to π radians because these GPOs interact only through anti-phase coupling. Thus the maximal f to generate Δt_{cod} is approximately 2.0 Hz. The value 2.0 Hz is somewhat low, in a behavioral sense, and the discrepancy between maximal f for onset and coda timing raises the question of whether the coupled oscillators model is always plausible for generating post-vocalic movement patterns.

There are a number of potential solutions to this problem. One is to allow for arbitrary values of relative phase targets for VC coupling, potentially even ones that exceed π . This is somewhat undesirable because the model becomes overly powerful and in effect a proxy for describing durations. Another solution is to relax the assumption of uniform intrinsic frequency. This solution also leads to an arguably over-powered model, unless some alternative source of constraint on f can be identified. One possible source of constraint is the stiffness parameter of gestures in the TD model. Recall that tract variables behave like damped mass-spring systems. Each gesture is associated with a restoring force for a tract variable. Although gestures specify critically damped trajectories, these imply an oscillation in the absence of damping. Hence we might take f_i as the natural frequency associated with the stiffness parameters of gestures, i.e. $f = k^{1/2}$. Given that onset and coda gestures are associated with similarly rapid movements compared to vocalic gestures, this suggests for $C_1C_2V_3C_4$ a relation $f_1 \approx f_2 \approx f_4 < f_3$. It is unclear whether these constraints will produce patterns that are consistent with empirical ones, and to my knowledge no systematic investigation of such constraints has been undertaken.

There are a few other issues in simulating empirical patterns with the coupled oscillators model. One is that in systems of three or more oscillators, there are combinations of initial relative phases which cannot converge to the desired pattern. For example, to model the c-center effect in C_1C_2V coordination, only the subset of initial conditions which satisfy $\phi_{C_1V} > 0 > \phi_{C_2V}$ can evolve to the empirical pattern in which gestural initiations obtain temporal order C_1, V, C_2 . Thus initial conditions of GPOs must be lexically stored or some other mechanism is needed for insuring appropriate initial conditions. One ambiguity which is not addressed in the above discussion (but see Tilsen, 2017; Tilsen & Goldstein, 2012) is whether gestural modeling applies only to the constriction formation phase of a consonant, or whether constriction releases should also be modeled as gestures. Another open question is whether there is any theoretical motivation for distinguishing between consonantal and vocalic gestures. The apparent differences between consonantal and vocalic gestures could derive from differences in the relative

importance of various modalities of sensory feedback associated with the achievement of gestural targets. In that case, the use of “C” and “V” symbols to refer to gestures should be seen as a matter of convenience, rather than a theoretical distinction.

3. Selection

The term *selection* encompasses a set of mechanisms which interact to regulate both the choice of which articulatory gestures to execute and the sequential ordering of those choices. Sensory feedback plays an important role in selection. One of the key tenets of any useful model of selection is that action plans are active in parallel, since this is necessary to account for behavioral phenomena such as sequencing errors and effects of sequence length on reaction time to initiate a response.

3.1 Associative chains vs. parallel planning

The early behaviorist conception of motor sequencing was based on classical conditioning: movement sequences were viewed as *associative chains*, i.e. “chains of reflexes, in which the performance of each element of the series provides excitation of the next”, and it was held that “verbal thought is a simple chain of central processes in which each element serves to arouse the next by direct association” (Lashley, 1951). Lashley (1951) identified a number of problems with the associative chain conception: (1) because movement elements can occur in arbitrary orders (e.g. *tire vs. right*), their order must be organized through a mechanism which does not rely on direct associative connections; (2) relations between non-adjacent units are observed, i.e. higher-order/long-distance relations, and these are not expected from first-order, local associations in a chain; (3) sequencing errors (e.g. Spoonerisms such as *pleech spanning*) are not expected from associative chains because such errors require the mechanism of ordering to be independent of content selection; (4) people can easily learn and generalize new sequencing patterns, as in pig Latin, and such ease of learning is unexpected if direct associations must be formed; (5) people can engage in subvocal speech in which feedback is unavailable. For an alternative, Lashley suggested that actions “are readied or partially activated before the order is imposed upon them” and that “some scanning mechanism must be at play in regulating their temporal sequence” (1951: 130). We refer to this conception as *parallel planning with a selection mechanism*.

A specific model of parallel planning with two main processes, subprogram selection and command, was developed by (Sternberg, Knoll, Monsell, & Wright, 1988; Sternberg, Monsell, Knoll, & Wright, 1978). In this model, a motor program (i.e. sequence of words) is comprised of subprograms (words), which differ in their complexity (i.e. number of syllables). When the program is prepared, each subprogram is marked for its serial order position and stored in an unordered buffer. Response production then involves an alternating series of selection and command processes, as shown Fig. 9A. Selection involves a search through the buffer for the next element of the sequence. Command involves execution of movements associated with the selected element. The duration of the selection process is held to depend on the number of units in the buffer (n), and the duration of each command process is held to depend on complexity of each unit (i.e. number of syllables in each word). The model predicts a linear dependence of reaction time (retrieval of first unit) on n (Fig. 9B) and additive, linear dependence of the average interunit duration on n and on unit complexity (Fig. 9C). In an extensive series of experiments on speech and typing using a delayed response paradigm, Sternberg and colleagues found compelling evidence for both of these effects (Sternberg et al., 1988, 1978).

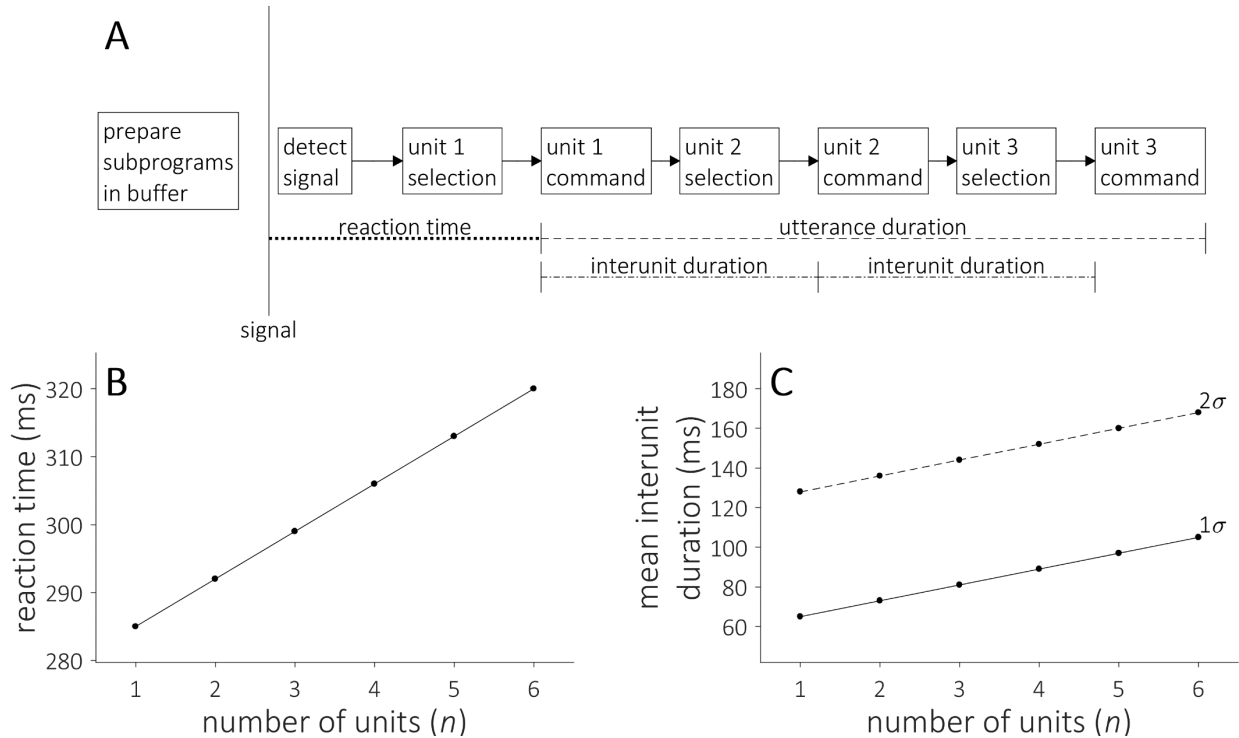


Fig. 9. Subprogram selection/command model and predicted empirical phenomena, adapted from Sternberg et al. (1988). (A) model schematization. (B) Linear effect of number of units on reaction time. (C) Linear effect of number of units on mean unit duration.

An important question which arises in the context of the subprogram selection model regards the size of each subprogram. Sternberg et al. (1988) presented data indicating that each subprogram corresponds to a metrical foot/stress group, and subsequent studies indicate that subprograms correspond to phonological words (Wheeldon & Lahiri, 1997, 2002). One shortcoming of the subprogram selection model is that it does not provide an explicit model of how selection operates.

3.2 Competitive queuing

A connectionist architecture with a more explicit model of selection was developed by Grossberg (1978, 1987). Models of this sort have been referred to as “competitive queuing” (CQ) models (Bullock, 2004; Bullock & Rhodes, 2002; Houghton & Hartley, 1995). The details of specific models are fairly complicated, so here we focus on several important features. Each motor program is associated with a node in a network, and the pattern of relative activation of nodes prior to the initiation of a sequence determines the order in which motor programs are selected. In other words, an relative activation pattern encodes temporal sequencing information. Fig. 10 shows how this works: upon initiation a of sequence, there is a competition process in which an external input causes node activations to grow until one of them exceeds a selection threshold. The motor programs associated with the selected node are then executed, and during this time competitor nodes are gated, i.e. their activation does not increase. Feedback regarding achievement of the target(s) of the selected motor programs subsequently induces suppression of the corresponding node, and the competition process resumes.

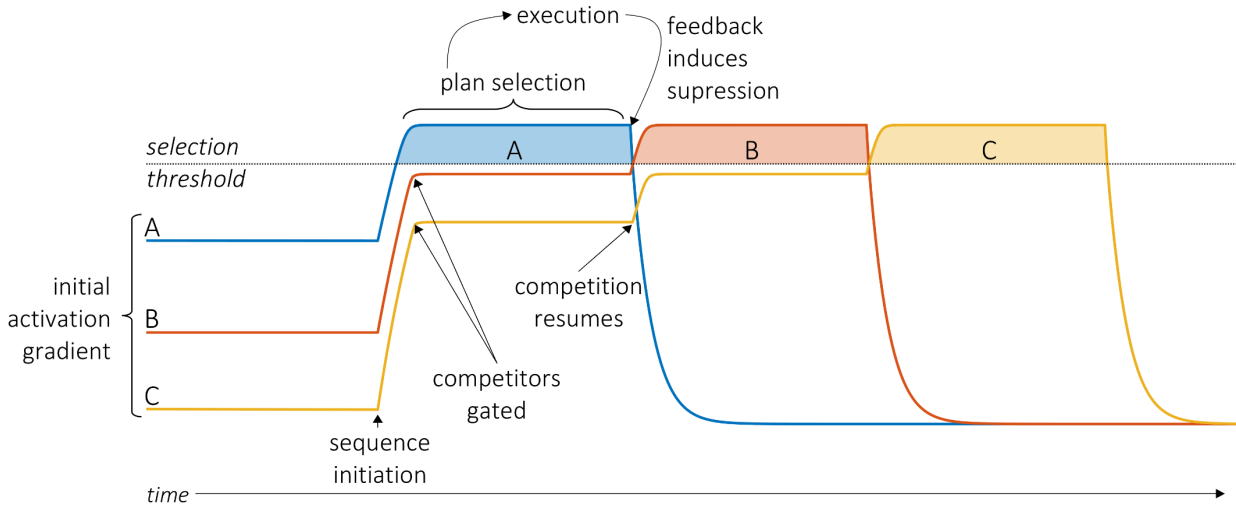


Fig. 10. Illustration of competitive queuing of three motor programs, {A}, {B}, and {C}. An activation gradient determines an order of selection. Competition, gating, feedback, and suppression processes govern sequential selection.

A variety of sequencing errors can be readily understood in the CQ framework. Substitution errors occur when the activation gradient deviates from an expected pattern, and this could happen for a number of reasons, such as noise or external influences on activation. Another type of pattern, called *errorful reselection*, is highly relevant to certain developmental phenomena that we examine later on. As shown in Fig. 11A, node {3} has deficient activation and node {2} is selected a second time, at the expense of {3}. Furthermore, Tilsen (2016) hypothesized that reselection could be an intentional mechanism for production of long vowels and geminates. As shown in Fig. 11B, the *intentional reselection* of {2} prolongs the period of time in which {2} exerts control over the vocal tract. The reader should note a reselection event does not entail a repetition of a movement, but rather a prolongation of the period of time during which a target state of the vocal tract is achieved.

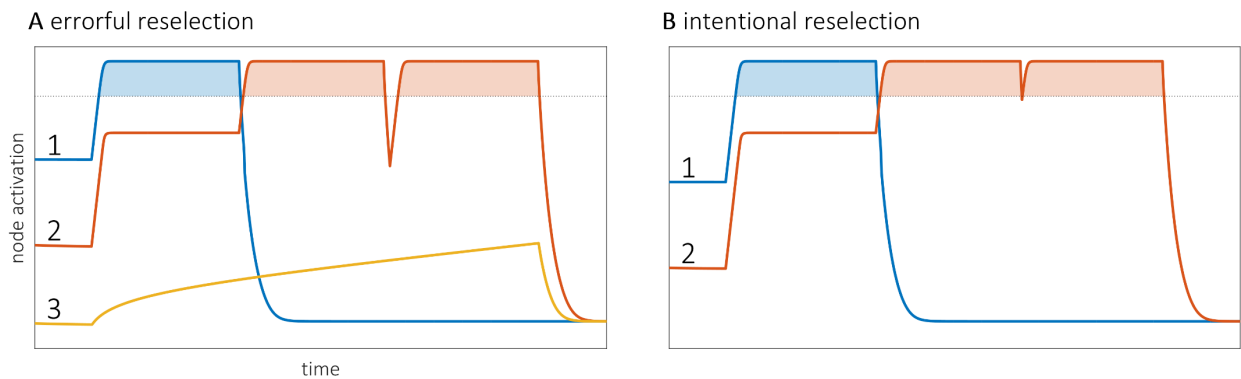


Fig. 11 Errorful reselection and intentional reselection. (A) Program {2} is errorfully reselected at the expense of plan {3}. (B) Program {2} is intentionally reselected.

In CQ models a normalization mechanism is responsible for organizing the relative activation of nodes. Grossberg (1978) argued that on-center off-surround networks can accomplish the normalization. In such networks, global increases in external input are balanced against increases in the inhibitory influences between units, and hence a constant pattern of proportional excitation can be maintained. However, some other mechanism of activation regulation may be required when external input varies for each unit.

A relatively abstract approach to modeling this activation-organizing mechanism is shown in Fig. 12, where unit activations are arranged in a quantal step-potential. As with the relative phase potential governing gestural planning oscillators (cf. Section 2.3), the opposite of the derivative of the potential specifies a force, which in this case acts on activation.

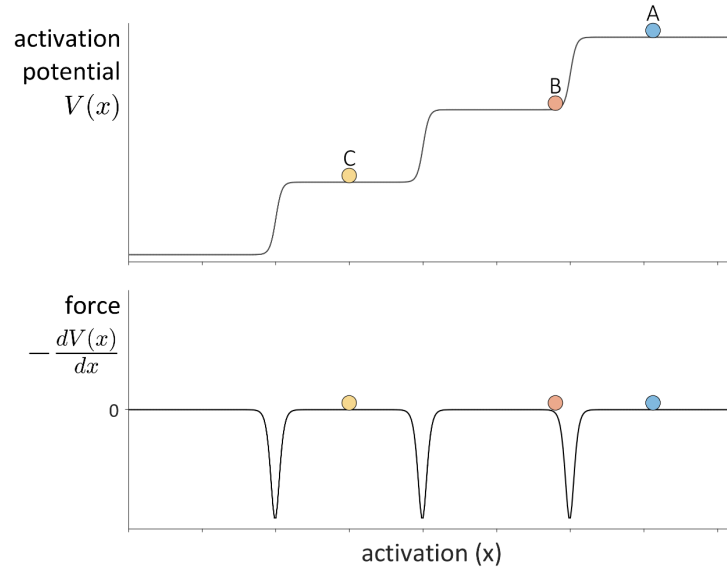


Fig. 12. Step potential conceptualization of activation organization. The opposite of the derivative of the activation potential specifies forces which promote a pattern of relative activation. In the above example, to increase the activation of plan B, a strong positive force on B would be required, in order to overcome the strong negative force described by the potential.

