



Research Article

Selection and coordination: The articulatory basis for the emergence of phonological structure



Sam Tilsen

Cornell University, 203 Morrill Hall, Ithaca, NY 14853, USA

ARTICLE INFO

Article history:

Received 27 May 2015

Received in revised form

27 October 2015

Accepted 30 November 2015

Keywords:

Speech production

Motor control

Motor feedback

Speech development

Language acquisition

Articulatory phonology

Phonological typology

ABSTRACT

Phonological theories commonly analyze speech utterances as composed of hierarchically organized units, such as features/gestures, segments, moras, and syllables, but it is not well understood why this hierarchical organization is observed. Moreover, current phonological theories and speech production models fail to explain cross-linguistic and developmental variation in the organization of units. This paper presents the *selection-coordination theory* of speech production, which attempts to unify our understanding of developmental and cross-linguistic variation in phonological structure. The theory holds that hierarchical organization emerges from a recurring trend in speech development whereby children acquire coordinative regimes of control over articulatory gestures that were previously competitively selected. In this framework, segments, moras, and syllables are understood as differently-sized instantiations of the same type of motor planning unit, and cross-linguistic and developmental phonological patterns are derived from distinguishing competitive and coordinative regimes of articulatory control. Evidence for the theory is drawn from research in motor control, speech development, and phonological and phonetic patterns in speech.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

This paper describes the selection-coordination theory of speech production, and explains how the theory helps us understand cross-linguistic and developmental phonological patterns. The gist of the theory is the interaction of two motor control mechanisms—selection and coordination—results in two prototypical articulatory control regimes: competitive control and coordinative control. Phonological structure involving segments, moras, and syllables emerges through developmental transitions from competitive to coordinative control, mediated by the internalization of sensory feedback.

The problems which motivate the development of selection-coordination theory can be summarized as three ways in which current approaches are insufficiently explanatory: (1) *Hierarchical structure* of utterances is not explained by formal representations or existing production models; these models presuppose structure rather than motivating it from cognitive mechanisms. (2) *Cross-linguistic variation* in phonological structure is described but not explained by current approaches; mechanisms for understanding how or why such variation emerges are absent. (3) *Developmental patterns* in speech are not explained, nor related to other forms of variation; existing approaches are either too rigidly structured or lack the necessary mechanisms to generate empirically observed phenomena.

Here we show that selection-coordination theory addresses these problems and provides a unified account of articulatory control, developmental patterns, and phonological structure. The theory elaborates on ideas which have been implemented computationally and to a limited extent tested experimentally in previous work (Tilsen, 2013, 2014a); here the ideas are developed in detail and applied to a broad range of phenomena, although the scope of the theory is such that treatment of all relevant issues is not possible. The manuscript is organized as follows:

Section 1 discusses how current frameworks are either too rigidly structured (in the case of hierarchical production models) or lack the necessary mechanisms for understanding hierarchical structure (in the case of Articulatory Phonology).

E-mail address: tilsen@cornell.edu

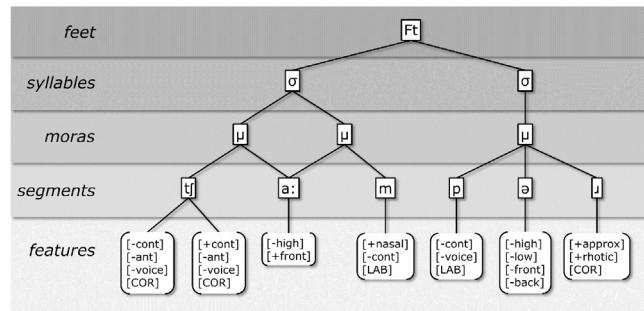


Fig. 1. Hierarchical representation of the word *chomper* illustrating various structural non-isomorphisms: multiple sets of features associated with one segment, multiple segments associated with one mora, and multiple moras associated with one segment.

Section 2 presents the theory in detail, describing the mechanisms of selection and coordination, their interaction, and the role of sensory feedback in developmental transitions from competitive to coordinative control.

Section 3 shows how the theory helps us understand phonetic and phonological patterns in development.

Sections 4 and 5, focusing on segments and moras/syllables respectively, provide a new understanding of phonological structure and apply selection-coordination theory to shed light on phonological and phonetic variation.

1.1. Hierarchical structure in phonological representations

Most phonological theories analyze speech utterances into hierarchically organized structures that consist of features or gestures, segments, moras, syllables, feet, and phrases. Problematically, current theories presuppose that these units are essential components of the representation, failing to satisfactorily motivate the existence of levels in the hierarchy. A hierarchically structured representation of the word “chomper” is shown in Fig. 1. The representation consists of units which are organized into spatially distinct levels or tiers. Representations of this sort rely on several conceptual metaphors (cf. Tilsen, 2009b): speech sounds and groups of speech sounds are objects, time is a space in which objects are located, and functional associations between sounds and groups of speech sounds are connections between objects. Consequently formal phonological theories have aimed to elucidate principles and constraints on the presence, absence, ordering, and connections between units.

Formal phonological representations dissociate the content, ordering, and representations of segments, thereby allowing for various non-isomorphisms as shown in Fig. 1. For the affricate [tʃ] a single segment is associated with multiple sets of features (Lombardi, 1990; Sagey, 1986), for the long vowel [a:] one set of features is associated with multiple moras, and all three moras are associated with multiple segments (Hyman, 1984, 1985; McCarthy & Prince, 1986).

Hierarchical phonological representations and non-isomorphisms between levels provide useful symbolic/geometric metaphors for describing phonological patterns but do not imply very much about cognitive mechanisms that could explain how such patterns arise. Phonological representations make only very limited predictions regarding the physical realizations of utterances; such predictions are generally viewed as the province of speech production models. Below we distinguish two general classes of such models.

1.2. Hierarchically structured production models

Most models of speech production are hierarchically structured and assume that segments and syllables play key roles in organizing speech motor planning. Importantly, none of these models incorporates a moraic level of motor organization, and they have no mechanisms for restructuring the network of units over time.

One example is the spreading activation model of Dell (1986), which views morphemes, syllables, syllable sub-constituents, segments, and features as nodes in a hierarchically structured network through which activation spreads. The selection of items on one level of representation is preceded by the selection of items on the next highest level. Thus the representation of a speech plan is constructed in a cascade, where the unit selected on a given level is the one with the highest amount of activation. The model provides an account of sequencing errors, and related models (Goldrick & Blumstein, 2006; Goldrick, Baker, Murphy, & Baese-Berk, 2011) have incorporated more detail regarding how spreading activation mechanisms interact with articulatory processes.

The gradient-ordering, directions into velocities of articulators model (GODIVA: Bohland, Bullock, & Guenther, 2010), an extension of the DIVA model (Guenther, Ghosh, & Tourville, 2006), views speech planning as a process in which segments are competitively selected to fill positions in syllable frames. Articulations are produced sequentially in association with segments via a mapping from sensory targets to motor commands, in conjunction with feedforward and feedback controllers.

The production model of Levelt (1993) presents a similar view of speech planning as the filling of timing slots by segmental material. Abstract plans are specified in the lexicon as a sequence of segments. A phonological encoding module groups segments into syllable-sized units and an articulatory module implements gestures with timing patterns determined by their syllabification. In recent versions of the model, moraic structure has been cited as “an emerging property of the syllabification process” (1999: 21), but without elaboration of how this occurs.

Although successful in some ways, the above models are limited in that they are not designed to accommodate changes may occur on a developmental timescale—these models describe a static organization of control that revolves around segments and syllables.

1.3. The non-hierarchical model of articulatory phonology

Articulatory Phonology (Browman & Goldstein, 1986, 1989, 1992), based on the task dynamic model of motor control (Kello, Saltzman, & Tuller, 1986; Saltzman & Munhall, 1989), rejects the assumption that the segment is a functional level of articulatory organization. Phonological representations are composed of primitive elements called *gestures*, which are dynamical systems that drive motor synergies to achieve target vocal tract geometries. For example, in a bilabial stop like [p], one gestural target is a bilabial closure which the upper lip, lower lip, and jaw can contribute to. The voicelessness of the bilabial stop is accomplished by a vocal fold abduction gesture. Hence the segment [p] can be viewed as a combination of a bilabial closure gesture and a glottal abduction gesture. Articulatory phonology also incorporates a set of hypotheses regarding the nature of syllable structure, based on the idea that articulatory timing can be derived from the relative phases of a network of phase-coupled planning oscillators (Browman & Goldstein, 2000; Goldstein, Byrd, & Saltzman, 2006).

By emphasizing local interactions between gestures, Articulatory Phonology imputes no special status to combinations of gestures that might correspond to segments, instead viewing segments and larger structures as constellations of gestures. Browman and Goldstein (1990) argue that “the basis for [segments] seems to be their utility as a practical tool rather than their correspondence to important informational units of the phonological system” (1990: 88), and hold that “the pattern of relations among a set of gestures, the gestural constellation, can serve the functions typically filled by other phonological structures, ranging from complex segments to syllables and their constituents” (1986: 225).

Thus Articulatory Phonology does not presuppose any form of hierarchical structure; segments, moras, and syllables are viewed as constellations within constellations, with none of them having any special role in organizing articulatory control. The notion that constellations can serve the functions of other phonological structures has not been further developed. Selection-coordination theory can be viewed as an extension of Articulatory Phonology that develops a more elaborated conception of sets of gestures.

2. Selection-coordination theory

Previous research in speech motor control supports distinguishing between a mechanism for choosing which movements to produce and a mechanism for precision control over the timing of movement execution. Here these mechanisms are called *selection* and *coordination*, respectively, although note that numerous researchers have drawn related distinctions with alternative terminology, e.g. chain-like vs. comb-like organization (Ohala, 1975), sequencing vs. timing (MacKay, 1982), succession vs. synchrony (Kent, 1983), subprogram selection vs. command (Sternberg, Knoll, Monsell, & Wright, 1988), serial ordering vs. parallelism (Jordan, 1997), response selection vs. timing adjustment (Sakai et al., 2000), sequential action vs. movement timing (Schöner & Santos, 2001).

Selection-coordination theory posits that the interaction between selection and coordination mechanisms gives rise to two prototypical regimes of control over speech movements: competitive control and coordinative control. For an analogy consider the piano score in Fig. 2A, where notes are played in a mutually exclusive sequence, corresponding to competitive control. In contrast multiple notes can be played together as a chord or as a principle note co-produced with grace notes, various examples of which are shown in Fig. 2B and C.

Distinguishing competitive and coordinative control provides insight into the structure of speech utterances and phonological patterns; in Sections 5.1 and 5.2 we will see how variation in the organization of control accounts for cross-linguistic distinctions such as simplex vs. complex onsets, and moraic vs. mora-sharing codas. It is important to keep in mind that selection and coordination are *mechanisms* involved in articulation, while competitive control and coordinative control are *regimes* or *modes* of articulatory control that arise from the interaction of selection and coordination. Sections 2.1 and 2.2 below describe prototypical competitive and

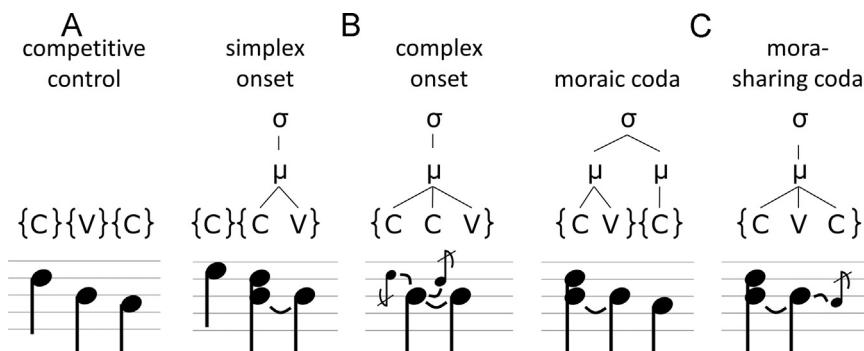


Fig. 2. Analogy between articulatory timing and piano notes. (A) competitive control, where notes/articulatory gestures are produced in a mutually exclusive sequence. (B,C) Examples of piano sequences with chords and/or grace notes, which correspond to co-selected, coordinatively controlled articulatory gestures. The correspondences shown between scores and syllable structures are discussed in Sections 5.1 and 5.2.

coordinative control; Section 2.3 discusses how coordinative control develops from competitive control through the internalization of feedback.

2.1. Competitive control: selecting some movement plans to the exclusion of others

Competitive control describes control in which a motor plan or a set of plans is selected to the exclusion of others, with feedback governing selection and deselection. The concept of competitive control originates with Lashley (1951), who argued that representations of actions in a sequence must be active in parallel before and during the production of the sequence, with some mechanism responsible for selecting those actions in the appropriate order. These arguments were motivated by observations of anticipatory and perseveratory error patterns in movement sequences and free combinatoriality of movements. Several models developed in this paradigm have been based on computer memory buffer metaphors (Shaffer, 1976; Sternberg et al., 1988), but for current purposes we focus on a more dynamical approach (Bullock & Rhodes, 2002; Grossberg, 1978, 1987).