The activation organization mechanism depicted in Fig. 12 should be interpreted as global characterization of node-specific activation potentials, each of which represents a sum over the potentials which are associated with inhibitory interactions between competing nodes. The step potential for a given node is modeled as a sum over sigmoid functions as in Eq. 6. The midpoint of each sigmoid is $x_j - \sigma$. In other words, there is an activation barrier between nodes i and j , and the location of that barrier in the activation coordinate is $x_j - \sigma$. The parameter σ spreads the steps out in activation space.

$$\text{Eq. 6} \quad V(x_i) = \sum_{j \neq i} -c_{i,j} \left[1 + e^{-b(x_i - (x_j - \sigma))} \right]^{-1}$$

To facilitate subsequent discussion of how selection and coordination interact, an abstract form of representation is adopted here in which changes in activation associated with competition and feedback-induced suppression correspond to abrupt reorganizations of the quantal potential. Thus the activation trajectories of Fig. 10 are re-envisioned as the output of the discrete sequence of ordered states as shown in Fig. 13. Prior to initiation [1], all nodes are below a selection threshold, which corresponds to a state in which the top-most activation level is unoccupied. After competition begins and an initial plan is selected, the state in [2] obtains. Subsequently a reorganization operation applies iteratively: the selected node is suppressed and all other nodes which have not been selected are promoted to the next highest level of the potential.

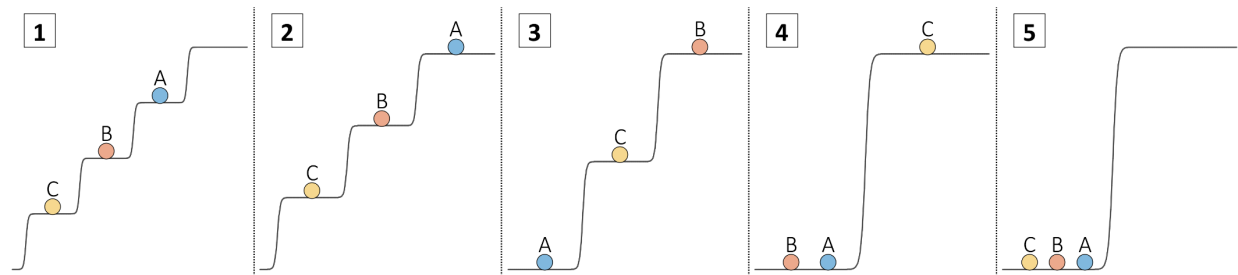


Fig. 13. Discrete sequence of organized relative activation states.

Another important feature of CQ is that feedback is responsible for suppressing the selected node and for gating/de-gating non-selected nodes, i.e. for inducing a reorganization of the quantal potential. Without feedback, a selected node would remain active indefinitely. Indeed, in a CQ model, the timecourse of feedback processes determines the duration of time during which a selected motor program exerts an influence on effectors. Recall that in the gestural scores of the AP/TD framework, the durations of gestural activation intervals are specified in an ad hoc manner, or as an arbitrary phase of the virtual cycle of a gesture. If we associate the nodes of the CQ framework with gestures in AP/TD, then the duration of gestural activation can be seen as a consequence of feedback-induced suppression.

3.3 External vs internal feedback

One source of feedback is from peripheral sensory organs, involving auditory, visual, tactile, and proprioceptive information. This feedback is called *external* because it derives from changes in the physical environment—the world outside of the central nervous system—caused by motor actions. For example, when our lips close the acoustic intensity of sound radiated from the vocal tract is greatly diminished, tactile nerves in the lips respond, and muscle stretch receptors in the oris obicularis muscles provide information regarding labial posture. All of this informs the nervous system regarding whether a labial closure target has been achieved.

The role of external feedback is schematized in Fig. 14. A motor controller transforms a sensory target to an outgoing motor command. The command effects a change in the external environment, which subsequently induces changes in the states of sensory systems. Comparison of the sensory system states with the target can be used to adjust outgoing motor commands. There is substantial evidence that this form of control occurs in speech (Hickok, Houde, & Rong, 2011; Houde & Jordan, 1998, 2002), and the GODIVA model explicitly incorporates feedback of this sort (Bohland, Bullock, & Guenther, 2010). In the CQ framework, comparison of external feedback with a target can also be used to suppress a selected node and degate other nodes.

However, external feedback is slow relative to the timescale on which movements are typically conducted and incurs sizeable temporal gaps between movements. A more timely mechanism for providing feedback information is therefore needed. A great deal of evidence indicates that the nervous system employs a predictive, anticipatory form of feedback, called *internal feedback* (Desmurget & Grafton, 2000; Hickok et al., 2011; Kawato, 1999; Kawato & Wolpert, 1998; Wolpert, Ghahramani, & Jordan, 1995). Fig. 14 schematizes internal feedback as the transformation of an efference copy of outgoing motor commands to the predicted/anticipatory sensory consequences of those commands. As with external feedback, internal feedback can be used to adjust outgoing motor commands. In a CQ framework, internal feedback can influence suppression of a selected node and gating of non-selected nodes.

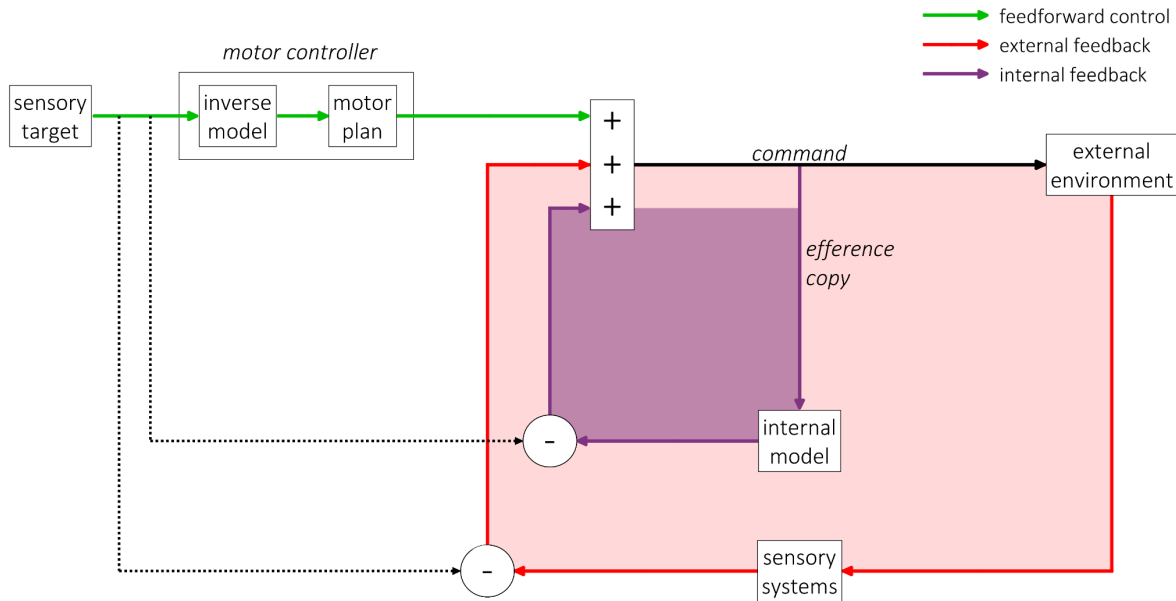


Fig. 14. Comparison of external and internal feedback in a generic motor control schema.

The utility of internal feedback derives from its potential use in adjusting the state of the motor system *before* a target is achieved in the physical environment. Consider the contrast between external and internal feedback for control of gestural selection in a CQ model. If suppression and degating are contingent solely on external feedback, as in Fig. 15A, then there is necessarily a gap in time between target achievement of a preceding gesture {1} and selection of a competitor gesture {2}. However, if internal feedback is used to de-gate {2} prior to target achievement of {1}, the gestural selection intervals can overlap, as in Fig. 15B. Pervasive overlap observed in speech indicates that anticipation/prediction of target achievement may be generally more influential than the peripheral sensation of achievement. It might also be expected that the internal regime of control would be associated with less variability in timing of selection than the external one, because external sensory information may be perturbed by contextual effects on movement targets or other environmental influences.

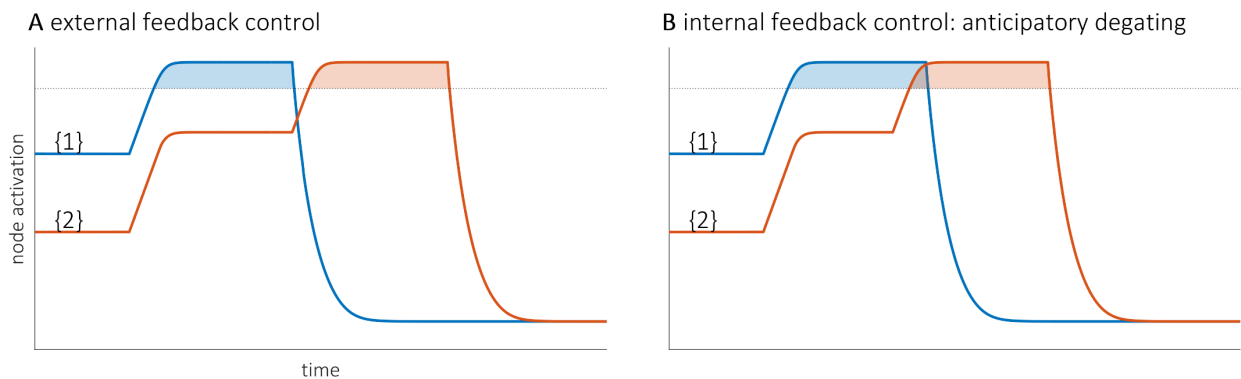


Fig. 15. Comparison of gestural overlap under external vs. internal feedback control. (A) External feedback entails a gap between intervals of gestural selection. (B) Overlap of selection intervals is possible with the use of internal feedback for anticipatory degating.

There are deep parallels between the external/internal feedback distinction and the extrinsic/intrinsic timing distinction (Fowler, 1980), and these suggest a basis for integrating CQ with AP/TD. Consider that

a CQ model relying solely on external feedback control is analogous to extrinsic control of timing: the relative timing of gestural activation (external feedback-induced suppression) derives from sensory systems that are *external/extrinsic* to the gestures themselves. In contrast, in a CQ model in which internal feedback control is used for anticipatory de-gating, the timing of gestural activation derives from a transformation of information associated with gestures themselves, i.e. the efference copy of motor commands. Internal feedback control can therefore be specified in a time-coordinate that is relative to the temporal evolution of the state of the motor system associated with the selected gesture, i.e. relative to its *phase*, an intrinsic time coordinate. Indeed, it has been pointed out that the original TD model in effect makes use of instantaneous, veridical internal feedback (Ramanarayanan, Parrell, Goldstein, Nagarajan, & Houde, 2016): specification of timing via phasing requires the control system to keep track of the phase states of active gestures, and it makes sense to attribute this phase-state tracking to an internal model. In light of these connections, one can view coordinative, feedforward control as control that relies solely on internal feedback.

The above considerations are relevant to the point that there are an important conceptual differences in how we can understand control of relative timing of differing sorts of gestural events. The precise control of timing of gestural initiation is understood in via the GPO phase-coupling mechanism in the AP/TD framework, but could alternatively be understood via anticipatory degating through internal feedback control. In contrast, the precise control of a gestural initiation relative to a preceding target cannot be understood with the GPO phase-coupling mechanism, because stabilization-triggering entails direct control over only the initiation of activation. The timing of target achievement is at best indirectly controlled to the extent that there is a stereotypical time from initiation to achievement. The CQ internal degating mechanism provides an alternative: the internal model representation of gestural phase, where target achievement is quasi-predictable, drives timing of subsequent gestural initiation. Neither model allows for direct control over the timing of a gestural target achievement to either a gestural initiation or another target achievement; this is because gestural targets cannot be achieved without first initiating a gesture. To directly control target achievement, the control system would have to “work backwards” from an estimation of initiation-to-target time, in order to determine when to initiate gestures.

The internal model, which maps outgoing motor commands to predicted sensory consequences, must be learned, and this entails that use of internal feedback for anticipatory degating is a learned behavior. A familiar (but not experimentally verified) example of this learning may be observed when a child learns to play a sequence of piano notes. Early on, the child depresses one key at time, releases it, then proceeds to the next one, apparently relying on external sensation of target achievement. However, this fairly slow performance is soon replaced by a more rapid one in which the movement to press a subsequent key occurs before the release of the preceding one—the child has proceeds to the next note *in anticipation of* having achieved the target of the preceding one.

An experimentally studied example involves reaching for and grasping an object, such as cup of water. Adults implement the grip formation (shape the hand in a manner appropriate for the object) *while* they transport the hand to the object. In contrast, very young children perform these tasks in a sequence: they first transport the hand, and only after receiving external sensory feedback for the target location of the hand do they initiate the grip formation (Jeannerod, 1986; von Hofsten, 1979). A similar developmental pattern applies to gripping and lifting (Forssberg, 1999; Forssberg et al., 1992). More generally, deficits in predictive motor control are hypothesized to underlie a wide variety of developmental movement disorders, and evidence along these lines has been observed for eye movements, reaching movements, grip force control, and dynamic postural control (Adams, 2018; Adams, Lust, Wilson, & Steenbergen, 2014; Wilson & Butson, 2007). If learning to use internal feedback also applies to sequencing articulatory gestures, which seems likely, then degree of reliance on internal feedback is expected to be a major source of variation in control, developmentally and perhaps in adult speech as well.

4. Intention

The AP/TD and CQ models are relatively impoverished when it comes to their conceptualization of movement targets. In TD for instance, a gestural target is a scalar value of a tract variable which specifies an equilibrium position. There are a number of phenomena which cannot be readily understood in this way—e.g. assimilatory and dissimilatory interactions between gestures which are not simultaneously selected, and pre-response anticipatory posturing of the vocal tract. This section discusses an elaborated model of *intention*, the planning of targets, in which target states vary over time and arise from integrating over spatially organized fields of neurons. This new conception of movement targets holds that AP/TD gestural targets are a form of long-term memory, while the time-varying target state of articulatory control is determined on-line in intentional planning from interactions of gestural targets. It should be noted that the term “intention” is used in this context because of its colloquial sense of an “aim” or “goal”.

4.1 Dynamic field model of movement preparation

The motivation for positing a spatial code for planning movement targets comes from studies of reaching and eye movement trajectories using a distractor-target paradigm. In this paradigm, a subject is presented with a distractor stimulus and subsequently a target stimulus, and then reaches or looks to the target. The distractor stimulus is understood to automatically induce planning of a reach/saccade to its location, and this planning subsequently influences the planning and execution of the reach/saccade to the target location. There are both assimilatory and dissimilatory phenomena observed in this paradigm which require an elaborated model of target planning.

When the distractor and target stimulus are sufficiently proximal in space, or are associated with similar movements, there is an assimilatory interaction in planning. Reaches and saccades to the target are observed to deviate toward the location of the distractor (Ghez et al., 1997; Van der Stigchel, Meeter, & Theeuwes, 2006; Van der Stigchel & Theeuwes, 2005). In speech, the analogous phenomenon of distractor-target assimilation was observed in Tilsen (2007, 2009). Formants in productions of the vowels [a] and [i] were assimilated toward distractor stimuli which were subcategorically shifted variants of [a] and [i], respectively.

Erlhagen & Schöner (2002) (cf. also Schöner, Kopecz, & Erlhagen, 1997) presented a dynamic field model capable of producing this assimilatory pattern (see also Roon & Gafos, 2016; Tilsen, 2009a). A simplified adaptation of a dynamic field model to the CQ/AP/TD framework is shown below, where tract variables are considered to be the relevant movement parameters. Each tract variable is associated with a distribution of activity in a field, which can be envisioned as a somatotopically organized population of neurons. As represented in Fig. 16A, a task-specific input to the field creates a Gaussian distribution of activation in the parameter field. Each input distribution can be associated with a gestural node, with the amplitude of the input proportional to the level of gesture node activation. The location and shape of the distribution of input from a gesture to the parameter field is a form of long-term memory of a gestural target. When a movement is initiated, the activation centroid of the field determines the value of the target parameter for the corresponding tract variable. If two gestures are simultaneously active and their activation distributions are sufficiently proximal, the activation centroid will be shifted in an assimilatory fashion, as shown in Fig. 16B. The dashed lines show the peaks in input from each of gesture, while the arrow indicates the target derived from intentional planning, i.e. the centroid of the field activation.

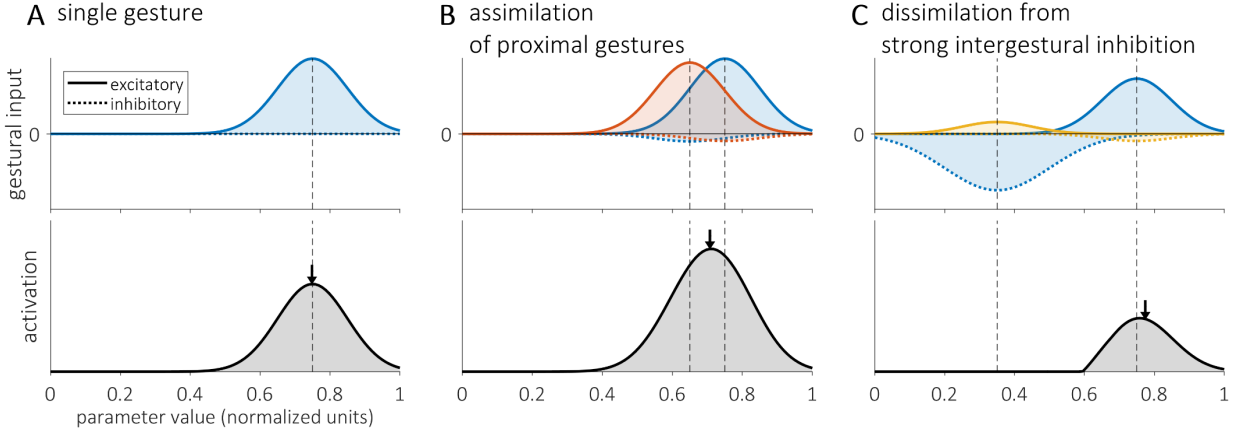


Fig. 16. Dynamic field models of intentional planning. (A) intentional planning of a single gesture. (B) assimilation of two gestures with proximal targets. (C) dissimilation when intergestural inhibition is strong.

In contrast to the assimilatory pattern, a dissimilatory pattern arises when the distractor and target are sufficiently distal in space. Eye movement trajectories and reaches are observed to deviate away from the location of the distractor in this case (Houghton & Tipper, 1994, 1996; Sheliga, Riggio, & Rizzolatti, 1994). In speech, the analogous effect was found in Tilsen (2007, 2009): productions of [a] were dissimilated from [i] distractors, and vice versa. A similar dissimilation was observed between Mandarin tone categories in a distractor-target paradigm (Tilsen, 2013b). The dissimilatory phenomenon has been explained by hypothesizing that inhibition of the subpopulation representing the distractor shifts the overall population distribution so that its centroid is further away from the target than it would otherwise be in the absence of the inhibited distractor (Houghton & Tipper, 1994). The effect can be modeled in the dynamic field framework through inhibitory input to parameter fields (cf. Tilsen, 2007). As shown in Fig. 16C, a gesture exerts inhibitory input centered on the target of a co-active gesture, and this inhibition shifts the centroid of the activation distribution away from the location of the inhibited target.

For a generic implementation of intentional planning, the dynamics of each parameter field $u(x,t)$ can modeled using a normalized coordinate x which ranges from 0 to 1 in small steps. Eq. 7 shows three terms that govern the evolution of the field. The first is an activation suppression term, with gain α , entailing that in the absence of input $u(x)$ relaxes to zero and that field activation saturates with strong excitatory input. The second term is excitatory input, where N is a Gaussian function of x with mean μ_i and standard deviation σ_i^{exc} associated with gesture i . The term g_i^{exc} is a sigmoidal function of gestural node activation and modulates the amplitude of the Gaussian. The gain term β_{exc} controls the overall strength of the excitatory input. The third term is inhibitory input, which is also a sum of Gaussians modulated by a sigmoidal function of gestural node activation, g_j^{inh} . This term differs from the excitatory one: each inhibitory input is centered on the gesture that is the target of inhibition, but the amplitude of the inhibition is a modulated by the activation of its source. The matrix I_{ij} specifies pairwise intergestural inhibition strengths. Note that excitatory and inhibitory inputs may differ in their spread (σ_i^{exc} vs. σ_i^{inh}), and the condition $u(x,t) \geq 0$ is imposed at each timestep. Eq. 8 shows the calculation of the target as the average activation-weighted parameter value, i.e. the field activation centroid.

$$\text{Eq. 7} \quad \frac{du(x)}{dt} = -\alpha u(x) + \beta_{exc} \sum_i g_i^{exc}(t) \mathcal{N}(x, \mu_i, \sigma_i^{exc}) - \beta_{inh} \sum_i \sum_j I_{ij} g_j^{inh}(t) \mathcal{N}(x, \mu_i, \sigma_i^{inh})$$

$$\text{Eq. 8} \quad T(t) = \frac{\sum_x x u(x,t)}{\sum_x u(x,t)}$$

Assimilatory and dissimilatory effects in distractor-target paradigms suggest a reinterpretation of the target parameters in the TD model. The empirical patterns can be understood if the calculation of targets is determined on-line in production, rather than pre-determined by a fixed gestural target. The target parameters of gestures are thus reinterpreted as a form of long-term memory which specifies distributions of input activity to a parameter field, while the target parameters in tract variable equations—i.e. equilibria of driving forces on tract variables—arise from integrating the excitatory and inhibitory input to the parameter fields.

4.2 Subthreshold intentional effects

When does intentional planning result in observable changes in the state of the vocal tract? One possibility is that tract variable driving forces are present only when some gesture is active, since gestural activation is associated with movement initiation in the AP/TD framework. However, studies of discrepancies between acoustic and articulatory measurements of verbal reaction time in delayed response paradigms have provided indirect evidence for changes in vocal tract state prior to the initiation of movement (Kawamoto, Liu, Mura, & Sanchez, 2008; Rastle & Davis, 2002). Indeed, substantial response-specific anticipatory posturing, up to several seconds prior to response initiation, was observed in a real-time MRI study designed specifically to test for such effects (Tilsen et al., 2016). This study showed that prior to the cued initiation of a response, speakers often adopt a vocal tract posture that partly assimilates to upcoming gestural targets. These findings suggest that gestures need not be selected in order for movement to occur.

The anticipatory posturing effect can be modeled in a CQ/AP/TD framework with several revisions. First, sub-selection-threshold activation of a gestural node must provide input to the corresponding tract variable parameter field. Second, the activation centroid of each tract variable parameter field must continuously determine the current equilibrium value of that tract variable. Third, the stiffness parameter in the tract variable equations must depend on the amount of activation in the field. In effect, these alterations lead to reinterpreting the gestural activation intervals of AP/TD as periods of time during which the gesture-associated nodes of CQ obtain above-selection-threshold activation.

Fig. 17 exemplifies how these changes give rise to anticipatory posturing in a single vocal tract variable. Throughout the response, the parameter field receives constant input which is associated with a speech-ready posture/neutral attractor. Thus before any gesture becomes active to any degree, the parameter field maintains an intentional target consistent with the speech-ready posture (Fig. 17A). When a speech plan is retrieved or cued by a stimulus, some gestures obtain a subthreshold level of activation. They provide input to the parameter field and thereby shift the centroid of the field (Fig. 17B), inducing movements which result in a potentially subtle change in vocal tract posture. Subsequently, when the response is “initiated”, the gesture obtains suprathreshold activation (“is selected”) and exerts a strong excitatory input on the parameter field (Fig. 17C), resulting in more extensive and rapid changes in vocal tract state.

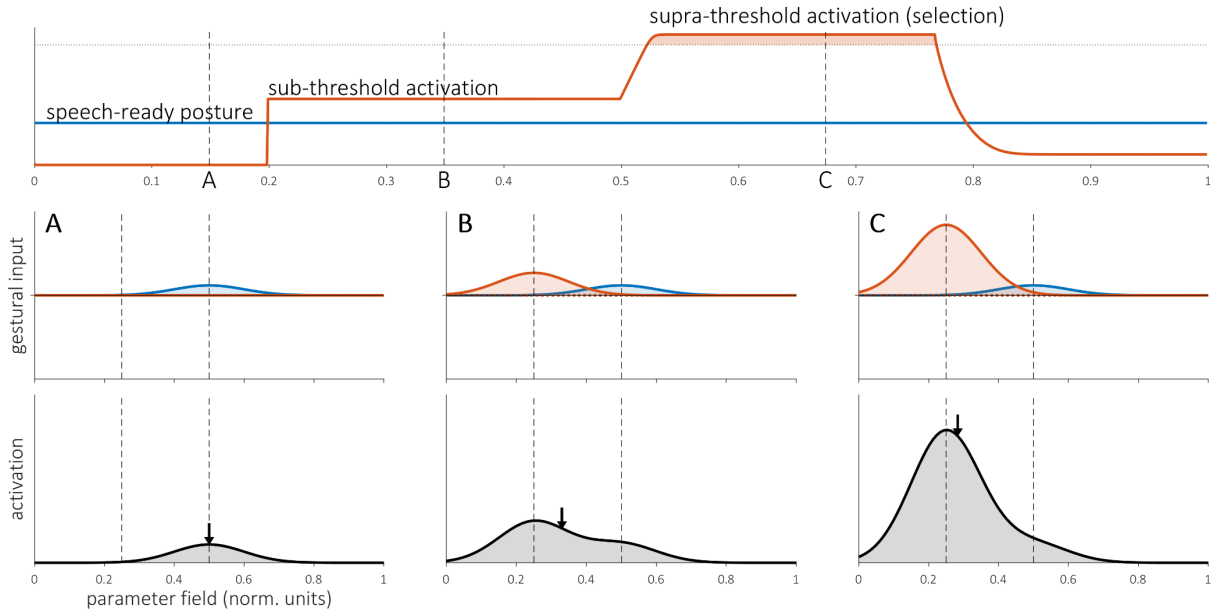


Fig. 17. Influence of subthreshold gestural activation on intentional planning. (A) Parameter field activation prior to response retrieval. (B) Pre-response field activation drives subtle changes in vocal tract state. (C) Field activation during gestural selection induces strong forces on a tract variable.

Subthreshold effects of gestural activation on intentional planning may be subtle and easily obscured by above-threshold (i.e. selected) gestures in typical speech. It is also unclear when subthreshold effects on target tract variable target states are expected to be predominantly excitatory or inhibitory, i.e. to result in assimilation or dissimilation. Dissimilatory effects are a proposed general mechanism of contrast preservation and maximization (Tilsen, 2013b), but as we consider in section 6, the balance of subthreshold excitatory and inhibitory input to parameter fields may give rise to long-distance harmonies.

The above considerations suggest that a disambiguation of the term *initiation* is in order. Gestures are “initiated” in two senses: gestures conceptualized as CQ nodes become active, to some subthreshold degree, and this “initiation” may or may not result in observable movement. Subsequently, gestural nodes are selected, i.e. their activation exceeds a threshold, and this results in a nonlinear increase in the stiffness of the vocal tract variable. Thus selection leads to rapid, observable movements. However, this revised conception raises new questions regarding what role coordination plays in controlling movement timing. In particular, how does the stabilization-triggering hypothesis of the coupled oscillators model interact with the notion of a continuously evolving intentional planning forces on tract variables?

5. Interaction of selection, coordination, and intention

In the preceding sections, we motivated the need for three mechanisms in modeling articulatory control. Gestural coordination is a mechanism for precise control of the timing of movement initiation, operating via a phase code. Selection is a mechanism for the choosing which gestures to execute and temporally ordering those choices via an activation code and sensory feedback. Intention is a mechanism for determining the forces which govern the state of the vocal tract, operating via a spatial code. Below we construct a picture of how these mechanisms fit together, and elaborate specific proposals regarding developmental changes and spontaneous variation in the interaction of the mechanisms.

5.1 An integrated framework

As hypothesized in Selection-coordination theory, the interaction of selection and coordination mechanisms is such that only gestures which are selected together (*co-selected*) can be coordinated (Tilsen, 2016). More specifically, phase coupling forces between GPOs are assumed to be negligible for gestures which are not co-selected. In other words, the selection mechanism organizes gestures into sets, and the precise timing of the movements associated with the gestures in each set is governed via phase coupling of GPOs. There are a number of ways to implement this in a computational model. For example, in Tilsen (2013a) the coordinative control of timing was implemented via a GPO/triggering mechanism which was constrained to operate subsequent to suprathreshold activation of gestural nodes. Here an alternative is described in which GPO oscillations of selected gestures, upon reaching a particular phase, provide a nonlinear boost to the stiffness of the associated tract variables. One motivation for pursuing this alternative is that in the integrated model, gestural node activation provides continuous input to intentional planning fields. Hence a nonlinearity in the influence of planning fields on tract variables is important to effect the rapid changes in driving forces on tract variables which are associated with movement initiation.

Simulations of co-selected CCV and VC syllables are shown in Fig. 18 below. There are several features to take note of. First, the beginning of the interval of time during which a gesture is selected (i.e. the *selection interval*) is not equivalent to the gestural activation interval, which is defined as the period of time during which the gesture exerts a stiffness boost on the corresponding tract variable. This dissociation occurs because a selected gesture augments tract variable stiffness only after its associated GPO reaches an arbitrarily chosen *boost phase* (here the oscillation peak). Thus the GPO triggering mechanism is reinterpreted as an augmentation of tract variable stiffness.

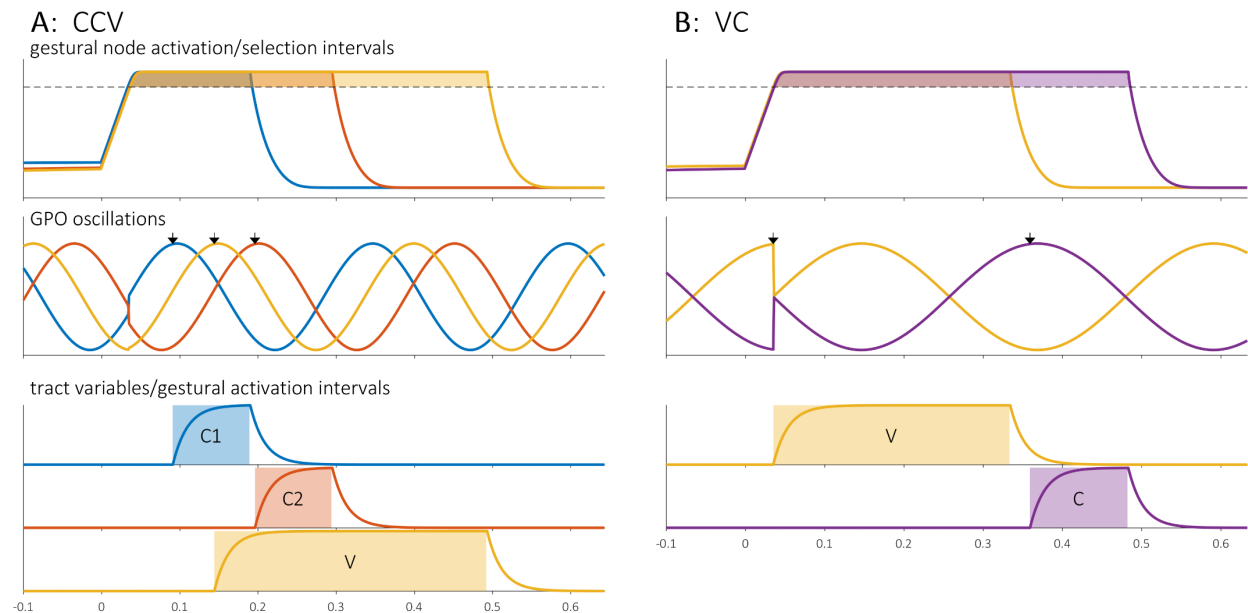


Fig. 18. Simulations of CCV and VC syllables with a model that integrates selection, coordination, and intention mechanisms.