A dynamical model of competitive control over three action plans is shown in Fig. 3A. Plans X, Y, and Z are associated with activation variables that have initial relative activations $X > Y > Z$. When the response is initiated, a competition process begins and activation levels grow. When the activation of X reaches a selection threshold, X is selected for execution and competing plans are temporally gated, i.e. prevented from being selected. When feedback regarding successful completion of movements associated with X is received, the selected plan is suppressed. When the activation of X falls below the selection threshold, X is deselected and a new round of competition is initiated. Fig. 3B shows how this process can be conceptualized as a cycle of competition, selection and gating of non-selected competitors, feedback-induced suppression, and deselection with deactivation of gating.

Competitive selection models are an innovation over stipulating memory for serial order, because the conceptual device for representing motor plans—activation—plays a direct role in the mechanism that sequences movements. Competitive selection models are successful in accounting for response initiation latencies and error patterns (Bullock, 2004; Bullock & Rhodes, 2002). When more plans are co-active, inhibitory interactions between them diminishes activation thereby increasing latencies to selection. This accounts for the robust finding that reaction time to initiate a movement sequence increases linearly with the length of the sequence (Henry & Rogers, 1960; Klapp, 1995; Rosenbaum, Cohen, Jax, Weiss, & van der Wel, 2007; Sternberg et al., 1988; Sternberg, Monsell, Knoll, & Wright, 1978). Neural recordings in primates trained to perform sequential actions also provide support for competitive selection dynamics (Averbeck, Chafee, Crowe, & Georgopoulos, 2002; Averbeck, Crowe, Chafee, & Georgopoulos, 2003; Cisek & Kalaska, 2005).

Various types of errors can be readily understood in a competitive selection framework. Transposition sequencing errors can arise when noise or external influences alter the relative activations of plans, causing plans to be selected in a nonstandard order, as in Fig. 3C. Furthermore, a previously selected movement plan can be errorfully re-selected at the expense of a subsequent one, as in Fig. 3D. This can occur for several reasons: suppression of the first plan may be insufficient, the first plan may be hyperactive, or the second plan may be only weakly activated. Errorful reselection associated with deficient feedback may also be the mechanism behind stuttering (Max, Guenther, Gracco, Ghosh, & Wallace, 2004).

Selection-coordination theory hypothesizes that reselection of a previously selected plan, as shown in Fig. 3E, can be used to prolong the period of time that a target posture is maintained. No apparent discontinuity need be evident in the articulatory or acoustic manifestation of an intentionally reselected plan, since no competing plan is selected between the first selection and the reselection. Long vowels and geminate consonants are viewed here as intentionally reselected vocalic or consonantal gestures, and we will see later on how this augments our understanding of various developmental and phonological patterns.

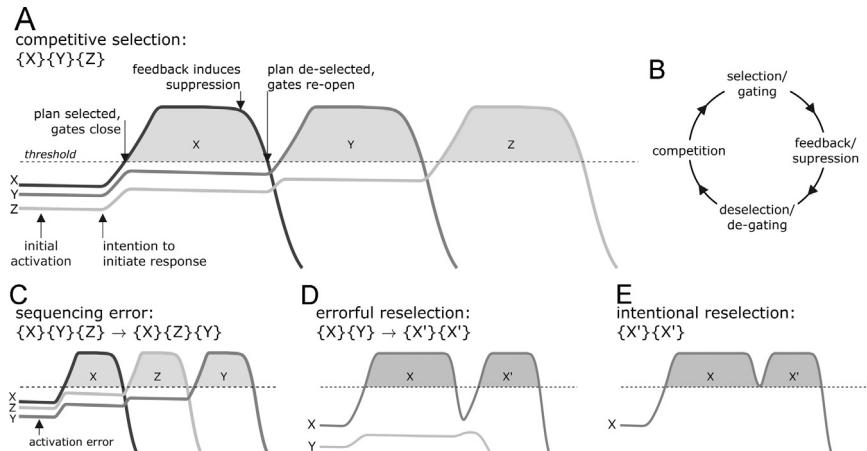


Fig. 3. Illustration of competitive selection dynamics governing the sequencing of three action plans. (A) Prototypical competitive selection. (B) Cyclic conceptualization of competitive selection. (C) Sequencing error resulting from non-standard activation gradient. (D) Errorful reselection of a previously selected plan, at the expense of a competitor. (E) Intentional reselection as a strategy to achieve length. See text for further detail.

2.2. Coordinative control: precision control of movement timing

Many actions involve multiple movements conducted in a precisely timed, overlapping manner; competitive control cannot readily model such patterns, and hence an alternative mechanism is required. For example, when adults reach to grab an object, they adjust hand-shape while transporting the hand to a desired location: the reaching and hand-shaping movements occur in parallel rather than being sequenced (Jeannerod, 1986). In speech, the consonantal and vocalic movements in CV syllables overlap (Kozhevnikov & Chistovich, 1965; Öhman, 1967; Perkell, 1969) and are initiated with nearly synchronous relative timing (Brownman & Goldstein, 1988; Krakow, 1989). Because the vocalic movements of a CV syllable are executed before feedback is received from the consonantal movements, competitive selection is unable to generate the empirical timing pattern. Coordinative mechanisms, on the other hand, provide a useful account of overlapping, precisely timed movements.

The concept of coordination was developed by Bernstein (1967) to address the "many degrees-of-freedom problem": accomplishing a given movement typically involves control over many joints and muscles, and given the dependence of movements on initial postural conditions and other perturbations, control cannot be specified straightforwardly in terms of relations between goals and individual degrees of freedom. Bernstein proposed that groups of muscles become linked in synergies or "coordinative structures", thereby reducing the degrees of freedom (Kelso, 1982). Early research from this perspective led to some basic principles of coordination (cf. Turvey, 1990): plans for movements refer to goals rather than muscles, joints, or trajectories; control is accomplished by interactions among multiple, relatively autonomous subsystems; muscles can become linked in a task-specific manner into coordinative structures that respond immediately to perturbations; and ballistic movements operate like critically damped springs. Further theoretical work has related motor coordination to self-organization in physical and biological systems: degrees of freedom organize into low-dimensional collective variables, the dynamics of which are governed by general principles.

The task-dynamic model of articulatory phonology holds that articulatory gestures are the basic units of speech, and views gestures as coordinative structures controlling multiple articulators to accomplish some goal or task (Brownman & Goldstein, 1989; Saltzman & Munhall, 1989). Gestures are associated with target values of task variables, which are typically defined in coordinates of vocal tract geometry. For example, a bilabial closure gesture is associated with the task of closing the lips, which corresponds to a zero or negative target value of lip aperture. This is accomplished by coordinating the movements of the lower lip, upper lip, and jaw. Articulatory gestures are hypothesized to be the fundamental units of both information and action in phonological systems (Goldstein et al., 2006). They are modeled as 1-dimensional dynamical systems with point-attractors, analogous to critically damped mass-spring systems, where the point attractor is some vocal tract variable target. Through an inverse model that maps changes in vocal tract variables to changes in articulator/effectuator positions, the model allows for gestures to reach their targets regardless of initial conditions.

In an important conceptual development, coordinative principles were extended to the control of relative timing of gestures; rhythmic movements have played a fundamental role in this regard. Periodic motions can be associated with limit-cycle oscillators, whose nature is determined by elastic (spring-like) forces, friction/damping, and an energy source to compensate for damping (Kay, Kelso, Saltzman, & Schöner, 1987). Only one degree of freedom, *phase*, is necessary to describe the state of a rhythmic movement, and the coordination of two rhythmic movements can be analyzed with a collective variable, *relative phase*. When strongly coupled,

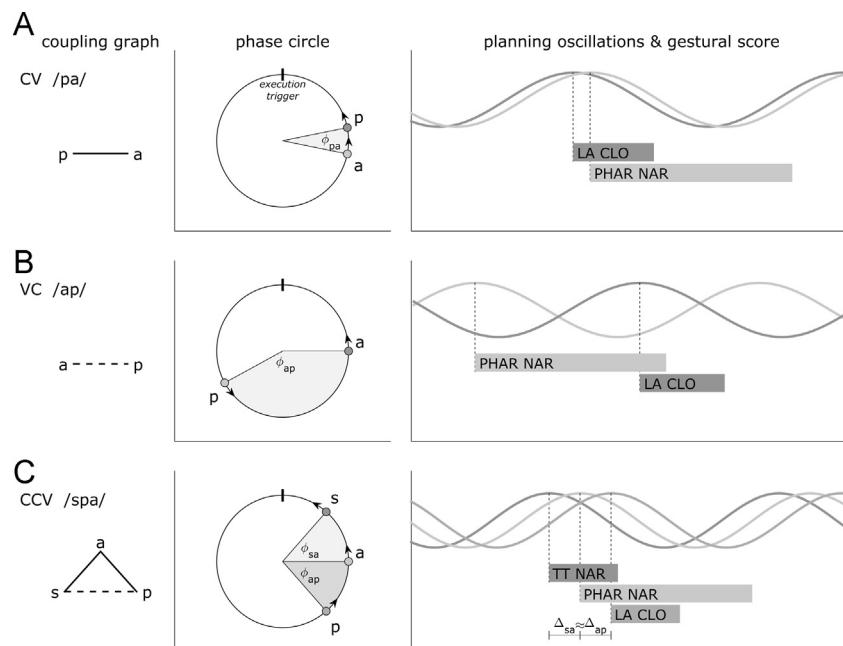


Fig. 4. Illustration of the coupled oscillators model of intergestural coordination for several different syllable shapes. For each syllable a hypothesized coupling graph is shown (solid lines: in-phase coupling; dashed lines: anti-phase coupling), along with a representation of phases of planning oscillators on a phase circle. Peaks in planning oscillations correspond to the top of the phase circle and determine when gestures become active.

oscillations with different intrinsic frequencies will entrain to a compromise frequency. Moreover, there are two preferred ways in which a pair of oscillations at similar frequencies can be synchronized: in-phase (0°) and anti-phase (180°), the former being more stable than the latter. In a classic experiment, Haken, Kelso, and Bunz (1985) showed that people attempting to wag their fingers in an anti-phase mode of coordination will exhibit a spontaneous transition from anti-phase to in-phase coordination as movement frequency is increased. Relative phase transitions of this sort have been demonstrated with a variety of limbs and between people (Schmidt, Bienvenu, Fitzpatrick, & Amazeen, 1998; Schmidt & Turvey, 1994), and also account for gait transitions in quadruped locomotion (Schöner, Jiang, & Kelso, 1990; Turvey, 2007).

The co-existence of relatively stable in-phase and anti-phase modes of coordination is observed in speech as well. Stetson (1951) found that when speakers repeat a VC syllable with increasing rate, the timing of the articulatory movements reorganizes to a CV-like pattern. Tuller and Kelso (1990, 1991) replicated this effect, observing relative phase transitions from anti-phase to in-phase coordination of consonantal and vocalic gestures when a critical value of repetition rate is reached (see also De Jong, 2001; Hertrich & Ackermann, 2000; Sato, Schwartz, Abry, Cathiard, & Loevenbruck, 2006). This transition is closely analogous to the phase transition in finger wagging, exhibiting critical fluctuations near the phase transition and hysteresis as the rate parameter is varied.

To apply principles of rhythmic coordination to the relative timing of speech movements, which are typically non-rhythmic, a conceptual leap is required. Drawing on the notion of intrinsic time (Kelso & Holt, 1980; Prigogine & Stengers, 1984; Richardson & Rosen, 1979), Kelso and Tuller (1987) argued that a “virtual cycle” for a gesture could be defined according to its stiffness parameter, which determines frequency of oscillation in an undamped response to displacement. Hence even non-rhythmic movements can be coordinated via the relative phases of their virtual cycles. These innovative ideas provided the basis for the concept of *phasing* in the gestural scores of articulatory phonology (Browman & Goldstein, 1986, 1992; Saltzman & Munhall, 1989). Control of phasing was subsequently associated with stabilized relative phases in systems of coupled planning oscillators (Browman & Goldstein, 2000; Saltzman & Byrd, 2000).

The coupled oscillators model constitutes a set of hypotheses regarding relative phase coupling between articulatory gestures in various syllable positions. Onset consonantal gestures are in-phase coupled to a following vocalic gesture (Fig. 4A) and coda consonantal gestures are anti-phase coupled to a preceding vocalic gesture (Fig. 4B) or to a preceding consonantal gesture in the case of coda clusters. Fig. 4 illustrates how these relations can be visualized in several ways: as graph topologies in which edges between gestures represent in-phase or anti-phase coupling forces (note that coupling graphs do not represent ordering in time); as configurations in phase space where coupling forces minimize (in-phase coupling) or maximize (anti-phase coupling) relative phases; and as oscillations in time with gestures being activated when oscillations reach their peaks.