Second, the time coordinates in Fig. 18 are expressed relative to response initiation, which can be construed as an external stimulus (i.e. go-signal) or an internally generated intent to initiate the response. The delay between response initiation and movement initiation is therefore comprised of two time

periods: the period during which gestural node activation, initially below selection threshold, grows to selection-level, and the period between the beginning of the selection interval and the subsequent time for GPOs to reach the boost phase. This latter period depends on the GPO frequencies and phases at the moment of gestural selection. Hence the model design is such that stabilization of GPOs is not required for movement initiation. Whether or not stabilization is required, there is a possibility for a misordering of gestural activation if gestural node selection occurs just after some but not all of the GPOs have reached the boost phase. Some mechanism is therefore required to avoid this circumstance. The solution employed here is to reset the phases of co-selected GPOs upon selection, preserving their relative phase pattern. This reset can be observed in Fig. 18B, where GPO oscillations exhibit a discontinuity when gestural activation exceeds a threshold.

Recall from section 2.3 that uniform frequency of GPOs creates a problem for modeling coordinative control over a CCVC form. The integrated framework offers an alternative conception of control over post-vocalic consonantal timing which avoids this problem. Rather than thinking of coda C movement initiation as coordinatively controlled via GPO phase-coupling, the selection-coordination framework offers the option of modeling coda C movement initiation as a consequence of anticipatory de-gating, which is mediated by internal feedback control. This timing pattern can be understood in comparison to a prototypically competitively selected post-vocalic C, which is shown in Fig. 19A. In the prototypical scenario, competitive gating prevents the {C} node from achieving selection-level activation while the {V} node is selected. Hence neither the selection intervals nor the gestural activation intervals overlap. In the anticipatory degating scenario (Fig. 19B), internal feedback is used to degate the {C} node while {V} is selected, and so the selection and activation intervals overlap.

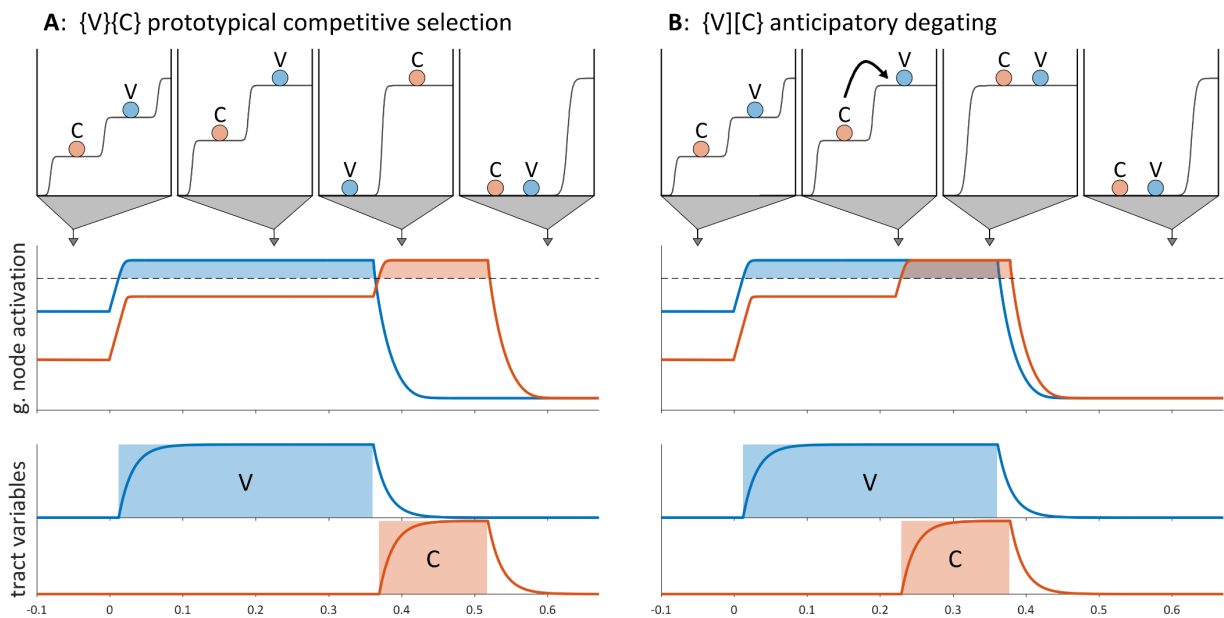


Fig. 19. Comparison of prototypical competitive selection and anticipatory degating. (A) The post-vocalic {C} node is selected after suppression of {V}. (B) Post-vocalic {C} is anticipatorily degated prior to suppression of {V}.

Thus the integrated framework offers two ways to model control of post-vocalic timing. One is via GPO anti-phase coupling (or possibly arbitrary phase-coupling)—i.e. *coordinative control*—and the other via internal feedback-driven anticipatory degating of the post-vocalic gesture—i.e. *competitive control*. Are these models of post-vocalic timing merely different conceptual models of the same phenomenon, or do they describe substantively different forms of organization? On one hand, the models may merely be

alternative descriptions: the timing of when internal-feedback induces anticipatory de-gating can be reinterpreted as an arbitrary phasing associated with GPOs, and vice versa. Because both models can generate empirically consistent timing patterns for post-vocalic consonants, it may be hard to differentiate them with behavioral studies. On the other hand, the models may describe the action of distinct neural mechanisms (Ziegler & Ackermann, 2017), and as considered below, the distinction is useful for the analysis of developmental patterns.

A third feature of the integrated model to emphasize is that the duration of a gestural activation interval is determined by feedback processes. In normal adult speech, internal feedback is primarily responsible for the deactivation of a gesture, which corresponds to the cessation of the stiffness boost associated with a triggered gesture. This entails that many durational phenomena should be understood as the consequence of feedback-induced gestural suppression. In the integrated model above, the suppression mechanism is implemented by incorporating gesture-specific feedback systems which accumulate activation when the state of vocal tract variables match the target associated with a gesture. The accumulation is weighted by the similarity of the target and current state. Each gesture is also associated with a threshold parameter that specifies how much activation is accumulated before the gesture is suppressed.

5.2 Developmental transition from competitive to coordinative control

The interaction of selection and coordination gives rise to two *prototypical* regimes of control: competitive control and coordinative control. Note that competitive and coordinative control regimes are conceptualized as *prototypes*, and we can view them as endpoints of a continuum, as shown for a vocalic gesture (V) and post-vocalic gesture (C) in Fig. 20. In the prototypical form of competitive control, the selection of the C is governed solely by external sensory feedback: the gesture is only degated after external sensory feedback regarding the achievement of the V target results in the suppression of V. However, internal sensory feedback can be used to anticipate V target achievement and de-gate C prior to V suppression. We can therefore imagine a continuum of internalization, i.e. reliance on internal sensory feedback, and an associated continuum of anticipation of the C gesture.

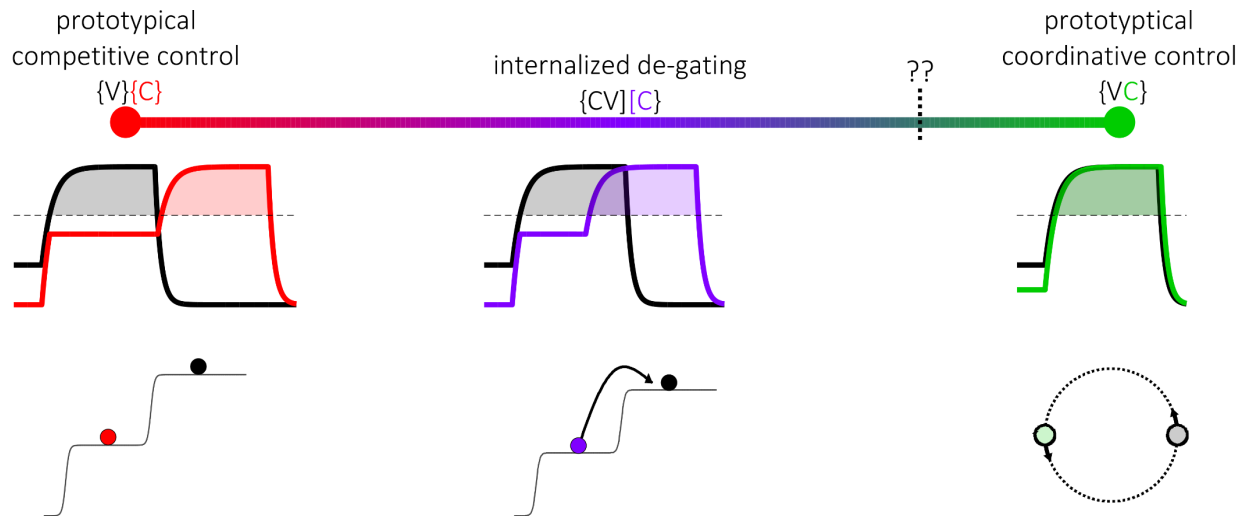


Fig. 20. Continuum of internalization between prototypical endpoints of competitive control and coordinative control.

The *developmental internalization hypothesis* holds that in first-language acquisition, primarily in the early word stage (typically from 1-3 y.o.), children increasingly rely on internal sensory feedback for control, and hence transition from the prototypical competitive regime toward a more prototypically coordinative regime. This hypothesis predicts different forms of organization for simple onset and coda consonants in this stage.

Control over a prevocalic consonantal gesture, even in the early word stage, is predominantly coordinative. In babbling children learn to co-select and coordinate gestures which are organized as {CV} sets. One argument for viewing babbles as coordinatively controlled is that babbling co-emerges with other oscillatory behaviors (Iverson & Thelen, 1999; Thelen, 1979), although there may be reasons not to identify babbling with general oscillatory behaviors (Jürgens, 2009). The oscillatory view of babbling also relates to the frame-content hypothesis (B. L. Davis & MacNeilage, 1995; MacNeilage & Davis, 2000; MacNeilage, Davis, & Matyear, 1997), which interprets babbling as a process of developing control over labial and lingual constriction gestures that occur in a frame of jaw oscillation.

In contrast to {CV} organization, post-vocalic constriction gestures are competitively controlled in the early word stage. When children attempt to produce adult-like word forms with post-vocalic consonants, they competitively select the post-vocalic consonant, i.e. {V}{C}. Moreover, over the course of development children internalize control, i.e. use {V}[C] organization, and hence transition along the continuum from prototypical competitive to coordinative control. The same sort of pattern applies to onset consonantal gestures which are not immediately vowel-adjacent (henceforth *marginal onset gestures*): organization of CCV word forms is initially {C}{CV}, subsequently becomes internalized as {C}[CV], and then potentially becomes coordinative {CCV}. Sections 6.1 and 6.2 review various forms of evidence for these hypotheses.

One important question that can be addressed in the context of the current framework is why the particular developmental asymmetry between pre-vocalic and post-vocalic organization has the form that it does. Specifically, why is organization in the beginning of the early word stage predominantly {CV}{C}, as opposed to {C}{VC}? The explanation involves differences in consonantal vs. vocalic sensory feedback in combination with the fact that movement initiation necessarily precedes target achievement. Consonantal gestures, particularly those which are acquired in babbling, involve constrictions, closures, and/or valvings (in the case of nasals). The target vocal tract states of consonantal gestures may be generally associated with relatively robust sensory feedback compared to vowels, and perhaps relatively more somatosensory feedback. Because of differences in the robustness and primary domain of sensory

feedback, vocalic gestures must remain active for longer periods of time than consonantal constriction gestures. Furthermore, C gestures are generally associated with more rapid movements than those which achieve vocalic targets. A consequence of these circumstances is that it is possible to initiate a C and V gesture simultaneously and, assuming the C gesture is suppressed relatively quickly, the synchronous initiation will not compromise the achievement of the target of the vocalic gesture. The same does not hold if the V gesture is initiated first and the two gestures are initiated asynchronously; when the C gesture is initiated midway through the V gesture activation, it will interfere with the achievement of the vocalic target. Hence the temporal circumstances of movement initiation and target achievement, in conjunction with differences in sensory feedback, ultimately manifest in an onset/coda asymmetry.

Internalization is not the only change in control which must occur in early speech motor development. Another important change that must occur is dissociation of {C} gestures from {CV} sets. If we assume that predominantly {V} and {CV} organizations have been learned in babbling, then children must learn to independently organize individuated {C} gestures, in order to produce organizations such as {V}{C} and {C}{CV}. Furthermore, it seems evident that for various reasons, some constriction gestures are less easily acquired than others, and this influences the propensity for these gestures to be dissociated from {CV} organizations and/or controlled through internal feedback. For example, sibilants are generally acquired later than stops, and this is plausibly due to greater difficulty in (i) learning the subgestural coordination of tongue and jaw posture necessary to achieve a narrow constriction required for generating turbulence, and/or (ii) learning a sensory-motor intentional mapping which is consistent with the most salient acoustic properties of adult sibilants (Reidy, Beckman, Litovsky, & Edwards, 2015; Smit, 1993). For another example, the English approximant /r/ may be difficult to acquire because it involves coordinating three oral gestures: labial protrusion, tongue tip elevation or bunching, and tongue root retraction (Gick, Bernhardt, Bacsfalvi, & Wilson, 2008; Smit, 1993; van Lieshout, Merrick, & Goldstein, 2008).

The above considerations lead to the conclusion that the internalization hypothesis should not be viewed as a monolithic prescription regarding developmental changes in the organization of consonantal and vocalic gestures. Instead, the organization of control should be viewed as gesture-specific and subject to multifactorial contingency: anatomical details, biomechanical interactions, consequences of overlap for perceptual recoverability, language-specific frequency, functional load, morphosyntactic function, and presumably other factors can influence the timecourse and degree of internalization.

As we examine in section 6, nonuniformity in internalization is a powerful source of explanation for cross-linguistic differences and other forms of variation. Before we examine these applications of the framework, it is helpful to emphasize a couple generic predictions. The first relates to the durational consequences of internalization. All things being equal, two gestures which are subject to prototypical competitive control will obtain a supra-threshold level of activation for a longer period of time than gestures which are subject to some intermediate degree of internalized feedback control. In other words, internalization increases gestural overlap and thereby decreases overall duration. A second generic prediction is that selectional organization, in the prototypical scenario, is transitive. If gesture A is co-selected with gestures in set X, and gesture B is competitively selected with A, then gesture B cannot be co-selected with gestures in set X. Instead, B must comprise or be a member of an alternative set, Y.

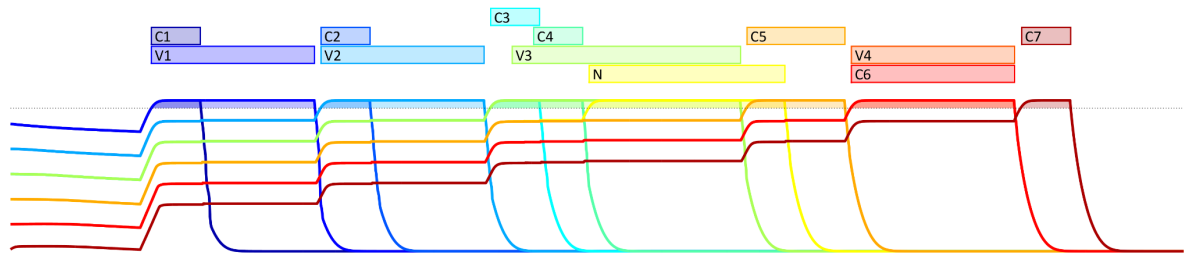
5.3 Spontaneous variation in organization

Whereas the internalization hypothesis posits a developmental trend in organization, another useful hypothesis relates to intraspeaker variation in adult speech, derived from variation in attention to external feedback. The *attentional modulation hypothesis* holds that reliance on internal feedback is modulated by the degree to which speakers attend to the sensory consequences of their speech. The hypothesis posits that competitive control dominates organization in emphatic and slow or hyperarticulated speech;

conversely in non-emphatic and rapid or hypoarticulated speech, coordinative control and anticipatory degating dominate. In other words, use of competitive control is a strategy for emphasis and hyperarticulation, while use of coordinative control is a strategy for rapid speaking and hypoarticulation.

The attentional modulation hypothesis implies that although the gestures in a given word or sequence of words can be organized in a wide variety of ways, there is a global modulation which biases organization to be more or less coordinative/internalized. For example, two different patterns of organization of oral and velar articulatory gestures in the phrase *coffee-drinkers* are compared in Fig. 21 below. These might be transcribed phonetically as [k^h:a:f:i::dɪ:ŋkə:z] and [kafitʃɪŋɪz] respectively, but the transcriptions belie the more fundamental differences in organization. To wit, in the more prototypically competitive example (Fig. 21A), gestures are longer in duration reflecting greater reliance on external sensory feedback for suppression, and gestural overlap is minimal reflecting degating which is minimally anticipatory. In the more prototypically coordinative example (Fig. 21B), gestures are shorter because they are suppressed in anticipation of target achievement and overlap is more substantial because of extensive anticipatory degating. The relative timing of initiation of onset consonantal gestures in the word *drinkers* is also compressed via an increase in the frequencies of GPOs.

A high degree of attention to sensory feedback: slow/hyperarticulated speech



B low degree of attention to sensory feedback: fast/hypoarticulated speech

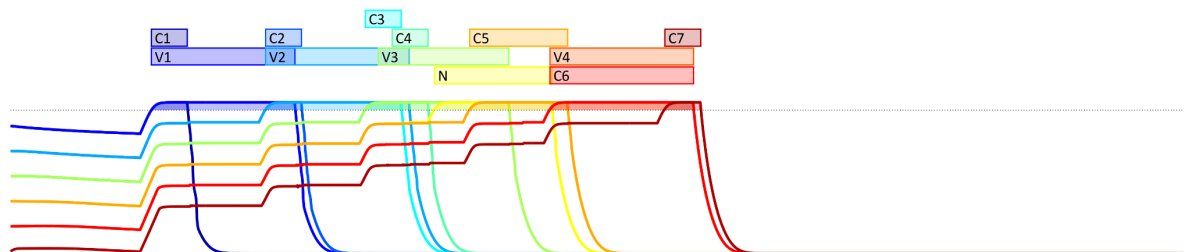


Fig. 21. Comparison of gestural timing in the phrase *coffee-drinkers*. (A) Production with a high degree of attention to sensory feedback. (B) Production with a low degree of attention to sensory feedback.

The attentional modulation of organization can also be seen as an alternative to the π -gesture analysis of prosodically induced durational phenomena. Byrd & Saltzman (2003) introduced π -gestures as a mechanism for modeling boundary-adjacent lengthening and kinematic augmentation (stiffness reduction or movement range increase), which are observed pervasively in speech (Byrd, Kaun, Narayanan, & Saltzman, 2000; Byrd & Saltzman, 1998; De Jong, Beckman, & Edwards, 1993). The π -gesture in effect works by slowing the evolution of a hypothetical “clock” time in which tract variables evolve, relative to the standard real-time evolution of the system.

In the current framework, durational lengthening of gestural activation intervals can be generated from increased thresholds for the accumulation of sensory feedback which induce degating, i.e. as the result of greater attention to sensory feedback. Kinematic augmentation on the other hand must be generated through some other mechanism such as a reduction in the stiffness boost associated with

gestural triggering and/or extremization of vocal tract targets in intentional planning. One of the advantages of this reconceptualization of prosodic effects on articulation is that the same feedback-related selection mechanisms which are held responsible for emphasis and hyperarticulation in adult speech—i.e. for greater duration and slower movement velocities—can be identified with mechanisms that produce those patterns in early development.

6. Applications

In this section a variety of applications of selection, coordination, and intention mechanisms are described. The aim here is to show the utility of the mechanisms by virtue of their explanatory breadth. To that end, the various phenomena are described in a relatively superficial manner. In many cases, the important point is that developmental and phonological patterns which have conventionally been analyzed via structural relations can be studied in light of their mechanistic origins.

6.1 Phonetic evidence for developmental changes in organization

Developmental asymmetry of CV and VC coarticulation

Developmental differences in CV vs. VC coarticulation can be understood as a consequence of $\{CV\}\{C\}$ organization in the early word stage. Findings from a number of studies show that formant transitions develop steeper, more adult-like slopes earlier in CV than VC environments (Goodell & Studdert-Kennedy, 1993; Hawkins, 1984; Katz & Bharadwaj, 2001; Kent, 1983; Repp, 1986; Sussman, Duder, Dalston, & Cacciatore, 1999; Turnbaugh, Hoffman, Daniloff, & Absher, 1985). This asymmetry is consistent with the hypothesized $\{V\}\{C\}$ organization in early words. As shown in Fig. 22A, $\{V\}\{C\}$ organization generates no gestural overlap; as feedback control is internalized, the degree of overlap increases.

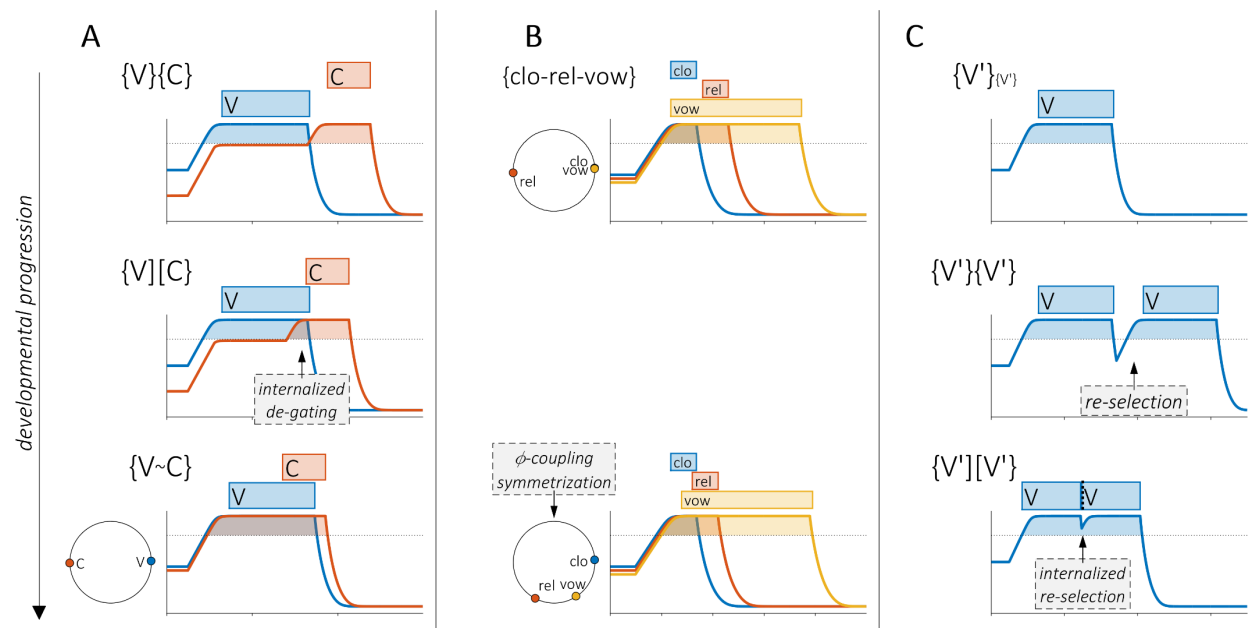


Fig. 22. Phonetic developmental patterns. (A) Internalization of control from $\{V\}\{C\}$ to $\{V\}[C]$ and $\{V\}\sim C$ generates degrees of gestural overlap. (B) Hyper-coarticulation of CV is associated with an imbalance in the strength of phase-coupling forces between vocalic and closure/release gestures. (C) Developmental

changes in long vowel duration arise from use of reselection to achieve length and subsequent internalization of reselection.

Hyper-coarticulation of CV

Hyper-coarticulation of {CV} in early development can be understood as a consequence of unbalanced and/or asymmetric phase coupling forces, i.e. immature coordinative control. A number of studies have found that onset consonants are hyper-coarticulated with vowels in early development (Goodell & Studdert-Kennedy, 1993; Nittrouer, Studdert-Kennedy, & McGowan, 1989; Nittrouer, Studdert-Kennedy, & Neely, 1996; Noiray, Abakarova, Rubertus, Krüger, & Tiede, 2017). This phenomenon can be understood with the coordination mechanism: symmetric displacement of closure and release relative to the vocalic gesture obtains when the strengths of the in-phase coupling forces between V and closure and release GPOs are approximately the same. The hyper-coarticulation pattern arises from an asymmetry such that closure is more strongly coupled to V than release. As shown in Fig. 22B, the refinement of coordinative control involves symmetrization of coupling forces.

Inverted-U profile of long segment duration

Developmental changes in the duration of phonologically long segments such as diphthongs, long vowels, and geminates exhibit an inverted U profile which can be understood with the progression {X} → {X'}{X'} → {X'}[X']. Several studies have observed an interesting developmental pattern: initially durations of long segments are variable but more representative of short segments; children subsequently tend to progress through a stage in which long segments are produced with atypically long durations, and then settle in on a more adult-like intermediate (cf. Kehoe & Lleó, 2003; Naeser, 1970; Raphael, Dorman, & Geffner, 1980; Kehoe & Lleó, 2003; Stokes & Wong, 2002; Hurme & Sonninen, 1985; Kunnari, Nakai, & Vihman, 2001; Vihman & Velleman, 2000). As illustrated for a V gesture in Fig. 22C, this inverted-U durational pattern arises from three stages of organization. In the first, the V gesture is selected just once: no length contrast is achieved; next, children learn to intentionally reselect the gesture, i.e. {V'}{V'}; subsequently, they learn to anticipatorily degate the reselection, i.e. {V'}[V'], thereby shortening overall duration.

6.2 Developmentally transient nonstandard sequencing patterns

Post-vocalic consonants

Anomalous sequencing patterns that are associated with post-vocalic consonants in the early word stage can be understood with {CV}{C} organization. The review of studies conducted in (Tilsen, 2016) found that compensatory vowel lengthening, reduplication, and vowel epenthesis are associated with post-vocalic consonants but not pre-vocalic ones (compensatory lengthening: (Bernhardt & Stemberger, 1998; Ota, 2001; Song & Demuth, 2008); reduplication: (Fee & Ingram, 1982; Ferguson, 1983; Schwartz, Leonard, Wilcox, & Folger, 1980); vowel epenthesis: (Demuth, Culbertson, & Alter, 2006; Matthei, 1989). Fig. 23A illustrates how {CV}{C} organization accounts for this asymmetry. Compensatory vowel lengthening is understood as the reselection of {V} at the expense of {C}. Reduplication is reselection of the preceding C and V gestures, i.e. {C'V'}{C'V'}, at the expense of the target {C}. Vowel epenthesis is selection of a {CV} set that includes the target {C}; this is expected on the basis of the assumption that {C} gestures are weakly dissociated from {CV} sets in early development.

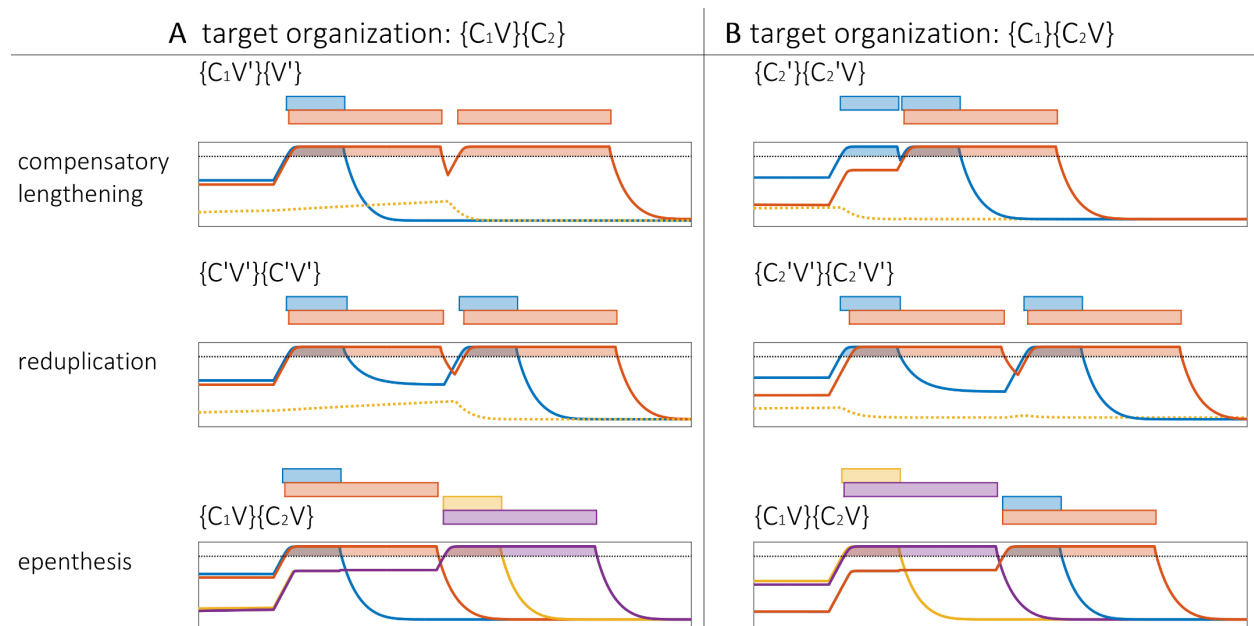


Fig. 23. Nonstandard sequencing patterns in development. (A) Errors associated with competitive organization of post-vocalic consonant. (B) Errors associated with competitive organization of marginal onset consonant.

The fact that such patterns are observed in association with post-vocalic consonants, but not immediately prevocalic consonants, indicates that children have implicit knowledge of a target organization with two distinct sets of gestures. In contrast, when an immediately prevocalic C gesture is omitted, it does not affect the number of sets that are selected, and hence there is no basis for reselection of any gestures. In all cases, the number of gesture-sets which are selected in the anomalous production is the same as the number of sets in the hypothesized target organization.

Word-initial clusters

The error patterns observed in association with marginal consonantal gestures (i.e. gestures which are not immediately prevocalic) are similar to those observed post-vocalically, and can be understood as a manifestation of $\{C_1\}\{C_2V\}$ organization. Several studies reviewed in (Tilsen, 2016) reported lengthening of C_2 , reduplication of C_2V , or epenthesis between consonants. Fig. 23B shows how these patterns arise from competitive organization. Lengthening of the vowel-adjacent consonant with the omission of the marginal one arises from selection of C_2 instead of C_1 and subsequent reselection of C_2 . Reduplication is the selection of $\{C_2V\}$ instead of C_1 , with subsequent reselection of $\{C_2V\}$. Epenthesis is selection of an alternative set, $\{C_1V\}$, perhaps due to insufficient differentiation of $\{C_1\}$ from $\{C_1V\}$. These patterns are directly analogous to those associated with post-vocalic consonants, and likewise follow from the assumption that children have implicit knowledge regarding the number of sets of gestures to select.

6.3 Cross-linguistic differences in organization

The internalization hypothesis predicts that gestures which are competitively organized in the early word stage may, through increasing reliance on internal feedback, become co-selected and coordinately controlled. However, this progression is not obligatory—it is contingent on a variety of factors—and the so emergence/non-emergence of a coordinative regime can account for various cross-linguistic phonetic and phonological differences.