One of the most successful applications of the coupled oscillators model involves the C-center effect (Browman & Goldstein, 1988), which describes the empirical observation that initiations of consonantal gestures in a complex onset tend to be equally displaced in opposite directions in time from the vocalic gesture, as in Fig. 4C. Browman and Goldstein (2000) conceptualized this pattern as a result of pairwise coupling interactions between the virtual cycles of the gestures: consonantal gestures are anti-phase coupled to each other but coupled in-phase with the vowel; the C-center pattern arises from minimizing the energy of the corresponding relative phase potentials. Nam and Saltzman (2003) implemented a dynamical model of a system of three planning oscillators that produces this pattern.

To summarize, the coupled oscillators model of Articulatory Phonology relies solely on coordinative control to model articulatory timing in various syllable shapes. The concept of phasing of virtual cycles is one particularly powerful way of modeling coordinative control, although other models of precision control may also be empirically adequate. The key point is that unlike competitive control, coordinative control does not rely on sensory feedback, gating, or deselection; instead, coordinatively controlled articulatory gestures are co-selected, i.e. selected together, and phasing mechanisms govern the relative timing of their execution.

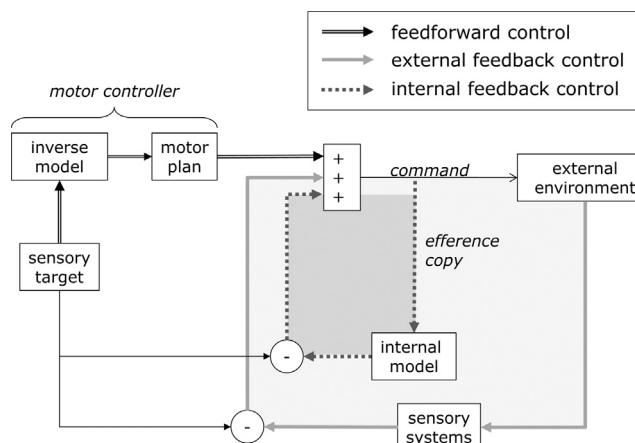


Fig. 5. Schematic illustration of feedforward control, external feedback control, and internal feedback control. The internal model maps an copy of outgoing motor commands to a sensory representation, which can be compared with a target to generate an error correction signal. For expository purposes different modalities of feedback are not distinguished.

2.3. Feedback internalization and development of coordinative control from competitive control

Selection-coordination theory hypothesizes that developmental transitions from competitive to coordinative control arise from a process of *feedback internalization*. Pursuing the piano analogy from this perspective, consider that children first learn to play songs note by note, which corresponds to competitive control; subsequently children learn to play chords and patterns with more precise control over timing, which corresponds to coordinative control. In general, competitive control developmentally precedes coordinative control, and the reason has to do with the role of feedback.

In models of competitive selection, feedback is necessary for suppression of motor plans. Grossberg (1978) cited two potential sources of suppression: inhibition from external feedback or recurrent inhibition, the latter of which can be associated with a different form of feedback called *internal feedback*. Indeed, in all models implementing competitive selection there is some mechanism akin to internal feedback that serves the purpose of regulating when plans are deselected, such as a “deactivation time” parameter (Rumelhart & Norman, 1982), “postselection negative feedback” (Dell, 1986), or a “a non-specific response suppression signal” (Bohland et al., 2010).

Fig. 5 schematizes relations between feedforward control, external feedback control, and internal feedback control. Feedforward control maps a sensory target to a motor plan via an inverse model, and the motor plan generates motor commands. These motor commands result in changes in the external environment that peripheral sensory organs can detect (e.g. generation of sound and changes in muscle tension). External feedback control compares this sensory information to the target, and can either adjust outgoing motor commands or suppress motor commands upon successful target achievement. However, external feedback is undesirably slow, since information regarding the sensory consequences of actions becomes available only after those actions have been performed.

Whereas external feedback represents the actual consequences of movements, internal feedback represents the *anticipated* sensory consequences of motor commands. Internal feedback control utilizes a correction signal derived from comparing the predicted consequences of motor commands to sensory targets. This prediction involves mapping a copy of the outgoing motor commands (“efference copy”) to a sensory representation through an internal, forward model. Hence internal feedback bypasses peripheral sensory systems, thereby making feedback available earlier. Much recent research has emphasized the role of internal feedback in motor control and motor learning (Desmurget & Grafton, 2003; Todorov & Jordan, 2002; Wolpert & Kawato, 1998; Wolpert, Ghahramani, & Flanagan, 2001).

Selection-coordination theory hypothesizes that coordinative control develops from competitive control through feedback internalization, i.e. through increasing reliance on internal feedback control. **Fig. 6** schematizes the progression from competitive to coordinative control between two gestures, X and Y, using activation variables modeled in Tilsen (2013) and incorporating the assumption that feedback internalization allows for anticipatory opening of gating variables. In prototypical competitive control (**Fig. 6A**), the gating variable for Y opens only after external feedback leads to the suppression and deselection of X. An intermediate degree of internalization (**Fig. 6B**) allows the opening of Y gating variables to occur before, i.e. in anticipation of, the suppression and deselection of X. More extensive internalization (**Fig. 6C**) is hypothesized to lead to the loss of competitive gating, resulting in co-selection of gestures and prototypical coordinative control. The loss of competitive gating can be motivated on the grounds that increased co-activation of movements plans (which is a direct consequence of internalization) opposes competitive interactions between plans.

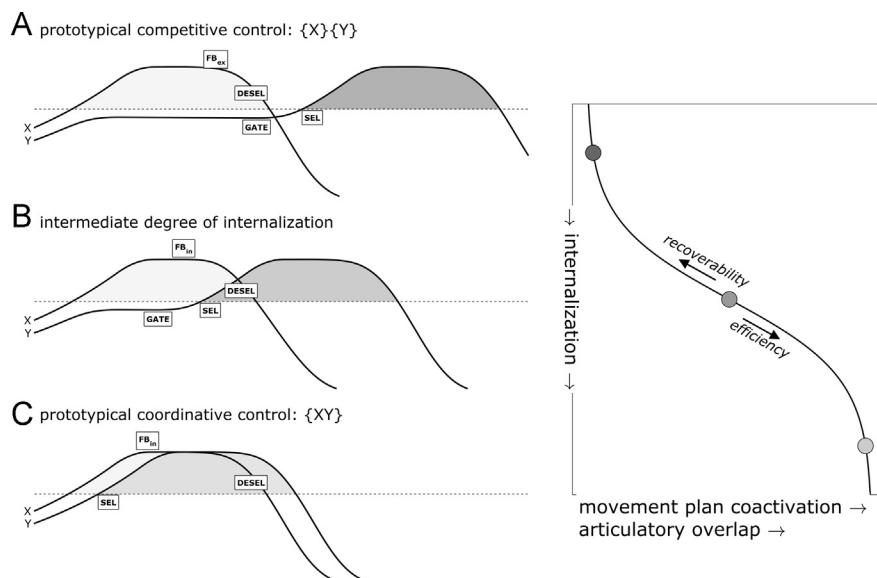


Fig. 6. Schematization of the consequences of feedback internalization. (A) Prototypical competitive control: plans X and Y are competitively selected, with external feedback governing selection and deselection. (B) An intermediate degree of internalization allows for anticipatory opening of the gating variable for Y, resulting in greater coactivation of movements plans and hence a greater degree of articulatory overlap. (C) With further internalization competitive gating is lost, allowing for prototypical coordinative control in which gestures are co-selected and coordinated.

Thus the development of coordinative control from competitive control depends on the degree of feedback internalization between gestures. In general we can conceptualize the degree of internalization as the result of competing forces, as represented in Fig. 6. On one hand, increased coactivation of plans and articulatory overlap allows for more rapid transmission of information; efficiency forces of this sort promote internalization of feedback. Internal models also make control more robust to contextually predictable perturbations, further promoting efficiency of information transmission. On the other hand, a high degree of overlap between gestures may threaten their perceptual recoverability (Chitoran, Goldstein, & Byrd, 2002); recoverability forces therefore disfavor internalization of feedback. From an alternative perspective, these forces can be reinterpreted as perceptual parsing, articulatory effort, and speech rate costs in a function that speakers optimize (Simko & Cummins, 2011). The relative strengths of forces/weights of costs will necessarily differ according to the nature of the biomechanical and aero-acoustic interactions between gestures; moreover, language- and context-specific modulation of gestural targets likely influences the strengths of these forces.

Because of the multitude of factors influencing internalization, selection-coordination theory emphasizes that competitive and coordinative control are *prototypical* regimes of control. In typical speech intermediate degrees of feedback internalization may be common. These intermediate degrees of internalization result from circumstances in which recoverability and efficiency forces are balanced such that control is neither prototypically competitive nor prototypically coordinative.

A more detailed discussion of the mechanisms involved in feedback internalization is beyond the scope of this paper, but several neural models and experimental studies implicate plasticity in connections within and between cerebellar and cortical areas to be responsible for learning internal feedback (Imamizu et al., 2000; Kawato, Furukawa, & Suzuki, 1987; Krakauer, Ghilardi, & Ghez, 1999; Thoroughman & Shadmehr, 1999). Investigations of behavioral and neural aspects of acquisition of automaticity in motor control (Doyon & Benali, 2005; Schmidt & Lee, 1999) further support the notion that internalization of feedback drives the development of coordinative control. In contrast, reliance on external feedback associated with competitive control may be involve cortical–basal ganglia circuits, which numerous models implicate in action selection (Doya, 2000; Gurney, Prescott, & Redgrave, 2001; Mink, 1996).

Further evidence in support of the internalization model involves progressions from competitive to coordinative control that occur in nonspeech domains. For example, when adults grip and lift an object, their gripping and lifting movements are initiated closely in time and overlap (Westling & Johansson, 1984). In contrast, children who are just learning to use a precision grip apply the gripping and lifting forces sequentially; the youngest children appear to wait for feedback that the gripping movement has achieved its target before selecting the lifting movement (Forssberg, 1999; Forssberg et al., 1992). A similar example involves reaching and grasping: adults form an appropriate hand/finger posture for the grasp during the reaching movement, but infants do not exhibit this anticipatory hand posturing, instead waiting until they have transported the hand to the object (Jeannerod, 1986; von Hofsten, 1979).

2.4. Co-selection sets: integrating competitive and coordinative control

Competitive and coordinative control play integrated roles in organizing the selection and execution of articulatory gestures. To represent how control is organized we introduce the concept of a *co-selection set*—a set of gestures that are selected together and coordinated. Selection-coordination theory posits a specific interaction between the mechanisms of selection and coordination: only co-selected gestures can be coordinatively controlled, or equivalently, competitively selected gestures cannot be coordinated. Thus the gestures in a given co-selection sets are competitively controlled relative to those in other co-selection sets, while within each set gestures are coordinatively controlled.

Fig. 7 shows one possible organization of control for the word *bandana*. Here the syllables /bæn/, /dæ/, and /na/ are mapped to co-selection sets {bæn}, {dæ}, and {na}; and the gestures in each set are coordinatively controlled. Note that words and utterances

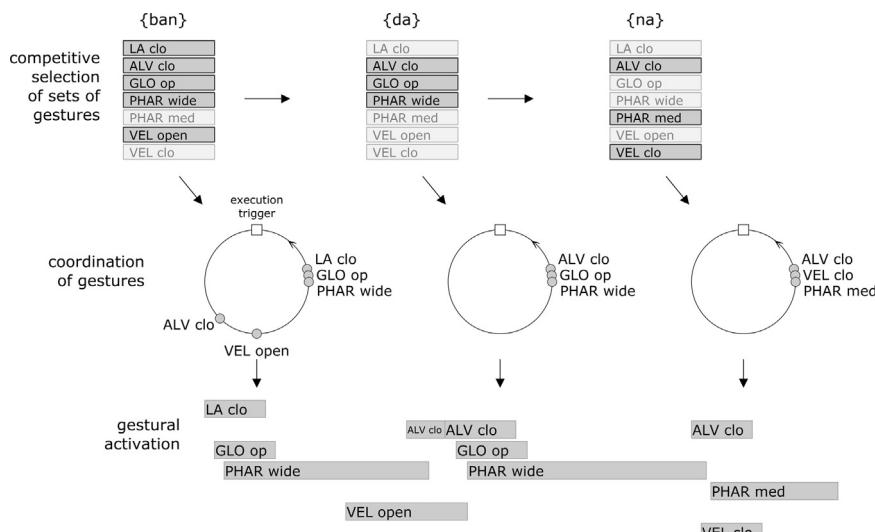


Fig. 7. Conceptualization of speech as a sequence of co-selection sets, using the word *bandana*. Gestures are grouped into co-selection sets. Within each set the gestures are coordinatively controlled, resulting in precise control over the relative timing of gestural activation in the gestural score.

may incorporate higher-level coordinative mechanisms that influence the selection of co-selection sets (cf. Tilsen, 2009a, 2009b, 2013), but our present concerns are restricted to the organization of control within the domain of syllables.