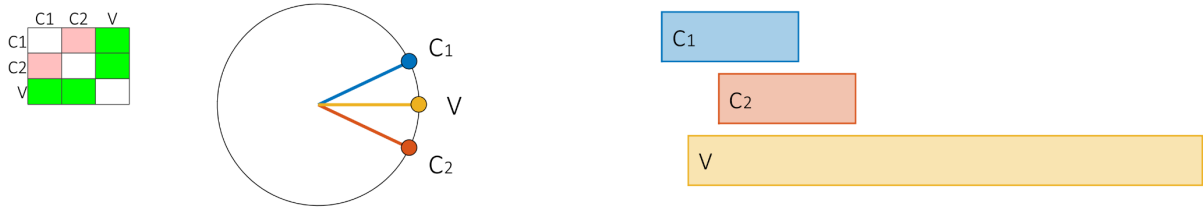
Simplex vs. complex onset typology

The typological distinction between simplex and complex onsets can be understood as a consequence of developmental contingency of {CCV} vs. {C}{CV} organization. In the *complex onset* case, phonological patterns treat all onset consonants and the vowel as part of the same unit, and the C-center effect is observed (cf. English: (Browman & Goldstein, 1988; Byrd, 1995; Honorof & Browman, 1995; Marin & Pouplier, 2010; Sproat & Fujimura, 1993); French: (Kuhnert, Hoole, & Mooshammer, 2006); Italian: (Hermes, Grice, Mücke, & Niemann, 2008; Hermes, Mücke, & Grice, 2013); Georgian: (Goldstein, Chitoran, & Selkirk, 2007); Serbian: (Tilsen et al., 2012)). In the *simplex onset* case, which has been observed in Tashlyhiyt Berber, Moroccan Arabic, Slovak, and Hebrew, cf. (Goldstein, Chitoran, et al., 2007; Pouplier & Beňuš, 2011; Shaw, Gafos, Hoole, & Zeroual, 2009, 2011; Tilsen et al., 2012), no C-center effect is observed and phonological patterns often indicate extraprosodicity of the marginal consonant (S. Davis, 1990; Dell & Elmedlaoui, 1985; Kaye, 1992). Note that in Italian, sC clusters alone exhibit the simplex onset pattern (Hermes et al., 2008), which shows that organization can be gesture-specific rather than homogenous within a language. The differences between simplex and complex onset gestural timing follow from a difference between {C}{CV} and {CCV} organization: the former is associated with conservation of competitive control; the latter employs phase coupling of onset consonant GPOs and hence is expected to exhibit a c-center effect.

Perceptual recoverability and overlap in word-initial consonant clusters

Relations between perceptual recoverability and gestural overlap in complex onsets can be understood as the result of a developmental interaction with acoustic sensory feedback and internalization. In word-initial clusters in Georgian, clusters with a front-to-back order exhibit more overlap than clusters with back-to-front order, which is presumably a consequence of more anterior closures masking the acoustic release of more posterior ones (Chitoran & Goldstein, 2006; Chitoran, Goldstein, & Byrd, 2002). Such effects suggest that perceptual recoverability has gradient influences on parameters of coupling strength that arise through internalization. The relevant parameter here is $[\phi^-/\phi^+]$, the strength of the anti-phase coupling force relative to the strengths of the in-phase coupling forces. When overlap of word-initial consonantal gestures has less severe consequences for perceptual recoverability, $[\phi^-/\phi^+]$ is relatively low and this results in a greater degree of overlap, as shown in Fig. 24A. In contrast, when overlap threatens C_1 release recoverability, $[\phi^-/\phi^+]$ is relatively high, and this results in a less overlap, as in Fig. 24B.

A C₁C₂V more overlap



B C₁C₂V less overlap

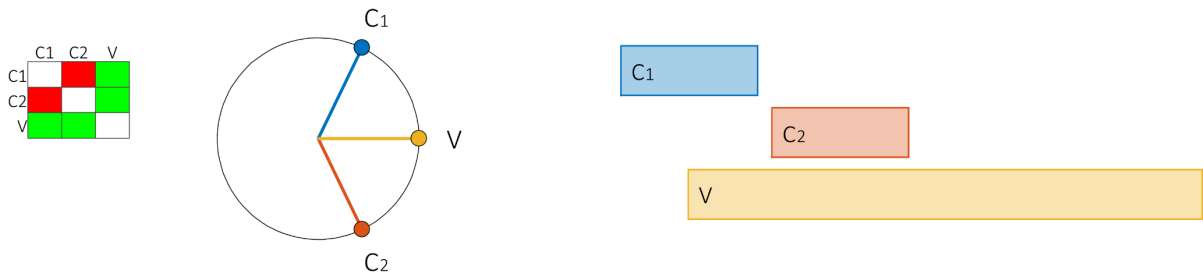


Fig. 24. Differences in gestural overlap interpreted as strength of anti-phase coupling.

The internalization hypothesis holds that the coordinative regime {C₁C₂V} is preceded by a competitive one: i.e. {C₁}{C₂V}, and then by an internalized one: {C₁}[C₂V]. Consequently, the coordinative organization that emerges can be inferred to exhibit strong anti-phase coupling, as in Fig. 24B. The extent to which [φ⁻/φ⁺] is subsequently weakened is expected to depend on perceptual recoverability.

Durational correlates of coda moraicity

The durational correlates of coda moraicity can be analyzed as a consequence of internalization of {V}{C} control. Moraic codas can be associated with {V}{C} or {V}[C] organization, i.e. a propensity to retain prototypical competitive control or minimal anticipatory de-gating. In contrast, non-moraic codas can be associated with {VC} organization, i.e. a substantial degree of anticipatory de-gating or anti-phase coordination. Studies of rime durations in languages with moraic and non-moraic codas (as assessed by phonological weight effects) show that rimes with non-moraic codas are shorter than ones with moraic codas (Broselow, Chen, & Huffman, 1997; Cohn, 2003; Duanmu, 1994). The contrast follows from the greater degree of overlap expected in {VC} vs. {V}{C}/[V][C] organization, which was shown previously in Fig. 15. Similarly, a distinction between geminates organized as {VC}{CV} vs. {V}{C}{CV} predicts durational differences observed between mora-sharing and moraic geminates (Ham, 2001). A prediction of this analysis is that regardless of which adult pattern emerges in a given language, young children will exhibit a transient stage in which durational measures of VC have a universal pattern that is consistent with {V}{C} organization.

6.4 Origins of phonological patterns

The selection-coordination-intention framework provides unified understandings of not only durational and temporal correlates of phonological organization, but also of more abstract associational organization and nonlocal patterns. Contour tone restrictions and quantity sensitivity have a common origin which

derives from the association of moraic structure with competitive organization and the transitivity of competitive organization. Whereas local assimilations can arise through several mechanisms, non-local assimilations can be understood to originate from interactions in intentional planning.

Contour tone restrictions

Contour tone restrictions arise because contour tones are competitively organized in early development and thus must be co-selected with different sets of gestures. In a variety of languages, contour tones are restricted to heavy syllables (Gordon, 2002), and this has been interpreted formally as the one-to-one association of tonal autosegments with moras (Hyman, 1985). Recent studies in the AP framework have argued that tones are analogous to articulatory gestures and that there is a corresponding tract variable coordinate in which tonal gestures specify F0 targets (Gao, 2008; Yi, 2017; Yi & Tilsen, 2015). If a contour tone such as HL is organized as {H}{L} in early development, then transitivity of selectional organization dictates that {H} and {L} cannot be co-selected and coordinated with the same set of oral articulatory gestures. Thus two sets, e.g. {V}{C} or {V}{V}, are necessary for production of a contour tone. Presumably in many languages control over the tonal gestures is eventually internalized and coordinative {HL} organization is learned; the contour tone can then be extended to other environments, such as {CV}. Nonetheless, the early restriction may give rise to a bias against production of contour tones in syllables which constitute just one set of co-selected gestures. The predictions of this analysis, which remain to be evaluated empirically, are that in the relevant developmental stage children will either truncate in contour tones in a monomoraic environment, or reselect the vocalic gesture to facilitate tone gesture selection.

Quantity sensitivity

The influence of syllable composition on the location of stress can also be viewed a consequence of selectional transitivity. In some languages, bimoraic syllables “attract” stress: a primary stress which would otherwise occur in a predictable position relative to a word boundary occurs on a heavy syllable (cf. Hayes, 1995). Here we conceptualize stress as an associational property which corresponds to the potential for a set of co-selected gestures to bear a pitch accent (Beckman & Edwards, 1994; Liberman & Prince, 1977), as opposed to viewing stress as a phonetic feature. As with lexical tones, pitch accents can be analyzed as gestures, and many pitch accents are bitonal (Mücke, Nam, Hermes, & Goldstein, 2012; Niemann, Mücke, Nam, Goldstein, & Grice, 2011; Tilsen, Burgess, & Lantz, 2013). In an early developmental stage, post-vocalic consonants, diphthongs, and long vowels are organized competitively as {V}{C}, {V}{V}, and {V}{V'}. The tonal gestures comprising a bitonal pitch accent such as H*+L are also organized competitively, i.e. {H}{L}. The production of such pitch accents is only possible in syllables with two sets of oral articulatory gestures. This circumstance creates a bias for selecting pitch accents on heavy syllables, which may lead to long-term memory associations which deviate from a regular stress pattern.

Gestural overlap and phonological assimilation

One of the most common cross-linguistic phonological patterns is assimilation of adjacent sounds. This often occurs in a sequence of intervocalic consonants, VC₁C₂V, for example. It is commonly understood that categorical assimilations such as *impossible* /in + p/ > [imp] arise on diachronic scales from hypocorrection of phonetic perturbations such as gestural overlap (Byrd, 1992; Gafos, 1999; Gafos & Goldstein, 2012; Ohala, 1993). Examining this account from the perspective of the intention mechanism, it is evident that gestural overlap can induce perturbations of targets in at least three distinct ways. The most direct perturbation occurs in the circumstance that two gestures C₁ and C₂ simultaneously provide different distributions of input to the same tract variable field. The weighted average of field activity, which determines the current vocal tract variable target, is thereby expected to obtain a value which is intermediate between the targets of the gestures (see Fig. 16B).

A less direct form of perturbation occurs in the circumstance that the targets of two gestures are planned in different tract variable fields and yet those tract variables are coupled to same model articulators. This form of perturbation is associated with many varieties of CV co-production. For example, the TR, TB, and JAW model articulators are employed to achieve vocalic postural targets, but JAW is employed in controlling lingual and labial consonantal constrictions as well (cf. the matrix of non-zero partial derivatives in Fig. 2B). Thus any circumstance in which consonantal and vocalic gestures are simultaneously active results in a biomechanical interaction between consonant and vowel. The immediate manifestation of this phenomenon is an influence on tract variable states. Presumably through long-term learning mechanisms, the intentional targets of the gestures may be altered to reflect these influences.

The least direct form of perturbation occurs in the circumstance that gestural overlap results in changes in the acoustic manifestations of gestures. Ohala (1993) proposed that no sound change arises if language learners perceptually compensate for context-conditioned acoustic perturbations of this sort, but if learners fail to compensate, they can reinterpret the production target. In other words, learners may form an alternative representation which is comprised of gestures with different targets and hence different patterns of excitation to parameter fields, which may involve the same or distinct tract variables. It is important to emphasize that the three forms of perturbation described above are not mutually exclusive, and the determination of their relative influence for a given assimilatory pattern is a worthwhile endeavor.

Long-distance phonological patterns

Phonological patterns which involve temporally distal sounds, such as vowel harmony (Nevins, 2010; Van der Hulst, 2011) and consonant harmonies (Hansson, 2010) can be understood to arise from subthreshold interactions between gestures which provide input to the same parameter field. Previous explanations of the origins of long distance patterns have invoked gestural overlap (Gafos & Dye, 2011). An alternative understanding, which can also shed light on the distinction between spreading vs. non-spreading (non-adjacent) harmonies and blocking phenomena (cf. Finley, 2015, 2017) is possible in the current framework (see also Smith, 2016). Fig. 25A shows how an anticipatory vowel harmony could arise. The quantal potential represents the activation states of gestural nodes of a $CV_1.CV_2.CV_3$ word form during the epoch of time in which the first syllable is selected. The vocalic gesture V_3 exerts abnormally strong excitatory and inhibitory input on a relevant parameter field. This results in a shift in the centroid of activation toward the target of V_3 even during the production of V_1 . A different sort of pattern involves the anticipatory or perseveratory activation of a gesture due to hypoinhibition. Fig. 25B shows anticipatory nasal harmony (velum opening) where the nasal (N) gestural node is selected early due to an absence of inhibition. Transparency and blocking phenomena can be derived from the extent to which co-active gestures exert sufficient inhibition on each other. The pattern can subsequently be phonologized as the early selection of a gesture or persistence of gestural selection. No artificial locality need be imposed to understand the origins of long-distance patterns in this approach, and featural tiers can be reinterpreted as independent parameter fields.

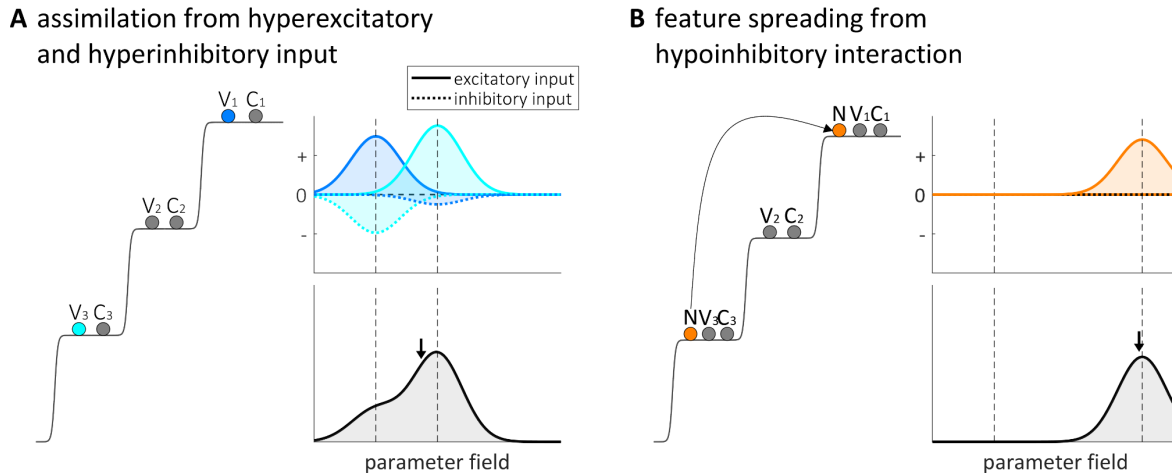


Fig. 25. Mechanisms for the origins of long-distance assimilatory phonological patterns. (A) Assimilation arises from hyper-excitatory input to a parameter from a gesture with subthreshold activation. (B) Feature spreading arises from hypo-inhibitory input to a parameter field.

6.5 Experimental manifestations of organization

Experimentally imposed constraints on speech influence how selection, coordination, and intention mechanisms interact with one another. It is therefore important to consider the potential task-specificity of organization when investigating speech behavior. Evidence also suggests that differences in organization provide a basis for speaker intuitions regarding phonological structure.

Syllable count intuitions

Cross-speaker variation in syllable count judgments can arise from the extent to which a post-vocalic gesture is competitively or coordinatively controlled. English words with long vowel/diphthong nuclei and liquid codas (e.g. *pool*, *peel*, *pail*, *pile*, *pear*, *pyre*) are unusual because intuitions about the number of syllables in such words vary across speakers and word forms (Cohn, 2003; Lavoie & Cohn, 1999). Tilsen & Cohn (2016) examined whether production of such forms correlates with syllable count intuitions. Acoustic evidence was observed for a greater degree of coarticulation between the vocalic and post-vocalic consonantal gesture in productions of words associated with monosyllabic judgments compared to those associated with disyllabic judgments. The correlation suggests that monosyllabic judgments are associated with a more prototypically coordinative regime in which the liquid coda and preceding vocalic gestures are co-selected, while disyllabic judgments are associated with more prototypically competitive control over the liquid gesture and hence involve relatively less extensive anticipatory de-gating.

Coordinative control bias in repetition tasks

Anti-phase to in-phase transitions in syllable repetition tasks provide evidence for coordinative control of post-vocalic consonantal gestures, but it is important to emphasize that repetition tasks by their very nature may create a bias for coordinative control. In rate-increasing syllable repetition, an abrupt transition from anti-phase (VC) to in-phase (CV) organization is observed (Sato, Schwartz, Abry, Cathiard, & Loevenbruck, 2006; Tuller & Kelso, 1990, 1991). Evidence that the transition involves an articulatory reorganization was discovered more than half a century ago (Stetson, 1951). The parallels between this phenomenon and the anti-phase to in-phase transition in finger wagging provide support for the AP model of control in which coda consonantal GPOs are anti-phase coupled to a vocalic GPO. However, repetition

tasks necessarily impose a periodic rhythm on gestural selection and suppression processes, and this may bias speakers to employ a control regime in which timing is coordinatively controlled via phase-coupling.

When gestural selection is not entrained to an external rhythm, post-vocalic gestures may exhibit timing patterns which are indicative of competitive organization. Evidence of this comes from an unusual experimental paradigm in (Tilsen, 2014), where speakers were randomly cued to insert an [s] into context syllables [pa] and [a], which were repeated in entrainment to metronome. Although the gestures of the context syllables were rhythmically selected, the inserted gesture was not. The relation between gestural overlap of the inserted [s] and the timing of the insertion cue was examined. It was observed that this relation differed between onset- and coda-syllabified insertions: the coda-syllabified insertions were substantially more resistant to intergestural compression. This finding is interpreted as evidence that post-vocalic gestures were selected after suppression of the vocalic gesture, rather than through relative phase-coupling.

Feedback perturbation

Whereas experimental tasks which entrain selection to an external rhythm appear to promote coordinative control, feedback perturbation paradigms may diminish reliance on internal feedback and hence promote prototypical competitive control. For example, mechanical perturbations of the lower lip during a bilabial closure in /p/ result in delayed and lengthened laryngeal abduction (Munhall, Löfqvist, & Kelso, 1994). Coordinative control of the laryngeal abduction gesture does not predict this effect because phase-coupling control over timing of gestural initiation does not depend on external sensory feedback. It seems plausible that speakers adapt to the expectation of a perturbation by reverting to reliance on external sensory feedback for gestural suppression and selection. More generally, it is observed that segmental durations are prolonged in auditory feedback perturbation studies (e.g. Cai, Ghosh, Guenther, & Perkell, 2011; Houde & Jordan, 1998, 2002; Larson, Burnett, Bauer, Kiran, & Hain, 2001; Purcell & Munhall, 2006; Tourville, Reilly, & Guenther, 2008; Villacorta, Perkell, & Guenther, 2007). This durational effect is expected if auditory feedback perturbations decrease reliance on internal feedback and thereby yoke gestural suppression and selection to external feedback regarding target achievement.

Gradient gestural intrusions

Some articulatory abnormalities can be understood as a consequence of hyperactive subthreshold intentional planning. In a two-syllable repetition task, Goldstein, Pouplier, Chen, Saltzman, & Byrd (2007) found that gradient gestural intrusion errors occur in which a gesture associated with one of the repeated syllables appears to be produced in both syllables. For example, in repetitions of *cop top*, the velar closure gesture associated with the onset of *cop* may be errorfully produced during the execution of the alveolar closure gesture associated with the onset of *top*. One of the interesting characteristics of such gestural intrusions is that they occur in a gradient manner: the vocal tract variable (a constriction degree) associated with the intrusion is partway between the value associated with a non-errorful target and the value observed in the absence of the intrusive gesture. This gradient is unexpected if the intrusive gesture is selected and executed in the canonical manner. The current framework provides a novel way of understanding the phenomenon: intrusive gestures result from hyperactive input to a planning field from a gestural node that either fails to be triggered or obtains subthreshold activation. As shown in Fig. 26, the input from C_1 , which represents the TD gesture of [k] in *cop top*, provides increasingly strong input to its associated parameter field, even in epochs 4 and 6 when the C_1 gesture is not selected.

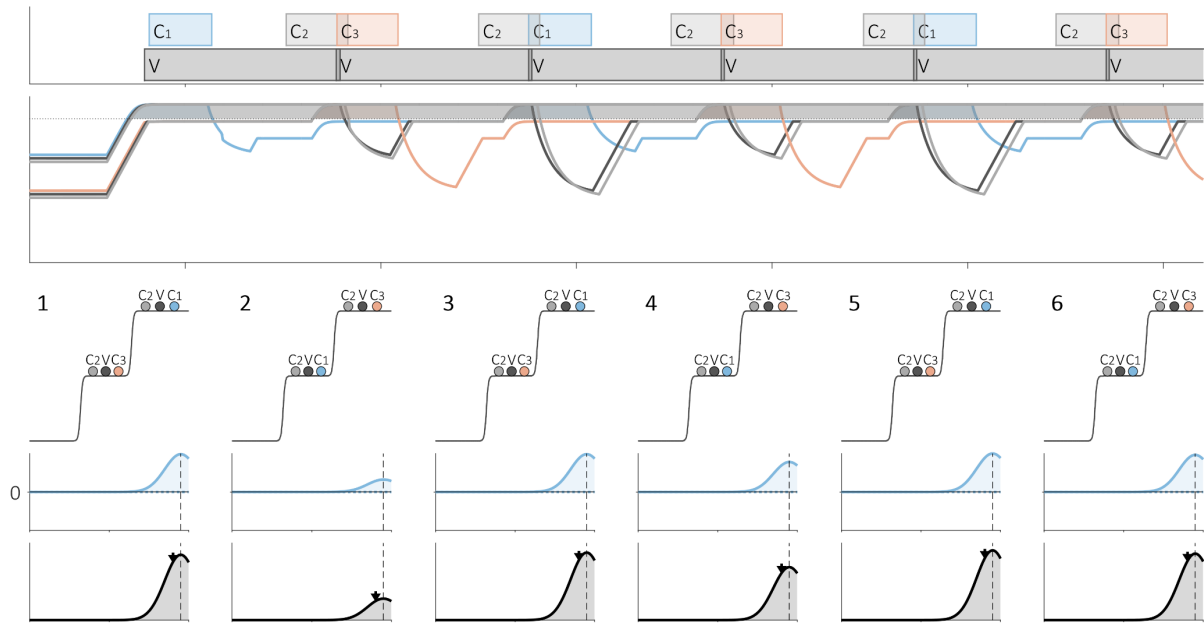


Fig. 26. Gradient gestural intrusions in *cop top* result from hyperactive input to a parameter planning field.

The observation that intrusive movements become more likely with each repetition suggests several possible causes of the hyperactivation of the parameter field. The inhibitory interactions which suppress the C_1 gestural node may decrease over time, the gain of gestural node input to the parameter field may increase, and/or the decay rate of parameter field activation in the absence of input may decrease. One thing to note about the above model of gestural intrusions is that the anomalous movements associated with the intrusion are in fact not “gestures” in any conventional sense; instead, the intrusive movements are a consequence of changes in the regulation of forces which gestural nodes exert on the vocal tract through their input to parameter fields.

7. Conclusion

This article explored the utility of three mechanisms—selection, coordination, and intention—for understanding a variety of speech phenomena. Selection regulates an activation code for choosing and ordering articulatory gestures, coordination regulates timing of co-selected gestures via a phase code governed by coupled gestural planning oscillators, and intention determines vocal tract parameter dynamics via a spatial code. Variation in how these mechanisms interact provides a powerful basis for understanding a diverse range of patterns.

It is important to reiterate that this aim of the article was to present a unified conceptual framework which integrates several existing conceptual frameworks. These frameworks address different aspects of articulatory control, and each contains layers of more specific hypothesis and implementational details. What is worth emphasizing here is that there are many possible approaches to computational implementation of mechanisms, but regardless of those details, it is evident that the mechanisms must be integrated in some way. The integration of mechanisms is necessary to achieve sufficient empirical coverage of phenomena.

An important consequence of taking a mechanistic approach to modeling articulation is that one must call into question the appropriateness of the traditional structural models. Considering the sources of variation examined herein, i.e. developmental, cross-linguistic, experimentally induced, and attentionally driven variation, leads to the conclusion that the organization of control may be quite flexible. The

hierarchical structures of units which are commonly employed in phonological theories therefore cannot be viewed as representations of a cognitive organization of any particular utterance. Rather, these conventional structures are best interpreted as summaries of different ways in which control can be organized, and as projections of such differences onto a time-independent space. This view calls for a widespread re-examination of experimental studies of units such as segments, moras, syllables, and feet within a more general paradigm. Rather than drawing conclusions about structure, investigation of speech in the selection-coordination-intention framework seeks insight regarding interactions of mechanisms.

References

- Abbs, J. H., & Connor, N. P. (1989). Motor coordination for functional human behaviors: Perspectives from a speech motor data base. In *Advances in psychology* (Vol. 61, pp. 157–183). Elsevier.
- Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology*, *51*(4), 705–723.
- Adams, I. L. J. (2018). *Predictive motor control in children with developmental coordination disorder: Mechanisms and intervention* (PhD Thesis). [SI: sn].
- Adams, I. L. J., Lust, J. M., Wilson, P. H., & Steenbergen, B. (2014). Compromised motor control in children with DCD: a deficit in the internal model?—A systematic review. *Neuroscience & Biobehavioral Reviews*, *47*, 225–244.
- Balasubramaniam, R., & Turvey, M. T. (2004). Coordination modes in the multisegmental dynamics of hula hooping. *Biological Cybernetics*, *90*(3), 176–190.
- Beckman, M., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. *Phonological Structure and Phonetic Form*, 1–33.
- Bernhardt, B., & Stemberger, J. (1998). *Handbook of phonological development: From the perspective of constraint-based nonlinear phonology*. San Diego: Academic Press.
- Bernstein, N. A. (1967). *The co-ordination and regulation of movements*. London: Pergamon.
- Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience*, *22*(7), 1504–1529.
- Broselow, E., Chen, S.-I., & Huffman, M. (1997). Syllable weight: convergence of phonology and phonetics. *Phonology*, *14*(1), 47–82.
- Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*(2–4), 140–155.
- Browman, C., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, *18*(3), 299–320.
- Browman, C., & Goldstein, L. (1991). Gestural structures: Distinctiveness, phonological processes, and historical change. In *Modularity and the motor theory of speech perception: Proceedings of a conference to honor Alvin M. Liberman* (pp. 313–338). Erlbaum Hillsdale, NJ.
- Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3–4), 155–180.
- Browman, C., & Goldstein, L. (1995). Gestural syllable position effects in American English. *Producing Speech: Contemporary Issues*, 19–33.
- Browman, C., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de La Communication Parlée*, *5*, 25–34.
- Browman, C., Goldstein, L., Saltzman, E., & Smith, C. (1986). GEST: A computational model for speech production using dynamically defined articulatory gestures. *The Journal of the Acoustical Society of America*, *80*(S1), S97–S97.
- Bullock, D. (2004). Adaptive neural models of queuing and timing in fluent action. *Trends in Cognitive Sciences*, *8*(9), 426–433.

- Bullock, D., & Rhodes, B. (2002). Competitive queuing for planning and serial performance. *CAS/CNS Technical Report Series*, 3(003), 1–9.
- Byrd, D. (1992). Perception of assimilation in consonants clusters: A gestural model. *Phonetica*, 49(1), 1–24.
- Byrd, D. (1995). C-centers revisited. *Phonetica*, 52, 285–306.
- Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. *Papers in Laboratory Phonology V*, 70–87.
- Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, 26, 173–199.
- Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31(2), 149–180.
- Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2011). Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *The Journal of Neuroscience*, 31(45), 16483–16490.
- Chitoran, I., & Goldstein, L. (2006). Testing the phonological status of perceptual recoverability: Articulatory evidence from Georgian. In *Proc. of the 10th Conference on Laboratory Phonology, Paris, June 29th–July 1st* (pp. 69–70).
- Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. *Laboratory Phonology*, 7, 419–447.
- Cohn, A. (2003). Phonological Structure and Phonetic Duration: The Role of the Mora. *Working Papers of the Cornell Phonetics Laboratory*, 15, 69–100.
- Davis, B. L., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech, Language and Hearing Research*, 38(6), 1199.
- Davis, S. (1990). Italian onset structure and the distribution of il and lo. *Linguistics*, 28(1), 43–56.
- De Jong, K., Beckman, M. E., & Edwards, J. (1993). The interplay between prosodic structure and coarticulation. *Language and Speech*, 36(2–3), 197–212.
- Dell, F., & Elmedlaoui, M. (1985). Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal Of African Languages and Linguistics*, 7, 105–130.
- Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, 49(2), 137–173.
- Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, 4(11), 423–431.
- Duanmu, S. (1994). Syllabic weight and syllabic duration: A correlation between phonology and phonetics. *Phonology*, 11(1), 1–24.
- Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review*, 109(3), 545.
- Fee, J., & Ingram, D. (1982). Reduplication as a strategy of phonological development. *Journal of Child Language*, 9(01), 41–54.
- Feldman, A. G. (1986). Once more on the equilibrium-point hypothesis (lambda model) for motor control. *Journal of Motor Behavior*, 18(1), 17.
- Feldman, A. G., & Levin, M. F. (2009). The equilibrium-point hypothesis—past, present and future. *Progress in Motor Control*, 699–726.
- Ferguson, C. A. (1983). Reduplication in child phonology. *Journal of Child Language*, 10(2), 239–243.
- Finley, S. (2015). Learning nonadjacent dependencies in phonology: Transparent vowels in vowel harmony. *Language*, 91(1), 48.
- Finley, S. (2017). Locality and harmony: Perspectives from artificial grammar learning. *Language and Linguistics Compass*, 11(1).

- Fitch, H. L., Tuller, B., & Turvey, M. T. (1982). The Bernstein perspective: III. Tuning of coordinative structures with special reference to perception. *Human Motor Behavior: An Introduction*, 271–281.
- Forsberg, H. (1999). Neural control of human motor development. *Current Opinion in Neurobiology*, 9(6), 676–682.
- Forsberg, H., Kinoshita, H., Eliasson, A. C., Johansson, R. S., Westling, G., & Gordon, A. M. (1992). Development of human precision grip. *Experimental Brain Research*, 90(2), 393–398.
- Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8(1), 113–133.
- Gafos, A. I. (1999). *The articulatory basis of locality in phonology*. Taylor & Francis.
- Gafos, A. I., & Dye, A. (2011). Vowel harmony: transparent and opaque vowels. *The Blackwell Companion to Phonology*, 4, 2164–2189.
- Gafos, A. I., & Goldstein, L. (2012). Articulatory representation and organization. In In: A. Cohn, C. Fougeron, & M.K. Huffman (Eds.), *The handbook of laboratory phonology* (pp. 220–231). New York: Oxford University Press.
- Gao, M. (2008). *Tonal Alignment in Mandarin Chinese: An Articulatory Phonology Account*. Doctoral Dissertation, Yale University, New Haven, CT.
- Ghez, C., Favilla, M., Ghilardi, M. F., Gordon, J., Bermejo, R., & Pullman, S. (1997). Discrete and continuous planning of hand movements and isometric force trajectories. *Experimental Brain Research*, 115(2), 217–233.
- Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. *Phonology and Second Language Acquisition*, 36, 315–328.
- Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. *Action to Language via the Mirror Neuron System*, 215–249.
- Goldstein, L., Chitoran, I., & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlihyt Berber. In *Proceedings of the XVth International Congress of Phonetic Sciences* (pp. 241–244).
- Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, 159–207.
- Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, 103(3), 386–412.
- Goodell, E. W., & Studdert-Kennedy, M. (1993). Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: A longitudinal study. *Journal of Speech, Language and Hearing Research*, 36(4), 707.
- Gordon, M. (2002). A typology of contour tone restrictions. *Studies in Language*, 25(3), 423–462.
- Greene, P. H. (1972). Problems of organization of motor systems. *Progress in Theoretical Biology*, 2, 303–338.
- Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. *Progress in Theoretical Biology*, 5, 233–374.
- Grossberg, S. (1987). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. *Advances in Psychology*, 43, 313–400.
- Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, 51(5), 347–356.
- Ham, W. H. (2001). *Phonetic and phonological aspects of geminate timing*. Routledge.
- Hansson, G. Ó. (2010). Consonant harmony: long-distance interaction in phonology. *UC Publications in Linguistics*.