Of crucial importance in the current theory is the question of how gestures are organized into co-selection sets. In general we cannot assume a one-to-one mapping between co-selection sets and any particular linguistic unit (i.e. gestures, segments, moras, syllables). For example, the organization of control shown in Fig. 7 assumes that all of the gestures in each syllable are co-selected, i.e. {bæn}{dæ}{na}. However, an alternative possibility (not shown) is that the moraic units /bæ/ and /n/ are competitively selected, i.e. {bæ}{n}{dæ}{na}.

In addressing how control is organized, we emphasize that co-selection sets are representations of control *in a particular utterance*. Co-selection sets do not exist outside of utterances. When we refer to co-selection sets without indexing a particular utterance, we are either referring abstractly out of convenience or making a statistical generalization regarding how control tends to be organized in an ensemble of utterances.

This leads to a fundamental hypothesis of selection-coordination theory: the organization of control over articulatory gestures is dynamic, changing over the course of development and changing in response to task demands. Accordingly, linguistic units of the “segment”, “mora”, and “syllable” can be reconceptualized as differently-sized instantiations of a more general type of motor unit, the co-selection set. Segments, moras, and syllables may describe the dominant organization of gestural selection at some stage of development, or in some task-specific context, but only one of these units is associated with the organization of control in any particular utterance.

To set the stage for subsequent sections, Fig. 8 depicts several ways in which control can be organized in CVC and CCV syllables. Note that co-selection sets are denoted with combinations of “C” and “V” symbols in curly brackets, and that the “C” and “V” symbols do not refer to segments but rather to oral articulatory gestures associated with consonants and vowels—e.g. a bilabial closure for /p/ and a pharyngeal constriction gesture for /a/. Moreover, the order of symbols within a bracketed co-selection set does not indicate a sequential ordering of movements.

The purely competitive regimes, {C}{V}{C} and {C}{C}{V} (Fig. 8A), can be ruled out on empirical grounds because they predict no substantial overlap between a prevocalic consonantal gesture and vocalic gesture (here “prevocalic” is used to mean both preceding and adjacent, i.e. immediately prevocalic). In these models each oral articulatory gesture is competitively selected relative to the others. Note that purely competitive regimes correspond to the hierarchical selection models discussed earlier.

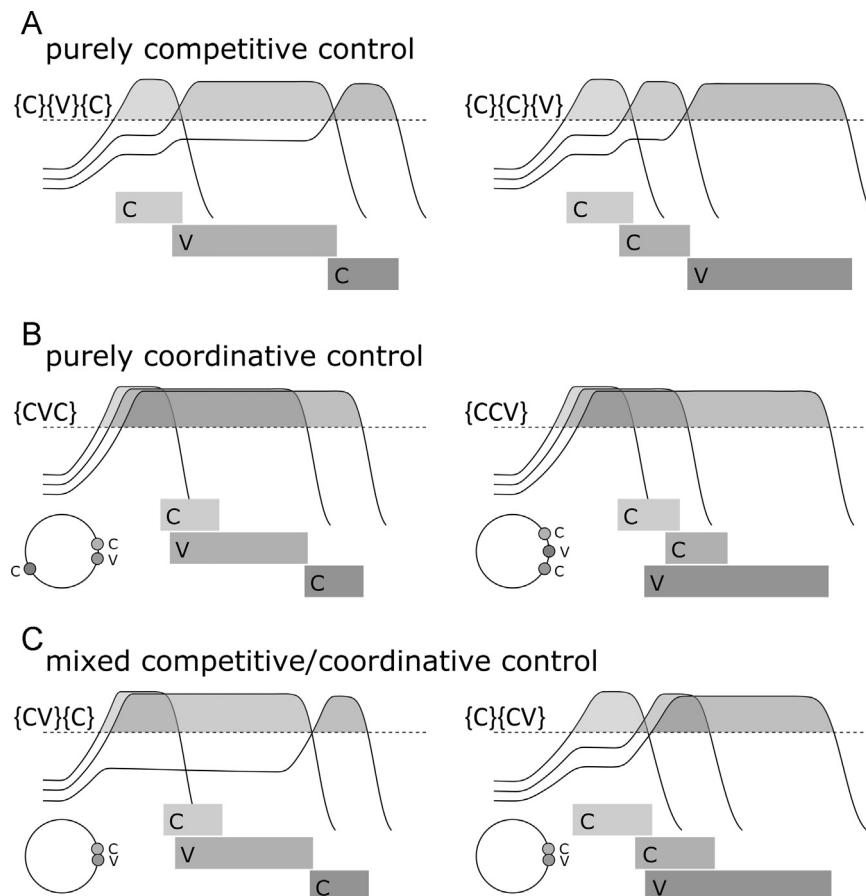


Fig. 8. Alternative control regimes for CVC and CCV syllables. (A) purely competitive control in which all gestures are competitively selected; (B) purely coordinative control in which all gestures are co-selected and coordinated; (C) mixed control: non-prevocalic consonantal gestures are competitively selected, prevocalic consonantal and vowel gestures are co-selected and coordinated.

The purely coordinative regimes, {CVC} and {CCV} (Fig. 8B), can adequately model empirical timing patterns associated with post-vocalic consonantal gestures and timing patterns in which a C-center effect occurs in complex onsets. These models utilize coordinative phasing mechanisms to govern the initiation of gestures and thereby allow for substantial gestural overlap, depending on how phasing is parameterized. Purely coordinative control corresponds to the hypotheses of articulatory phonology, which can be taken to imply that the gestures in each syllable are co-selected and coordinatively controlled.

The mixed competitive/coordinative control regimes, {CV}{C} and {C}{CV} (Fig. 8C), reflect organizations of control in which a prevocalic consonantal gesture and vocalic gesture are coordinatively controlled while other consonantal gestures are competitively controlled. The {CV}{C} model accounts for CV timing in the same way as the purely coordinative model, but provides an alternative interpretation of how VC timing is accomplished: the post-vocalic consonantal gesture is competitively selected relative to the preceding vocalic gesture. Mixed {CV}{C} and purely coordinative {CVC} models can generate similar timing patterns for post-vocalic consonants, but they do so with differently organized control. Thus alternative forms of evidence need to be considered to resolve between them. In contrast, the mixed {C}{CV} and purely coordinative {CCV} models generate different timing patterns: the former generates no C-center effect.

Which control regime—purely coordinative or mixed competitive/coordinative—best accounts for empirical patterns in CVC, CCV, and other syllable shapes? In the following sections we will see that mixed control regimes better describe the organization of control in early development, and that a variety of cross-linguistic differences in phonological patterns can be associated with one or the other of these regimes.

3. Developmental transitions from competitive to coordinative control

3.1. Overview of hypothesized transitions

Selection-coordination theory holds that transitions occur in early development in which articulatory gestures that were previously competitively controlled become coordinatively controlled. This hypothesis is not entirely novel, as Kent (1983) observed: “mastery of sequencing is perhaps logically prior to mastery of phasing...the motor patterns in children’s speech tend to have a more segmental or discrete character than the motor patterns of adults. Phasing is thus developmentally later than sequencing...” (1983: 73). The hypothesis can also be related to the more general phenomena of motor chunking and automaticity (Logan, 1985; Miller, 1956; Posner & Snyder, 1975): the “chunks” are co-selection sets, control over execution is automatized within a co-selection set, and the sets may become larger in the course of development.

This section focuses on evidence for competitive-to-coordinative control transitions in early development, beginning with the early word stage. Treatment of babbling is deferred for subsequent work and it will be assumed that children enter the early word stage having acquired {CV} co-selection sets (cf. Tilsen, 2014b). Furthermore, because the transitions of interest are hypothesized to occur early in development, we omit discussion of the extensive body of work on the prolonged refinement of coordinative control that occurs through adolescence.

Fig. 9 schematizes the hypothesized developmental transitions. The reader should keep in mind that the use of generic “C” and “V” symbols does not imply that transitions occur simultaneously for all consonantal or vocalic gestures in some configuration; rather, a large number of factors related to perceptual recoverability and communicative efficiency (cf. Section 2.3) induce gesture-specific variation in when competitive-to-coordinative control transitions occur.

Entering the early word stage (normally around 12 months), children only use coordinative control for a prevocalic consonantal gesture and vocalic gesture, i.e. {CV}. Children are hypothesized to use mixed control regimes such as {CV}{C} and {C}{CV} in this stage, and subsequent transitions to {CVC} and {CCV} coordinative regimes are hypothesized to occur in some languages. Support for these hypotheses is found in non-standard sequencing and coarticulatory patterns reviewed below.

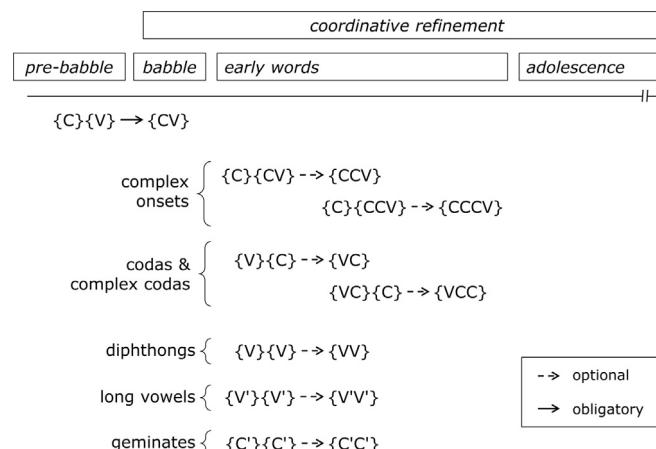


Fig. 9. Hypothesized developmental transitions in selection-coordination theory. C and V symbols represent generic consonantal and vocalic gestures respectively, curly brackets represent co-selection sets.

Table 1

Nonstandard sequencing patterns in singletons and clusters.

	singleton consonants		consonant clusters	
	onset	coda	onset	coda
omission	CVC → VC	VC → V	CCV → CV	VC ₁ C ₂ → VC ₁ VC ₁ C ₂ → VC ₂
substitution	CVC → C'VC	VC → VC'	CCV → CC'V	VCC → VC'C VCC → VCC'
comp. lengthening	*	VC → V:	? C ₁ C ₂ V → C ₂ V	VC ₁ C ₂ → VC ₁ : VC ₁ C ₂ → VC ₂ :
V epenthesis	*	C ₁ VC ₂ → C ₁ V.C ₂ V	CCV → CVCV	? VCC → VCVC
reduplication	*	C ₁ VC ₂ → C ₁ V.C ₁ V		
C lengthening	*	VC → VC:		
V truncation	*	V:C → VC V ₁ V ₂ C → V ₁ C		
local metathesis			*	VC ₁ C ₂ → VC ₂ C ₁
trans-V metathesis		*		C ₁ C ₂ VC ₃ → C ₃ C ₁ VC ₂
fusion			C ₁ C ₂ V → C ₁₂ V	? VC ₁ C ₂ → VC ₁₂

*: pattern is not attested, ?: pattern expected but not identified in review, shaded cells: pattern not applicable.

3.2. Non-standard sequencing patterns in early words

Children normally begin producing isolated CV and CVC words around 1 y.o. and CCV/VCC words around 2 y.o. (McLeod, Doorn, & Reed, 2001). During this period children are observed to produce many non-standard productions, which have been defined as “any systematic difference between the actual forms produced by a child and the (adult-like) forms the child is evidently attempting” (Rose & Inkelas, 2011). Here we focus on a subset of non-standard productions which involve sequencing of articulations, many of which are attributable to competitive control over non-prevocalic consonants.

Table 1 lists non-standard sequencing patterns identified in a review of literature on speech development. The first two—omission and substitution—are the most basic and can involve both prevocalic and non-prevocalic gestures. The remaining patterns exhibit asymmetries between prevocalic and non-prevocalic gestures. These include compensatory vowel 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), consonant lengthening/repetition (Camarata, 1989; Lebrun & Van Borsel, 1990), and vowel truncation (Bernhardt & Stemberger, 1998; Kehoe & Stoel-Gammon, 1997).

3.2.1. Omission with selection of a competitor

One common pattern that involves only non-prevocalic consonants is omission of the consonant with errorful reselection of a competing set of gestures. For example, when a coda is omitted the preceding gesture can be lengthened (compensatory lengthening) or both the preceding consonant and vowel can be repeated (reduplication), but neither of these patterns occur when a prevocalic consonant deletes.

Compensatory vowel lengthening can be understood to involve error reselection (cf. Fig. 3D) of a vocalic gesture at the expense of a post-vocalic one, i.e. {CV}{C} → {CV'}{V'}, which can occur if the vocalic gesture is not sufficiently suppressed and/or the coda consonantal gesture is not sufficiently active to outcompete the vocalic gesture for selection. Compensatory lengthening is also expected between codas in a cluster, i.e. {V}{C₁}{C₂} → {V}{C₁}{C₁}. Reduplication can be understood along the same lines: deficient activation of a coda gesture results in reselection of a preceding {CV} set at the expense of a post-vocalic {C}, i.e. {CV}{C} → {CV}{CV}. The difference between compensatory lengthening and reduplication amounts to whether only the preceding vocalic gesture is errorfully reselected (compensatory lengthening) or whether both a prevocalic consonantal and vocalic gesture are reselected (reduplication).

The crucial asymmetry to consider is that compensatory lengthening and reduplication patterns are not observed in association with the omission of prevocalic consonants. Loosely speaking, the patterns suggest that children know that a CVC word, for example, involves the selection of two sets of gestures, and that those sets are organized as {CV}{C} rather than {C}{VC}.

3.2.2. Hyperactive selection of consonant–vowel co-selection sets

Another common pattern involves the hyperactive selection of a {CV} co-selection set instead of {C}. In vowel epenthesis, e.g. {CV₁}{C} → {CV₁}{CV₂} and {C}{CV₁} → {CV₂}{CV₁}, a non-prevocalic consonantal gesture is correctly selected, but a vocalic gesture errorfully accompanies it. This can be understood to occur when the activation of a {CV} set outcompetes activation of the {C} set with

the same consonantal gesture. The hypothesized control asymmetry not only provides an intuitive interpretation of this pattern, but also correctly accounts for why vowel epenthesis of the form $CV_1 \rightarrow CV_2V_1$ do not occur: this would involve selecting a more complicated set $\{CV_2V_1\}$ over a more robustly learned one $\{CV\}$, or would require the selection of two sets, $\{CV_2\}\{V_1\}$, in place of one. Relatedly, the lengthening of a post-vocalic consonant can be viewed as the reselection of the consonantal gesture, in this case a consequence of a hyperactive representation.

3.2.3. Selection with omission of a competitor

Further evidence for competitive control of non-prevocalic gestures comes from productions in which selection of a coda gesture is accompanied by the truncation of a preceding long vowel or diphthong. In the early word stage, long vowels and diphthongs are hypothesized to involve two selection events: reselection in case of the long vowel, i.e. $\{V'\}\{V\}$, and selection of a second vocalic target in the case of the diphthong, i.e. $\{V_1\}\{V_2\}$. Hence the truncation pattern can be understood to arise from selection of a post-vocalic consonantal gesture at the expense of the selecting the second gesture of a diphthong $\{V_1\}\{V_2\}\{C\} \rightarrow \{V_1\}\{C\}$ or at the expense of reselecting the vocalic gesture, i.e. $\{V'\}\{V\}\{C\} \rightarrow \{V'\}\{C\}$.

Presumably co-exciting and maintaining an appropriate activation gradient is more difficult when there are more sets, and young children may have limited control over relative activation, restricting the number of elements they can correctly sequence. This accounts for the observations that coda consonants are acquired earlier and produced more frequently after short or lax vowels than after long vowels or diphthongs: $\{V\}\{V\}\{C\}$ targets involve three selection events, whereas $\{V\}\{C\}$ targets involve just two. Furthermore, selection of a prevocalic consonantal gesture is correctly predicted not to induce truncation in a following vowel, i.e. $\{CV_1\}\{V_2\} \rightarrow \{C\}\{V_2\}$ does not occur because C and V_1 are not associated with competing selection sets.

3.2.4. Non-standard ordering of selection and grouping of gestures

Metatheses provide another example of asymmetry between prevocalic and non-prevocalic gestures that can be explained with competitive control. Metatheses come in two varieties: the more common local metathesis involves transposition of two post-vocalic consonants (Bernhardt & Stemberger, 1998; Fikkert, 1994; Kirk & Demuth, 2005), i.e. $\{V\}\{C_1\}\{C_2\} \rightarrow \{V\}\{C_2\}\{C_1\}$; the relatively rare transvowel metathesis pattern (Jaeger, 2005; Menn, 1976) always involves transposition of a post-vocalic consonant with a non-prevocalic onset consonant, i.e. $\{C_1\}\{C_2V\} \rightarrow \{C_2V\}\{C_1\}$.

These patterns can be understood as the result of non-canonical activation gradients between sets, resulting in non-canonical ordering of selections (cf. Fig. 3C). The hypothesized difference in control over prevocalic and non-prevocalic gestures also accounts for why metatheses $\{C_1\}\{C_2V\} \rightarrow \{C_1\}\{V\}\{C_2\}$ and $\{C_1\}\{C_2V\} \rightarrow \{C_2\}\{C_1V\}$ are not observed—these would involve transpositions of parts of co-selection sets, which may be less likely than transpositions of entire sets. An explanation is warranted for why certain transpositions predicted to occur in three-set structures, such as $\{V\}\{C_1\}\{C_2\} \rightarrow \{C_1\}\{V\}\{C_2\}$ or $\{V\}\{C_1\}\{C_2\} \rightarrow \{C_2\}\{V\}\{C_1\}$, were not identified in the review. One possibility is sets containing a vocalic gesture are relatively highly active, creating a bias against transpositions which delay the selection of a vocalic gesture.

Many of the above patterns suggest that co-selection sets—particularly $\{CV\}$ —exhibit a form of integrity which discourages dissociations of gestures within the same set. However, non-standard productions in which gestures subvert this expected organization are not disallowed, particularly when children are in the process of learning to coordinate gestures that were previously competitively controlled. For example, fusions such as $C_1C_2V \rightarrow C_{12}V$ occur in which an onset cluster is reduced to a single consonant that combines features of both segments. This can be viewed as the result of a merger of gestures from different sets into a single set, with errorful omission of some gestures or non-canonical timing. Fusions are the product of attempts to develop coordinative control, and hence are predicted to arise later in development than errors which reflect competitive control over the non-prevocalic consonant.

3.2.5. Principles of selection in early development

Early word stage asymmetries in non-standard sequencing patterns are well explained by the hypothesis that children coordinatively control prevocalic consonants but competitively control non-prevocalic consonants, i.e. $\{C\}\{CV\}\{C\}$. This leads to the question of whether one can identify general principles that govern production in this stage.

Some of the patterns described above have been attributed to a highly ranked constraint to adapt productions to a bimoraic template (Demuth, 1995; Demuth & Fee, 1995; Levelt & Van de Vijver, 2004; Levelt, Schiller, & Levelt, 2000). Such accounts are contradicted by observations of short vowel CV productions and by compensatory lengthening for omitted codas in both CVC and CVVC syllables (Song & Demuth, 2008). Selection-coordination theory offers a more mechanistic interpretation without imposing overly restrictive templatic constraints on production; a cognitively motivated mechanism—competitive selection—can account for the apparent influence of templates.

Several useful generalizations arise in considering the non-standard sequencing patterns reviewed above. First, many of the asymmetries serve to respect the integrity of co-selection sets, since patterns which would decouple gestures in the same co-selection set, e.g. $\{CV\} \rightarrow \{C\}\{V\}$, do not seem to occur. Second, some errors appear to involve hyperactivation of $\{CV\}$ selection sets, as in vowel epenthesis $\{CV_1\}\{CV_2\}$. Hyperactivation can also explain the compensatory lengthening and reduplication patterns: in these cases gestures from an initial co-selection set may be insufficiently suppressed on account of abnormal activation. Third, difficulty of production in this stage appears to depend on the number of co-selection sets in a word: competitively selecting $n+1$ sets is more difficult than competitively selecting n sets.

Because competitive to coordinative control transitions are hypothesized to arise from internalization of feedback, gesture-specific factors involving biomechanics, perception, and articulatory dynamics likely play an important role in governing when in development these transitions occur. In the general case development of coordinative control over a specific combination of gestures does not entail development of coordinative control over the entire class, e.g. {a}{p} → {ap} does not entail {V}{C} → {VC}. However, gesture-specific transitions likely precipitate more generic restructuring of control. There probably also exist implicational generalizations regarding the timing of more complex coordinative developments relative to simpler ones, i.e. {C}{CV} → {CCV} likely precedes {C}{C} → {CV} → {CCV}. The extent to which these predictions are borne out is an empirical issue which only frequently sampled longitudinal acoustic data can answer.

3.3. Phonetic evidence of control transitions in development

Over the course of development generic trends of decreased segmental duration and more extensive coarticulation are observed. Both of these presumably arise from increased overlap of articulations, which is a predicted consequence of transitions from competitive to coordinative control. For example, as feedback internalization promotes a {V}{C} → {VC} transition, the coda gesture is selected earlier relative to the vocalic one and the two overlap to a greater degree (cf. Fig. 6). This causes increased gestural overlap which results in shortened vowel duration and in greater acoustic manifestations of coarticulation.

Developmental changes in segmental duration and coarticulation (Kent, 1976; Kent & Forner, 1980; Smith, 1992; Smith & Goffman, 1998; Smith, Kenney, & Hussain, 1996; Vihman & Velleman, 2000; Weismer, 1984) are often supposed to arise from refinement of control that occurs gradually throughout development. In contrast, selection-coordination theory emphasizes the importance of more abrupt changes in how children organize control, with patterns asymmetrically involving prevocalic and non-prevocalic gestures being particularly relevant for understanding such changes.

Several studies of coarticulation have shown that adult-like CV coarticulation emerges earlier in development than adult-like VC coarticulation. CV coarticulation is similar to adults in children 3 years old or younger (Katz & Bharadwaj, 2001; Repp, 1986; Sussman, Duder, Dalston, & Caciato, 1999; Turnbaugh, Hoffman, Daniloff, & Absher, 1985), and some studies have even found that younger children hyper-coarticulate in CV (Goodell & Studdert-Kennedy, 1993; Nittrouer, Studdert-Kennedy, & McGowan, 1989; Nittrouer, Studdert-Kennedy, & Neely, 1996). VC coarticulation, in contrast, develops later: Kent (1983) found little evidence for anticipatory coarticulation in some VC syllables in 4-year-old children, Hawkins (1984) observed non-adult-like timing of VC articulations in 3 and 4 y.o. children, and Goodell and Studdert-Kennedy (1993) observed a greater degree of CV coarticulation than VC coarticulation in eCV sequences. The hypothesis that non-prevocalic consonants are competitively controlled in early development provides a straightforward understanding of this coarticulatory asymmetry.

Selection-coordination theory can thereby answer two fundamental questions regarding onset/coda asymmetry, which are (1) why does coordinative control develop earlier for CV than VC, and (2) why in adult {CVC} forms are onset gestures in-phase coordinated while coda gestures are anti-phase coordinated?

The developmental difference can be seen as a consequence of a temporal asymmetry in the relation between movement initiation and movement feedback. External sensory feedback regarding target achievement is temporally asymmetric, becoming available when a movement has reached its target, i.e. near the end of the movement rather than the beginning. Feedback regarding the attainment of a vocalic target is hence more susceptible to interference from a post-vocalic consonantal gesture than from a pre-vocalic one: a pre-vocalic consonantal gesture interferes with the acoustic consequences of vocalic movement initiation, while a post-vocalic gesture interferes with the acoustic consequences of vocalic target achievement. Thus the developmental delay of VC co-selection may be driven by the relative importance of feedback regarding target achievement compared to feedback regarding movement initiation.

The answer to the second question, regarding the association of onset/coda timing with in-phase/anti-phase coordination, follows from constraints on feedback internalization described in Section 2.3. Coda consonantal gestures do not become in-phase coordinated with a preceding vocalic gesture because this would compromise the perceptual recoverability of the vocalic gesture. This idea can be interpreted in an optimization framework (Simko & Cummins, 2011) if the perceptual costs of such perturbations are allowed to change over the course of development, perhaps due to the development of perceptual models which better compensate for contextual perturbations.

Another observation in support of competitive control involves a common developmental pattern associated with long vowels, geminate consonants, and diphthongs. Early in development the durations of long vowels are variable but ultimately more representative of short segments; children subsequently progress through a transitional stage in which vowel length is only accurately produced in the absence of a post-vocalic consonant (Fikkert, 1994; Kehoe & Lleó, 2003). Eventually the length contrast is correctly produced but with atypically long durations that may subsequently shorten (Kehoe & Lleó, 2003; Naeser, 1970; Raphael, Dorman, & Geffner, 1980). Similar progressions can be inferred from studies on diphthongs (Kehoe & Lleó, 2003; Stokes & Wong, 2002) and geminates (Hurme & Sonninen, 1985; Kunnari, Nakai, & Vihman, 2001; Vihman & Velleman, 2000).

The observation that long segments lengthen before they shorten can be readily understood in the current framework. At first children produce long targets with short duration because they have not acquired the ability to intentionally reselect a vocalic or consonantal gesture (cf. Fig. 3E). When children do master intentional reselection, the long-to-short duration ratio will be large, reflecting more prototypical competitive control in which reselection requires external feedback. As feedback is internalized, reselection can be anticipated, resulting in a greater degree of overlap between selection periods and hence a shortening of duration.

In sum, phonetic patterns such as onset/coda coarticulatory asymmetry, and lengthening and shortening of long segments/diphthongs, can be readily understood within selection-coordination theory. In combination with evidence from non-standard sequencing patterns, these phenomena support the hypothesis that children employ mixed control regimes such as {C}{CV}{C} early in development.