- Hawkins, S. (1984). On the development of motor control in speech: Evidence from studies of temporal coordination. *Speech and Language: Advances in Basic Research and Practice*, 11. Retrieved from <http://www.haskins.yale.edu/Reprints/HL0472.pdf>
- Hayes, B. (1995). *Metrical stress theory: principles and case studies*. University of Chicago Press.
- Hermes, A., Grice, M., Mücke, D., & Niemann, H. (2008). Articulatory indicators of syllable affiliation in word initial consonant clusters in Italian. In *Proceedings of the 8th International Seminar on Speech Production* (Vol. 151, pp. 433–436). Strasbourg.
- Hermes, A., Mücke, D., & Grice, M. (2013). Gestural coordination of Italian word-initial clusters: the case of ‘impure s.’ *Phonology*, 30(1), 1–25.
- Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron*, 69(3), 407–422. <https://doi.org/10.1016/j.neuron.2011.01.019>
- Honorof, D. N., & Browman, C. (1995). The center or edge: How are consonant clusters organized with respect to the vowel. In *Proceedings of the XIIIth international congress of phonetic sciences* (Vol. 3, pp. 552–555).
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor Adaptation in Speech Production. *Science*, 279(5354), 1213–1216. <https://doi.org/10.1126/science.279.5354.1213>
- Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language and Hearing Research*, 45(2), 295.
- Houghton, G., & Hartley, T. (1995). Parallel models of serial behavior: Lashley revisited. *Psyche*, 2(25), 1–25.
- Houghton, G., & Tipper, S. P. (1994). A model of inhibitory mechanisms in selective attention.
- Houghton, G., & Tipper, S. P. (1996). Inhibitory mechanisms of neural and cognitive control: Applications to selective attention and sequential action. *Brain and Cognition*, 30(1), 20–43.
- Hyman, L. M. (1985). *A theory of phonological weight*. Dordrecht: Foris Publications.
- Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, 6(11–12), 11–12.
- Jeannerod, M. (1986). The formation of finger grip during prehension. A cortically mediated visuomotor pattern. *Behavioural Brain Research*, 19(2), 99–116.
- Jürgens, U. (2009). The Neural Control of Vocalization in Mammals: A Review. *Journal of Voice*, 23(1), 1–10. <https://doi.org/10.1016/j.jvoice.2007.07.005>
- Katz, W. F., & Bharadwaj, S. (2001). Coarticulation in fricative-vowel syllables produced by children and adults: a preliminary report. *Clinical Linguistics & Phonetics*, 15(1–2), 139–143.
- Kawamoto, A. H., Liu, Q., Mura, K., & Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language*, 58(2), 347–365.
- Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9(6), 718–727.
- Kawato, M., & Wolpert, D. (1998). Internal models for motor control. *Sensory Guidance of Movement*, 218, 291–307.
- Kaye, J. (1992). Do you believe in magic? The story of s+C sequences. *SOAS Working Papers in Linguistics and Phonetics*, 2, 293–313.
- Kelso, J. A. S., & Tuller, B. (1987). Intrinsic time in speech production: theory, methodology, and preliminary observations. *Sensory and Motor Processes in Language*. Hillsdale, NJ: Erlbaum, 203, 222.
- Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10(6), 812.

- Kent, R. (1983). The segmental organization of speech. In *The production of speech* (pp. 57–89). New York: Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4613-8202-7_4
- Kuhnert, B., Hoole, P., & Mooshammer, C. (2006). Gestural overlap and C-center in selected French consonant clusters. In *Proc. 7th International Seminar on Speech Production* (pp. 327–334).
- Lammert, A. C., Goldstein, L., & Iskarous, K. (2010). Locally-weighted regression for estimating the forward kinematics of a geometric vocal tract model. In *Eleventh Annual Conference of the International Speech Communication Association*.
- Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., & Hain, T. C. (2001). Comparison of voice f responses to pitch-shift onset and offset conditions. *The Journal of the Acoustical Society of America*, *110*, 2845.
- Lashley, K. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–135). New York: Wiley.
- Latash, M. L., Scholz, J. P., & Schöner, G. (2002). Motor control strategies revealed in the structure of motor variability. *Exercise and Sport Sciences Reviews*, *30*(1), 26–31.
- Latash, M. L., Scholz, J. P., & Schöner, G. (2007). Toward a new theory of motor synergies. *Motor Control*, *11*(3), 276–308.
- Lavoie, L., & Cohn, A. (1999). Sesquisyllables of English: the structure of vowel-liquid syllables. In *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp. 109–112). Retrieved from http://conf.ling.cornell.edu/pdfs/Lavoie_Cohn_1999.pdf
- Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*(2), 249–336.
- MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, *288*(5465), 527–531.
- MacNeilage, P. F., Davis, B. L., & Matyear, C. L. (1997). Babbling and first words: Phonetic similarities and differences. *Speech Communication*, *22*(2), 269–277.
- Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control*, *14*(3), 380–407.
- Matthei, E. (1989). Crossing boundaries: More evidence for phonological constraints on early multi-word utterances. *Journal of Child Language*, *16*(1), 41–54.
- Mooshammer, C., Goldstein, L., Nam, H., McClure, S., Saltzman, E., & Tiede, M. (2012). Bridging planning and execution: Temporal planning of syllables. *Journal of Phonetics*, *40*(3), 347–389.
- Mücke, D., Nam, H., Hermes, A., & Goldstein, L. (2012). Coupling of tone and constriction gestures in pitch accents. In *Consonant Clusters and Structural Complexity* (pp. 205–230). Berlin: Mouton de Gruyter. Retrieved from http://books.google.com/books?hl=en&lr=&id=xtKtil7W75sC&oi=fnd&pg=PA205&dq=gao+2008+speech+mandarin+tone+articulatory+phonology&ots=fwwBSI__PB&sig=LO4VCWXVDTVmDQ7xIOT-MoF1VHI
- Munhall, K. G., Löfqvist, A., & Kelso, J. A. S. (1994). Lip–larynx coordination in speech: Effects of mechanical perturbations to the lower lip. *The Journal of the Acoustical Society of America*, *95*, 3605–3616.
- Nam, H. (2007). Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure. *Laboratory Phonology*, *9*, 483–506.
- Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. *Approaches to Phonological Complexity*, 299–328.
- Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th International Conference on Phonetic Sciences* (pp. 2253–2256). Barcelona, Spain.
- Nevins, A. (2010). *Locality in vowel harmony* (Vol. 55). Mit Press.

- Niemann, H., Mücke, D., Nam, H., Goldstein, L., & Grice, M. (2011). Tones as Gestures: the Case of Italian and German. *Proceedings of ICPhS XVII*, 1486–1489.
- Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech, Language and Hearing Research*, 32(1), 120.
- Nittrouer, S., Studdert-Kennedy, M., & Neely, S. T. (1996). How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables. *Journal of Speech, Language and Hearing Research*, 39(2), 379.
- Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., & Tiede, M. (2018). How do children organize their speech in the first years of life? Evidence from ultrasound imaging.
- Ohala, J. J. (1993). The phonetics of sound change. *Historical Linguistics: Problems and Perspectives*, 237–278.
- Ota, M. (2001). Phonological theory and the development of prosodic structure: Evidence from child Japanese. *Annual Review of Language Acquisition*, 1(1), 65–118.
- Pouplier, M., & Beňuš, Š. (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. *Laboratory Phonology*, 2(2), 243–273.
- Purcell, D. W., & Munhall, K. G. (2006). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, 120, 966.
- Ramanarayanan, V., Goldstein, L., Byrd, D., & Narayanan, S. S. (2013). An investigation of articulatory setting using real-time magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 134(1), 510–519.
- Ramanarayanan, V., Lammert, A., Goldstein, L., & Narayanan, S. (2014). Are articulatory settings mechanically advantageous for speech motor control? *PLoS One*, 9(8), e104168.
- Ramanarayanan, V., Parrell, B., Goldstein, L., Nagarajan, S., & Houde, J. (2016). A New Model of Speech Motor Control Based on Task Dynamics and State Feedback. In *INTERSPEECH* (pp. 3564–3568).
- Rastle, K., & Davis, M. H. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology: Human Perception and Performance; Journal of Experimental Psychology: Human Perception and Performance*, 28(2), 307.
- Reidy, P. F., Beckman, M. E., Litovsky, R. Y., & Edwards, J. (2015). The acquisition of English sibilant fricatives by children with bilateral cochlear implants. *The Scottish Consortium for ICPhS, 2015*.
- Repp, B. H. (1986). Some observations on the development of anticipatory coarticulation. *The Journal of the Acoustical Society of America*, 79, 1616–1619.
- Roon, K. D., & Gafos, A. I. (2016). Perceiving while producing: Modeling the dynamics of phonological planning. *Journal of Memory and Language*, 89, 222–243.
<https://doi.org/10.1016/j.jml.2016.01.005>
- Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, 19(4), 499–526.
- Saltzman, E., Goldstein, L., Browman, C., & Rubin, P. (1988). Modeling speech production using dynamic gestural structures. *The Journal of the Acoustical Society of America*, 84(S1), S146–S146.
- Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1(4), 333–382.
- Sato, M., Schwartz, J.-L., Abry, C., Cathiard, M.-A., & Loewenbrück, H. (2006). Multistable syllables as enacted percepts: a source of an asymmetric bias in the verbal transformation effect. *Perception & Psychophysics*, 68(3), 458–474.
- Schöner, G., Kopecz, K., & Erlhagen, W. (1997). The dynamic neural field theory of motor programming: Arm and eye movements. In *Advances in Psychology* (Vol. 119, pp. 271–310). Elsevier.
- Schwartz, R. G., Leonard, L. B., Wilcox, M. J., & Folger, M. K. (1980). Again and again: reduplication in child phonology. *Journal of Child Language*, 7(01), 75–87.

- Shaw, J., Gafos, A. I., Hoole, P., & Zeroual, C. (2009). Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation. *Phonology*, 26(01), 187–215.
- Shaw, J., Gafos, A. I., Hoole, P., & Zeroual, C. (2011). Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters. *Phonology*, 28(03), 455–490.
- Sheliga, B. M., Riggio, L., & Rizzolatti, G. (1994). Orienting of attention and eye movements. *Experimental Brain Research*, 98(3), 507–522.
- Smit, A. B. (1993). Phonologic error distributions in the Iowa-Nebraska articulation norms project: Word-initial consonant clusters. *Journal of Speech, Language and Hearing Research*, 36(5), 931–947.
- Smith, C. (2016). A gestural account of neutral segment asymmetries in harmony. In *Proceedings of the Annual Meetings on Phonology* (Vol. 3).
- Song, J. Y., & Demuth, K. (2008). Compensatory vowel lengthening for omitted coda consonants: A phonetic investigation of children’s early representations of prosodic words. *Language and Speech*, 51(4), 385–402.
- Sorensen, T., Toutios, A., Goldstein, L., & Narayanan, S. S. (2016). Characterizing Vocal Tract Dynamics Across Speakers Using Real-Time MRI. In *Interspeech* (pp. 465–469).
- Sproat, R., & Fujimura, O. (1993). Allophonic variation in English/l/and its implications for phonetic implementation. *Journal of Phonetics*, 21(3), 291–311.
- Sternberg, S., Knoll, R., Monsell, S., & Wright, C. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica*, 45(2–4), 175–197.
- Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. *Information Processing in Motor Control and Learning*, 117–152.
- Stetson, R. H. (1951). *Motor phonetics; a study of speech movements in action*. Amsterdam: North-Holland Publishing Company.
- Sussman, H. M., Duder, C., Dalston, E., & Cacciatore, A. (1999). An acoustic analysis of the development of CV coarticulation: A case study. *Journal of Speech, Language and Hearing Research*, 42(5), 1080–1096.
- Thelen, E. (1979). Rhythmical stereotypies in normal human infants. *Animal Behaviour*, 27, 699–715.
- Tilsen, S. (2007). Vowel-to-vowel coarticulation and dissimilation in phonemic-response priming. *UC Berkeley Phonology Lab 2007 Annual Report*, 416–458.
- Tilsen, S. (2009a). *Hierarchical spatiotemporal dynamics of speech rhythm and articulation*. University of California, Berkeley.
- Tilsen, S. (2009b). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics*, 37(3), 276–296.
- Tilsen, S. (2013a). A Dynamical Model of Hierarchical Selection and Coordination in Speech Planning. *PLoS One*, 8(4), e62800.
- Tilsen, S. (2013b). Is contrast maximization accomplished by perception or motor planning?
- Tilsen, S. (2014). Selection and coordination of articulatory gestures in temporally constrained production. *Journal of Phonetics*, 44, 26–46.
- Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, 55, 53–77.
- Tilsen, S. (2017). Exertive modulation of speech and articulatory phasing. *Journal of Phonetics*, 64, 34–50.
- Tilsen, S., Burgess, D., & Lantz, E. (2013). Imitation of intonational gestures: a preliminary report. *Cornell Working Papers in Phonetics and Phonology*, 2013, 1–17.
- Tilsen, S., & Cohn, A. (2016). Shared representations underlie metaphonological judgments and speech motor control. *Laboratory Phonology*.

- Tilsen, S., & Goldstein, L. (2012). Articulatory gestures are individually selected in production. *Journal of Phonetics*, 40(6), 764–779.
- Tilsen, S., Spincemaille, P., Xu, B., Doerschuk, P., Luh, W.-M., Feldman, E., & Wang, Y. (2016). Anticipatory Posturing of the Vocal Tract Reveals Dissociation of Speech Movement Plans from Linguistic Units. *PLoS ONE*, 11(1), e0146813. <https://doi.org/10.1371/journal.pone.0146813>
- Tilsen, S., Zec, D., Bjorndahl, C., Butler, B., L'Esperance, M.-J., Fisher, A., ... Sanker, C. (2012). A cross-linguistic investigation of articulatory coordination in word-initial consonant clusters. *Cornell Working Papers in Phonetics and Phonology*, 2012, 51–81.
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, 39(3), 1429–1443.
- Tuller, B., & Kelso, J. A. S. (1990). Phase transitions in speech production and their perceptual consequences. *Attention and Performance*, 13, 429–452.
- Tuller, B., & Kelso, J. A. S. (1991). The production and perception of syllable structure. *Journal of Speech, Language and Hearing Research*, 34(3), 501–508.
- Turnbaugh, K. R., Hoffman, P. R., Daniloff, R. G., & Absher, R. (1985). Stop-vowel coarticulation in 3-year-old, 5-year-old, and adult speakers. *The Journal of the Acoustical Society of America*, 77, 1256–1257.
- Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. *Perceiving, Acting and Knowing*, 211–265.
- Turvey, M. T. (1990). Coordination. *American Psychologist*, 45(8), 938–953.
- Van der Hulst, H. (2011). *Vowel harmony*. Oxford University Press.
- Van der Stigchel, S., Meeter, M., & Theeuwes, J. (2006). Eye movement trajectories and what they tell us. *Neuroscience & Biobehavioral Reviews*, 30(5), 666–679.
- Van der Stigchel, S., & Theeuwes, J. (2005). Relation between saccade trajectories and spatial distractor locations. *Cognitive Brain Research*, 25(2), 579–582.
- van Lieshout, P., Merrick, G., & Goldstein, L. (2008). An articulatory phonology perspective on rhotic articulation problems: A descriptive case study. *Asia Pacific Journal of Speech, Language and Hearing*, 11(4), 283–303.
- Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, 122(4), 2306–2319.
- von Hofsten, C. (1979). *Development of visually directed reaching: The approach phase*. Department of psychology, University of Uppsala.
- Wheeldon, L., & Lahiri, A. (1997). Prosodic Units in Speech Production* 1,* 2,* 3. *Journal of Memory and Language*, 37(3), 356–381.
- Wheeldon, L., & Lahiri, A. (2002). The minimal unit of phonological encoding: prosodic or lexical word. *Cognition*, 85(2), B31–B41.
- Wilson, P. H., & Butson, M. (2007). Deficits underlying DCD. *Developmental Coordination Disorder: A Review of Current Approaches*. Solal Editeurs, Marseille, 115–119.
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232), 1880–1882.
- Yi, H. (2017). *Lexical tone gestures*. Cornell University.
- Yi, H., & Tilsen, S. (2015). Gestural timing in Mandarin tone sandhi. *Proceedings of Meetings on Acoustics*, 22(1), 060003. <https://doi.org/10.1121/2.0000029>
- Ziegler, W., & Ackermann, H. (2017). Subcortical Contributions to Motor Speech: Phylogenetic, Developmental, Clinical. *Trends in Neurosciences*, 40(8), 458–468.