4. Reinterpretation of segmental structure

Hierarchical production models assume that segments are a functional level of motor organization, but the available evidence supports an alternative interpretation in which the units typically called segments are a diverse group of co-selection sets. This section discusses several ways in which the assumption of segmental organization is problematic: (1) segmental organization is inconsistent with formal phonological representations, where multiple levels of structure are necessary to uniquely characterize segments; (2) the entities classified as segments are fairly diverse from a developmental and motoric perspective; and (3) the primary argument for segmental organization—the occurrence of segmental sequencing errors—is unconvincing. The selection-coordination framework provides a more general understanding of the nature of these units and does not suffer the problems associated with the assumption that segments organize articulatory movements.

4.1. Inconsistency between phonological representations and production models

The assumption of segmental organization made by most production models is not consistent with formal phonological representations. Such representations treat segment length contrasts by associating a segment-root node with one or more timing slots, either autosegments or moras (cf. the long vowel in Fig. 1), and represent segment contourality by associating two root nodes with a single timing slot (cf. the affricate in Fig. 1). To uniquely characterize the difference between a long vowel and a short vowel, two levels of representation are required: a root node level and timing slot level. The same holds for the distinction between a diphthong and a sequence of vowels. Hence there is no single level on which segments are uniquely represented. Problematically, hierarchical production models treat phonetic differences between long segments, short segments, and contour segments as parameters of articulatory implementation, rather than manifestations of structural differences.

The problem with ignoring structural organization is that the models are unable to account for a variety of phonological patterns. For example, in many languages a syllable containing a long vowel influences the stress pattern of a word, whereas a syllable containing a short vowel does not. Explanations for these sorts of phenomena do not follow straightforwardly from production models which assume segmental organization. The non-hierarchical model of articulatory phonology suffers from the same problem: in the absence of a mechanism for organizing gestures, various phonological patterns cannot be readily explained.

4.2. Non-uniformity of segments from a developmental and motoric perspective

Another problem with the assumption of segmental organization is that, from developmental and motoric perspectives, segments are a diverse group of entities. If all segments were essentially the same type of unit, we would expect them to be similar in certain regards, but this is not the case. For expository purposes, Fig. 10 groups segments according to two factors: (1) when children tend to acquire adult-like production of the segment, and (2) the degree of asynchrony between articulations that comprise the segment.

Segments in group A (simplex vowels, plain stops, glides, and nasals) tend to be acquired early and involve highly synchronous articulations. For example, in plain stops the oral occlusion and glottal abduction gestures are initiated closely in time (Hoole, 2006; Löfqvist & Yoshioka, 1984). Velar opening and oral closure are highly synchronous in prevocalic nasals (Byrd, Tobin, Bresch, &

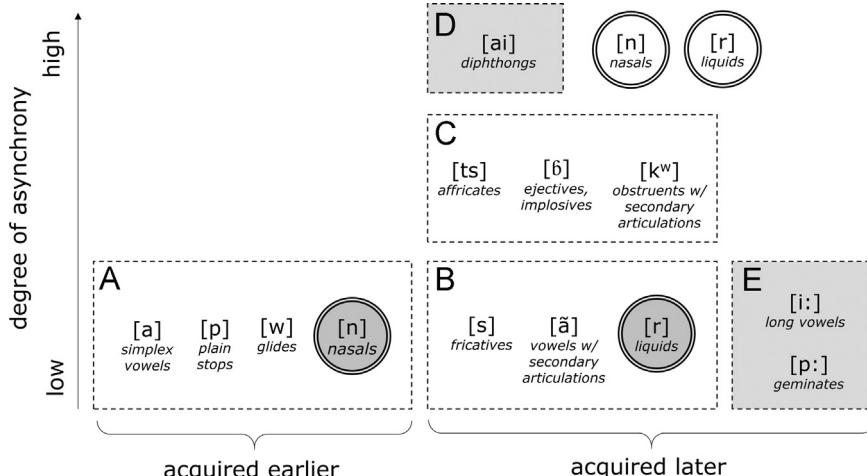


Fig. 10. Grouping of segments according to relative timing of acquisition and degree of asynchrony among component articulations. Nasals and liquids are highlighted because of their propensity to decohere post-vocally. Long segments and diphthongs are shaded because of their similarity with regard to control and phonological patterning.

Narayanan, 2009; Krakow, 1989), as are lip rounding and tongue body retraction in labiovelar glides (Gick, 2003). The segments in this group appear frequently in transcriptions of babbling and are typologically very common.

Segments in group B (fricatives, liquids, vowels with secondary features) also exhibit highly synchronous timing and yet are acquired subsequent to babbling. The relatively late acquisition of these segments may be attributable to the incorporation of a greater number of oral gestures or finer control requirements on individual effectors. For example, some fricatives such as /s/ and /ʃ/ involve labial rounding or retraction gestures in addition to a precisely controlled lingual constriction for producing turbulent airflow. Liquids and rhotics involve highly synchronous tongue blade and dorsum gestures in prevocalic environments (Browman & Goldstein, 1995; Gick, Campbell, Oh, & Tamburri-Watt, 2006; Proctor, 2009).

Segments in group C (affricates, ejectives, implosives, and consonants with secondary articulations) are acquired relatively late and involve moderately asynchronous gestures. The asynchrony normally results in salient acoustic effects. For example, affricates involve a fully occluded constriction gesture followed by a narrow constriction gesture that generates turbulent airflow. These two gestures cannot be highly synchronous, otherwise they would interfere mechanically and some aspect of the acoustic pattern would not be perceptually recoverable. Consonants with secondary articulations (e.g. labialized, palatalized, or pre-nasalized stops, ejectives, and implosives) probably also belong to this class, although more articulatory and developmental data are needed to confirm this.

Segments in group D (diphthongs) are quite different from other segments in that they involve a relatively high degree of asynchrony between their component articulations. This high degree of asynchrony is associated with competitive control over the vocalic gestures in early development. Segments in group E (long vowels, geminates) are likewise acquired relatively late, and this is also associated with competitive control in development; however, in the case of long segments the control involves intentional reselection of a gesture. The similarity in phonological and developmental patterning can ultimately be understood with regard to the similarity in controlled—group D and E segments involve two selection events, unlike most other segments. In Section 5 these relations between control organization and phonological patterning are considered in detail.

A further complication involves nasals and liquids, where the degree of articulatory synchrony can vary depending on whether these segments occur pre-vocally or post-vocally. In English nasal codas, velum lowering begins well before the associated oral closure gesture and reaches its target prior to the completion of the oral gesture (Byrd et al., 2009; Krakow, 1989). Similar patterns are observed for laterals, where the tongue tip constriction gesture and tongue dorsum constriction gesture are asynchronous in codas (Browman & Goldstein, 1995; Proctor, 2009; Sproat & Fujimura, 1993), and for rhotics, which involve a pharyngeal constriction gesture that precedes a tongue tip retroflexion or bunching gesture (Gick et al., 2006; Proctor, 2009). These patterns constitute a “decoherence” of gestures: articulations that in other contexts are highly synchronous are much less so post-vocally. Notably, decoherence patterns are acquired relatively late in development: adult-like timing of velar lowering and oral constriction in post-vocalic nasals emerges between 3 and 6 years of age (Thompson & Hixon, 1978).

The observation that segments are quite dissimilar with regard to when they are acquired, degree of synchrony between their articulatory components, and propensity of their components to decohere, calls into question the assumption that speech production is necessarily organized around the production of segments.

4.3. Error patterns do not justify the assumption of segmental organization

The most commonly cited rationale for segmental organization comes from sequencing errors, such as substitution, omission, addition, and transposition of elements. Many errors appear to involve segment-sized sets of articulations, and this has been used to justify the role of segments in hierarchical production models. The reasoning is that if errors can involve segment-sized units, then such units must constitute a level of organization in the production system. The logic behind such arguments is flawed; a more straightforward interpretation is simply that errors can involve sets of gestures of variable size and composition.

Several studies have reported that errors which involve segments, i.e. single vowels or consonants, are the most common type of sequencing error (Fromkin, 1971; Levelt, 1993; Shattuck-Hufnagel, 1983). This observation has been widely used as a justification for segmental organization in production models, although some researchers have taken a more cautious perspective: “few unambiguous conclusions about the mechanisms of speech production can be drawn from speech error data” (Frisch & Wright, 2002: 139). Sequencing errors involving units of various sizes are shown in Table 2, in which examples were drawn from Fromkin (1971).

Table 2

Examples of sequencing errors affecting units of varying sizes.

	Domain	Target	Error
(a)	segment	cup of coffee	cu[f] of coffee
(b)	segment	split pea soup	[pl]it [sp]ea soup
(c)	gesture	pity the new teacher	[m]ity the [d]ew teacher
(d)	gesture	pedestrian	[t]e[b]estrian
(e)	cluster	pay scale	[sk]ay [p]ale
(f)	cluster	shelter from the storm	[st]elter from the <u>s</u> torm
(g)	onset-vowel	pussy cat	[kae]ssy [po]t
(h)	vowel-coda	a heap of junk	a h[unk] of j[ip]
(i)	syllable	tremendously	tremenly
(j)	foot	butterfly and caterpillar	butterpillar and caterfly

Careful analyses in fact do not support the assertion that the most common errors are associated with segments. First, any error identified as segmental can be reanalyzed as an error affecting multiple gestures or features (Fromkin, 1971). For example, the apparent transposition of the [s] in Table 2(b) can be viewed as the simultaneous transposition of all of the features/articulatory components of the [s], rather than transposition of a segment. Second, more detailed analyses have found that the majority of substitution errors involve a single feature and that only a small percentage of errors involve all features of a segment (MacKay, 1970; Nooteboom, 1973). For example, the more typical sort of error is shown in (d), where individual oral gestures have exchanged between a [p] and [d] but laryngeal gestures have not. Third, sequencing errors commonly involve larger sets of elements, including segments clusters (e, f), vowel-consonant combinations (g, h), and syllables and feet (i, j).

Ambiguity in the domains of errors and variation in the size of these domains beg the question of whether error patterns really do justify the assumption of segmental organization. Biases and limitations inherent to phonetic transcription (Frisch & Wright, 2002) further call into question the assumption that many errors involve segments. Justifying segmental organization on the basis of error patterns is ultimately tautological. The logic of the argument is that because there are errors affecting “segments”, there must exist a level of organization in motor planning that represents segments. The tautology arises from the presumption that the errors being observed involve *segments* in the first place. In other words, the segment is presupposed in the argument, rather than being independently justified. The occurrence of errors involving segment-sized sets of articulatory gestures ultimately does not entail that segments are a domain around which planning is necessarily organized.

4.4. Segments in selection-coordination theory

Given the absence of compelling arguments for segmental organization in production, a reconceptualization of the notion of the segment is in order. One possibility is to treat segments as “practical tools” (Brownman & Goldstein, 1990), as opposed to units in a cognitive representation. A more extreme perspective views the concept of the segment as the product a scientific culture which emphasizes discrete computation. Port and Leary (2005) suggest that the tendency of linguists to posit symbolic units in analyzing language reflects a cultural predisposition to understand language as discrete computation. Another culprit behind the assumption of segmental organization may be the use of alphabetic writing systems (Faber, 1992; Linell, 2005). Pre-literate children and illiterate adults are generally not aware of phonemic categories in their language (Goswami & Bryant, 1990; Lukatela, Carello, Shankweiler, & Liberman, 1995); and segment-based language games are attested only in languages with alphabetic writing systems (Bagemihl, 1989). Along these same lines, Ladefoged (2001) suggested that “consonants and vowels are largely figments of our good scientific imaginations” (2001: 170), and wrote of “the phonemic conspiracy” in linguistic description (Ladefoged, 1984), which refers to the notion that because we *can* use segmental units, we feel compelled to do so.

Selection-coordination theory offers an alternative to outright rejection of segments without requiring a commitment to the assumption of segmental organization. The units referred to as “segments” are viewed as sets of gestures which organize control *in some stage of development, but not necessarily in adult speech*. Consider early word {a}{s} and {s}{pa} utterances, where the {s} represents a co-selected lingual constriction gesture and glottal abduction gesture, and {pa} represents a co-selected bilabial closure, glottal abduction, pharyngeal constriction (for the low vowel [a]), and glottal adduction gestures. The non-prevocalic co-selection set {s} happens to correspond a “segment,” and so in this developmental stage the segment characterizes how control of gestures is organized in some phonological environments. Note that in other environments, such as {sa}, the same articulatory gestures of {s} are present but do not exclusively comprise a co-selection set. Furthermore, after transitions to coordinative control, i.e. {a}{s} → {as} or {s}{pa} → {spa}, gestures associated with the non-prevocalic {s} no longer exclusively comprise a co-selection set in that environment: control has been reorganized into a larger unit.

The integration of gestures into co-selection sets during development does not necessarily remove all traces of developmentally prior organization. To the contrary, sets of gestures which were formerly exclusive members of a co-selection set, after being integrated into a larger set, may be more strongly phase-coupled to each other than to other gestures. For example, the lingual constriction and glottal abduction gestures of post-vocalic {s} are exclusive members of a co-selection set in the {a}{s} competitive regime; after a transition to an {as} coordinative regime, the lingual constriction and glottal abduction gestures may remain more strongly coupled to one another than to the vocalic pharyngeal constriction gesture. Thus the developmentally more primitive competitive organization of control leaves traces in coordinative control, in the form of gradient variation in phase-coupling between co-selected gestures. The adult representation, therefore, is quasi-holographic in the sense that coordinative phasing relations can retain information regarding the developmental time-course of transitions from competitive to coordinative control.

Selection-coordination theory thus incorporates developmental time as an important dimension of phonological representations. Sound patterns in speech—particularly those which evidence a hierarchical structure of phonological units—can be reinterpreted as byproducts of how control is organized in early stages of development, and so the hierarchical representations of formal phonology should not be viewed as isomorphic to a motoric representation which governs how adults control speech. As we show below, this has important consequences for our understanding of phonological patterns which involve relations between segments and “higher-level” units such as moras and syllables.

5. Reinterpretation of moraic and syllabic structure

By conceptualizing segments, moras, and syllables as differently sized co-selection sets which may organize control during different periods of development, selection-coordination theory offers a markedly different perspective on hierarchical phonological

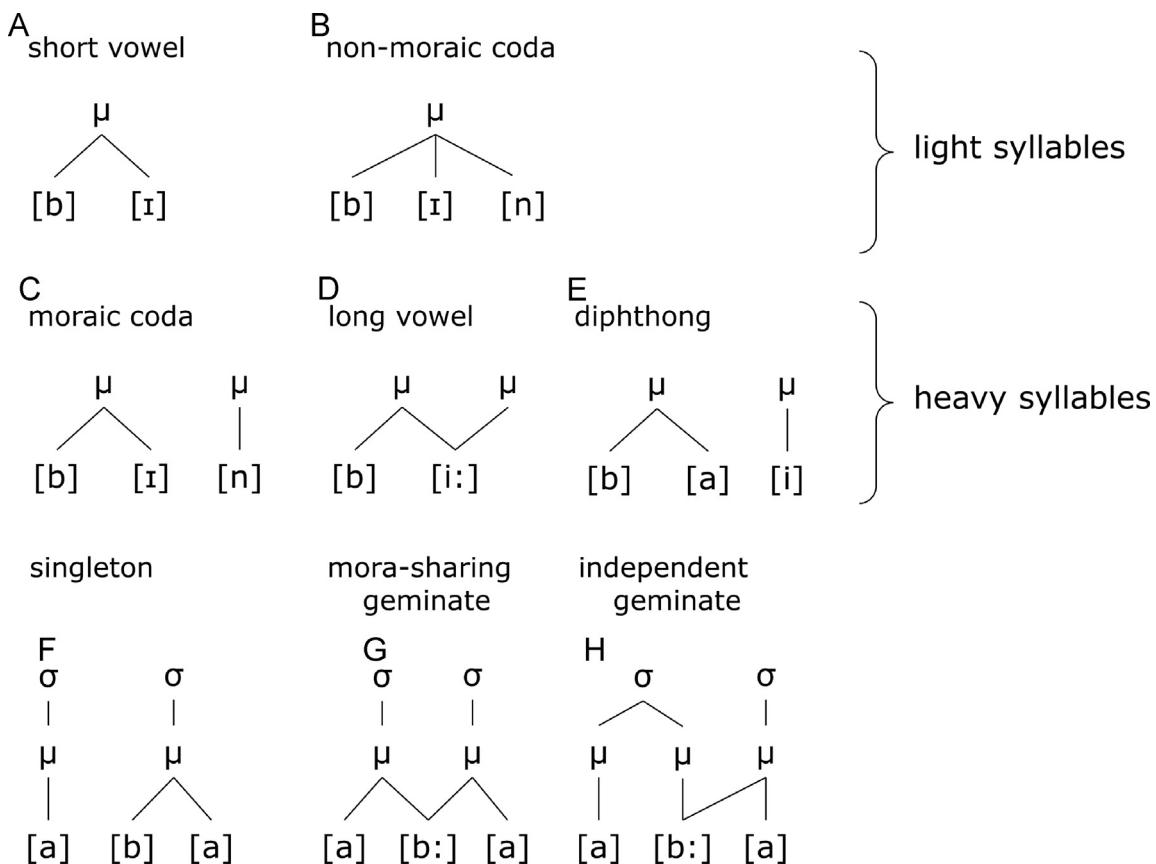


Fig. 11. Moraic representations. (A, B) light syllables have one mora; (C, D, E) heavy syllables have two moras; (F) intervocalic singleton consonant; (G) mora-sharing geminate; (H) independent geminate.

structure. The structural units which phonological analyses typically associate with a given utterance—particularly segments, moras, and syllables—do not simultaneously organize production in utterances, as far as selection is concerned. When an adult produces a given word with a {CVC} regime, control is organized into one set of co-selected gestures which corresponds to a syllable. Conversely, when a child in the early word stage produces that same word with a {CV}{C} regime, control is organized into two sets, and these happen to correspond to moras.

One consequence of this is that the notion that production involves the association of content to timing slots is misguided. Segments, moras, and syllables should not be expected to be isochronous and do not serve as temporal placeholders. The time-courses of selection and feedback-induced deselection in competitive control will differ from gesture to gesture, and hence from co-selection set to co-selection set. The phonological units are not occupiers of time independent of gestures; instead, the units represent current or prior organizations of gestures. They occupy time only by virtue of the fact that processes of selection and deselection take time.

Below we consider a variety of phonetic and phonological phenomena which involve moraic/syllabic organization. We show that these phenomena are a predicted consequence of variation in control regimes, which is attributable to the language-specificity or optionality of competitive-to-coordinative control transitions in development.

5.1. Cross-linguistic variation in coda moraicity

One important consequence of optionality in control regime transitions is cross-linguistic variation in coda moraicity, which follows from optionality of the transition {V}{C} → {VC}. Moras serve several descriptive functions in phonological representations (Hayes, 1989; Hyman, 1984, 1985; McCarthy & Prince, 1986), as exemplified in Fig. 11. One function is representing distinctions between short and long segments: a short vowel is associated with a single mora (A), a long vowel/diphthong is associated with two moras (D/E), an intervocalic singleton consonant shares a mora with the following vowel (F), and an intervocalic geminate either shares moras with flanking vowels (G) or is associated with an independent mora (H).

Another function of moras is to represent syllable weight, which serves to describe patterns of stress assignment in many languages (cf. Zec, 2003, 2007). Light syllables consist of one mora, heavy syllables two (cf. Fig. 11). A distinction is drawn between languages in which coda consonants are moraic (C), where syllables with codas are heavy and attract stress, and languages in which codas are non-moraic (B), where syllables with codas are “light” and do not attract stress.

There are two approaches to formal representation of onset consonants with regard to moras, one in which onset consonants link directly to moras (Fig. 11, cf. Hyman, 1984, 1985) and one in which onset consonants link to syllables, bypassing the moraic level

(not shown, cf. Hayes, 1989). The former is used here because it is more consistent with the selection-coordination model: onset consonantal gestures and vocalic gestures are typically co-selected, hence they should be directly associated with the same unit.

Phonetic evidence supports distinguishing between moraic and non-moraic codas (cf. Cohn, 2003). Broselow, Chen, and Huffman (1997) showed that in Malayalam, where phonological patterns indicate that codas are non-moraic, vowel durations are influenced by the presence of a coda; but in Hindi, where codas are moraic, vowel durations are not influenced by codas. Likewise, moraic-coda syllables in Mandarin Chinese are longer than non-moraic coda syllables in Shanghai Chinese (Duanmu, 1994). Moraic structural distinctions also have phonetic consequences for intervocalic geminates: Ham (2001) found that the geminate-to-singleton duration ratio is relatively small with the mora-sharing geminates in Madurese and Bernese (≈ 1.5), but relatively large with the moraically independent geminates in Levantine and Hungarian (≈ 2.0).

In the selection-coordination framework differences in coda moraicity and their phonetic consequences follow straightforwardly from the optionality of $\{V\}\{C\} \rightarrow \{VC\}$ transitions. Moraic codas are associated with a competitive $\{V\}\{C\}$ regime and non-moraic codas with a coordinative $\{VC\}$ regime. This predicts more gestural overlap in rimes with non-moraic codas and thereby accounts for the durational patterns cited above. Along these same lines the theory explains the cross-linguistic clustering of long segments into two groups, i.e. those with approximately twice the duration of singletons, and those which are substantially shorter than twice the singleton duration. In the former, control is more prototypically competitive, i.e. $\{V\}\{V\}$ and $\{C\}\{C\}$; in the latter, control is more prototypically coordinative, i.e. $\{VV\}$ and $\{CC\}$, with a relatively high degree of internalization.

Current approaches offer no satisfactory explanation for cross-linguistic variation in coda moraicity. In hierarchical production models there are no mechanisms for relating differences in structural representations to the observed differences in articulatory timing, short of incorporating ad-hoc parameters. The non-hierarchical model of articulatory phonology likewise has no direct mechanisms for modeling coda-moraicity distinctions, since it treats all post-vocalic consonantal gestures as coordinatively controlled.

5.2. Cross-linguistic variation in onset complexity

Another consequence of optionality in control transitions is variation in word-initial onset complexity, which follows from optionality of $\{C\}\{CV\} \rightarrow \{CCV\}$ transitions. Word-initial consonant clusters have been found to exhibit either a “complex onset” or “simplex onset” pattern. In the complex onset pattern all consonants in the cluster are tightly integrated with the following vowel: consonantal and vocalic gestures exhibit the C-center effect (Brownman & Goldstein, 1988; cf. Section 2.2), indicative of precise, coordinated control over timing. The C-center effect has been replicated in a number of studies (Byrd, 1995, 1996; Honorof & Brownman, 1995; Marin & Pouplier, 2010; Sproat & Fujimura, 1993), and has been observed in word-initial onsets in other languages, such as French (Kuhner, Hoole, & Mooshammer, 2006), Italian (Hermes, Grice, Mücke, & Niemann, 2008; Hermes, Mücke, & Grice, 2013), Georgian (Goldstein, Chitoran, & Selkirk, 2007), and Serbian (Tilsen et al., 2012).

In the simplex onset pattern only the prevocalic consonant is coordinated with following vowel; in this case no C-center effect is observed, and the relative timing of the prevocalic consonantal gesture and vocalic gesture does not vary in the presence of additional onset consonants. Simplex onsets have been observed in Tashylhiyt Berber, Moroccan Arabic, Slovak, and Hebrew (Goldstein et al., 2007; Pouplier & Beňuš, 2011; Shaw, Gafos, Hoole, & Zeroual, 2009, 2011; Tilsen et al., 2012), as well as in /s/-initial clusters in Italian (Hermes et al., 2008). Phonological treatments of syllabification often parallel the phonetic observations as in Tashylhiyt Berber (Dell & Elmedlaoui, 1985) and Italian (Davis, 1990; Kaye, 1992).

In selection-coordination theory differences in onset complexity follow directly from the optionality of $\{C\}\{CV\} \rightarrow \{CCV\}$ transitions. Complex onset languages are associated with a regime in which word-initial consonantal gestures are co-selected and coordinated, i.e. $\{CCV\}$, predicting the integrated control over timing observed in the C-center effect. In contrast, simplex onset languages are associated with the conservation of competitive control, i.e. $\{C\}\{CV\}$, predicting no C-center effect (recall that only co-selected gestures can be coordinated). Notably, the observation that Italian exhibits complex onset organization in all cases except for /sC/ clusters indicates that control regimes can be gesture-specific.

As with coda moraicity, hierarchical production models and articulatory phonology do not generate variation in onset complexity; control is too monolithically organized in hierarchical production models, articulatory phonology lacks the necessary mechanisms for organizing gestures, and neither provides a means of understanding the role of development in cross-linguistic variation.

5.3. Moraic phonological patterns and selection of pitch gestures

Within the selection-coordination framework, certain phonological patterns involving moraic structure—namely restrictions on contour tones and quantity sensitivity of stress—can be understood as the consequence of biases on the selection of pitch gestures. Contour tone restrictions involve lexical tones, and quantity sensitivity involves licensing of pitch accentuation by stress. Hence both patterns crucially involve the control of pitch.

A number of recent studies have concluded that lexical tones and intonational pitch accents are gestures. As such, lexical tones and intonational accents must be integrated into co-selection sets and coordinated with other gestures. Brownman and Goldstein (1989) originally suggested that lexical tones and pitch accents could be gestures on par with oral articulatory gestures, and a recent pitch accent imitation study indicates that speakers control precisely those parameters of pitch gestures which are controlled for oral articulatory gestures (Tilsen, Burgess, & Lantz, 2013). Mandarin lexical tone gestures have been observed to participate in a C-center effect with vocalic and consonantal gestures (Gao, 2008; Yi & Tilsen, 2015), and intonational pitch accent gestures have been found

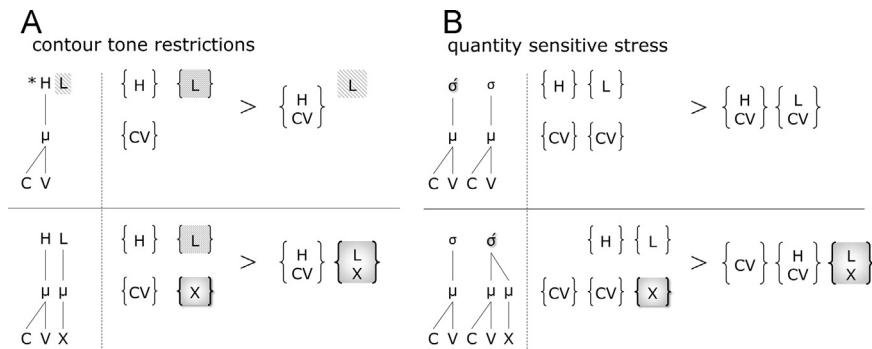


Fig. 12. Selection-coordination analysis of phonological patterns involving moraic structure. (A) Contour tone restrictions arise from difficulty in integrating a competitively controlled contour tone, {H}{L}, into a single co-selection set. (B) Quantity sensitivity arises from facilitated selection of competitively controlled bitonal pitch accent gestures in syllables associated with two co-selection sets. The asterisk “*” indicates that a structure is prohibited, and “X” represents a consonantal or vocalic gesture.

to be coordinated with vocalic gestures in German and Italian (Mucke, Nam, Hermes, & Goldstein, 2012; Niemann, Mücke, Nam, Goldstein, & Grice, 2011).

Furthermore, contour tones and bitonal/tritonal pitch accents, represented with two or more targets in the autosegmental-metrical approach (e.g. HL, LH), can be understood as multiple gestures. This suggests that competitive-to-coordinative control transitions may occur in development, such as {H}{L} → {HL}. Indeed, contour tones are acquired later than simple ones (Li & Thompson, 1977; Tse, 1978); this parallels the later acquisition of diphthongs, which have a similar structure and potential for control transition. Given the conceptualization of tones and accents as gestures, a mechanistic re-interpretation of the phonological patterns becomes possible.

In some languages contour tones are restricted to occur only in bimoraic syllables (Gordon, 2002; Zhang, 2004). As schematized in Fig. 12A, Hyman (1985) proposed that individual tones associate with moras, and hence restrictions on contour tones follow from the availability of moras for tones to associate with. This descriptively adequate generalization can be augmented with a mechanistic understanding in the selection-coordination framework. The pattern results from a bias involving the need to integrate each lexical tone gesture with a different co-selection set. This bias applies in a developmental stage in which the components of the contour tone are competitively controlled, i.e. {H}{L}, and so integrating both into the same set is not an option. As shown in Fig. 12A, when a syllable lacks two sets of oral articulatory gestures, the second component of the contour tone will not be integrated into the production, hence giving rise to the phonological restriction.

Many languages have quantity-sensitive stress systems in which syllable weight (i.e. mora count) influences the location of stress relative to word boundaries. Fig. 12B exemplifies this pattern in a language with default penultimate stress. In this example, stress appears to be “attracted” to a bimoraic word-final syllable. Standard accounts propose constraints on the alignment of stress in words or rules for locating stress on the rightmost (or in other cases leftmost) heavy syllable. Though descriptively adequate, such accounts fail to provide explanations for such patterns.

Stress should be understood as a structural configuration which enables pitch accentuation, rather than as a phonetic feature with articulatory/perceptual content (Liberman & Prince, 1977; Selkirk, 1980). Increased duration, intensity, and pitch excursions are often erroneously cited as “phonetic correlates” of stress. Yet these phonetic correlates are indirect and language-specific (Hayes, 1995), and in the absence of pitch accents, stressed syllables have not been associated with additional duration or intensity (Campbell & Beckman, 1997). Crucially, the one thing that stressed syllables clearly do have in common is a *potential* for intonational pitch accents to associate with them. Indeed, three of the four diagnostics Hayes (1995) proposed for identifying a stressed syllable involve a potential to bear an intonational pitch accent.

Given the above understanding of stress and the gestural conceptualization of intonational pitch accents, selection-coordination theory accounts for quantity sensitivity as a consequence of facilitated selection of pitch gestures. As illustrated in Fig. 12B, the presence of a word-final syllable with two selection sets allows the {H} and {L} components of a bitonal pitch accent gesture to be co-selected with the penultimate and final sets, respectively. In other words, heavy syllables “attract” stress because two competitively selected pitch gestures can be integrated into two distinct co-selection sets. Although not all pitch accents are bitonal, the potentiation of bitonal accents may result in analogization of the pattern to monotonal accents.

Similarly, the absence of an influence on stress when a word-final consonant is extraprosodic can also be readily understood in this framework. Extraprosodicity is a formal representational device for describing the failure of word-final consonants to influence stress (Hayes, 1980, 1982; Ross, 1969). Here extraprosodicity is viewed as a situation in which the post-vocalic consonantal gesture is co-selected with the preceding vocalic gesture, predicting no facilitation of pitch gesture selection and thus the default penultimate pattern.

Auditory/acoustic explanations for contour tone restrictions and quantity sensitivity have been proposed, but these are inconsistent with the empirical data in several ways. Gordon (2002) argued that contour tone restrictions arise because longer and louder intervals of voicing promote the perception of tonal information. Hence long vowels, vowel-sonorant rimes, and vowel-obstruent/short vowel rimes provide decreasingly suitable environments for perception of F0 changes. The cross-linguistic survey of tone languages in Gordon (2002) found some evidence for an implicational hierarchy of restrictions along these lines, but also found numerous cases not explainable from a perceptual perspective: these include languages in which contour tones are restricted in VO

but not V rimes (e.g. Cantonese, Vietnamese, Maru), or are restricted in V but not VO rimes (e.g. Hausa, Musey, and Luganda). More problematically, the perceptual account cannot explain why sonorant onset consonants fail to influence tone-bearing capacity.

A related account of quantity sensitivity was proposed in Gordon (2006). This perceptual account associates weight with “total perceptual energy,” which Gordon defined as the integration of loudness over time. Gordon argued that heavy syllables have more perceptual energy, and that this is responsible for the attraction of stress to these syllables. As before, this account predicts that onset consonants should contribute to syllable weight, since onset consonants also contribute loudness to a syllable. Yet onset consonants rarely influence stress patterns, and hence the perceptibility account is problematic without an ad-hoc stipulation that onsets are ignored.

One advantage of understanding contour tone restrictions and quantity sensitivity in the selection-coordination framework is that onset/coda asymmetries can be readily explained. Onset consonantal gestures do not constitute an independent co-selection set during the early word stage in development and hence do not have the potential to influence selection of lexical or intonational pitch gestures. Another advantage is that the underlying similarities between mora-related phonological patterns can be seen more clearly: such patterns arise from the need to integrate competitively selected pitch gestures with different sets of oral articulatory gestures. While moras are descriptively adequate for the patterns considered above, selection coordination theory provides a motorically grounded explanation for these patterns.

5.4. Dissociation of adult control from phonological patterning

Phonological patterns may emerge in association with regimes which describe control in an early period of development, but which do not necessarily characterize adult control. Accordingly, the presence of a phonological pattern associated with competitive control does not entail that adult speakers organize control as such. However, allowing for a dissociation between adult control and phonological patterning begs the question of how phonological patterns associated with competitive control in some phonological environment could be productive for adult speakers who use coordinative control in that environment.

Adults may retain the ability to use competitive control in phonological contexts where they typically use coordinative control, in which case analogical mechanisms may productively extend competitive regimes to new forms. Previous research has shown that when producing unfamiliar words or morphological constructions, adult speakers analogize from familiar forms to new ones on the basis of similarity, frequency, and probability (Albright & Hayes, 2003; Frisch, 1997; Gentner & Markman, 1997; Pinker & Prince, 1988). If control regimes are susceptible to analogization, there should be evidence that adults can employ either competitive or coordinative control in some phonological environments.

There are indeed several forms of evidence that speakers do retain competitive control as an option for organizing selection, even after coordinative control becomes the dominant strategy. Tilsen, Cohn, and Ricciardi (2014) observed variability in the control of liquid codas in a study which elicited syllable-count judgments and productions of English words with long vowel/diphthong nuclei and liquid rimes (e.g. pool, peel, pail, pile, pear, pyre). These forms are unusual because intuitions about the number of syllables in such words vary across speakers (Cohn, 2003; Lavoie & Cohn, 1999). Tilsen et al. (2014) found acoustic evidence for a greater degree of coarticulation between vocalic and liquid-dorsal gestures 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 coordinative control regime in which liquid and vocalic gestures are co-selected, while disyllabic judgments are associated with competitive control over the liquid gesture.

Comparisons of articulatory timing between experimental paradigms also indicate that speakers can engage competitive or coordinative control. For example, abrupt changes from VC- to CV-like articulatory timing (i.e. VC → CV) are observed in syllable repetition tasks as rate is increased (Sato et al., 2006; Stetson, 1951; Tuller & Kelso, 1990). This phenomenon seems to provide evidence in support of the Articulatory Phonology hypothesis that coda consonantal gestures are anti-phase coordinated with a preceding vocalic gesture. However, the repetition task imposes a periodic rhythm on selection processes, and this may bias speakers to use coordinative control. Regularity of metrical structure has indeed been associated with decreased segmental durations (Tilsen, 2011), suggestive of greater gestural overlap and reliance on coordinative control. Notably, when speakers entrain the production of a short phrase to a two-tone metronome pattern, those metrical patterns which impose simpler rhythmic relations between feet and phrases have been found to result in diminished variability in articulatory timing of an /sp/ onset cluster (Tilsen, 2009a). Thus tasks with entrainment to periodic rhythms appear to promote the use of coordinative control.

In contrast, articulatory and auditory feedback perturbation studies provide examples of task-induced competitive control. Munhall, Löfqvist, and Kelso (1994) found that mechanical perturbations of the lower lip during a bilabial closure in /p/ resulted in delayed and lengthened laryngeal abduction. A coordinative phasing model does not predict such effects because feedback does not modulate relative timing. The effects can be interpreted as the consequence of an adaptation to the perturbation task in which speakers engage competitive control over the labial and laryngeal gestures. More generally, auditory feedback perturbation studies (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) demonstrate that speakers can adopt a mode of production in which feedback governs the deselection of movement plans: segments produced in the context of auditory perturbations are typically prolonged.

Taking a different approach, Tilsen (2014a) developed a cued-insertion task to perturb the selection of gestures. The speaker in this task repeats a context syllable, and a randomly-timed insertion signal cues them to insert a gesture in the upcoming syllable. The relative timing of the inserted gesture was observed to compress as a function of signal timing, but crucially, compression of

articulatory timing was more substantial in onset-syllabified insertions than in coda-syllabified ones. This compressibility difference is predicted by a {CV}{C} mixed control model: coordinative processes are expected to exhibit relatively more variation as a function of the amount of time available for coordinative planning, whereas the time-courses of selection and deselection processes are not expected to exhibit such variation. Hence the externally imposed time pressure on execution in the task may have promoted the use of a {V}{C} regime for coda-syllabified insertions.

Differences in articulatory timing across experimental tasks indicate that speakers alter control regimes to adapt to task demands, but these findings do not necessarily inform our understanding of control in spontaneous conversational speech. Unfortunately direct elicitation tasks, for example of coda timing in English (e.g. Browman & Goldstein, 1995; Marin & Pouplier, 2010), are inconclusive. Such experiments show that a post-vocalic coda gesture is initiated either slightly before the physical target of a preceding vocalic gesture is reached, or at nearly the same time. Yet because the articulatory implementation of a post-vocalic gesture is biomechanically coupled to the tongue (even if indirectly through the jaw), the apparent synchronicity of coda gesture initiation and vocalic target attainment can be viewed as an artifact of their biomechanical coupling. The timing pattern can in fact be interpreted in one of two ways: either the vocalic and consonantal gestures are co-selected and phased with a high degree of asynchrony; or, the gestures are competitively selected but a small degree of internalization allows the second gesture to be initiated slightly before the target of the first is achieved.

The forms of variation described above are important because they lend support to the notion that speakers retain the ability to use competitive control even after developing coordinative control. This provides a basis for understanding how phonological patterns associated with competitive control could be productive. There are other reasons why retaining the option of competitive control may be desirable. Competitive control is hypothesized to be used for controlling speech rate, for producing hyperarticulated speech, for emphasizing sub-units of utterances, and for a variety of paralinguistic functions such as communicating emotional states. Empirical work should be conducted to test each of these hypotheses.

One final point worth making is that phonological patterning and control organization are probably related in a bicausal manner. For example, contour tone restrictions were hypothesized above to originate from biases associated with competitive control; however, children may also infer control regimes from phonological patterns. For example, contour tone distributions in a language may cue language learners that a syllable with a contour tone involves two co-selection sets. Hence biases on selection of contour tones and the presence of contour tones in certain phonological contexts may concurrently reinforce one another.

The reciprocity between control and phonological patterning applies to all of the patterns discussed above and speaks to the notion that phonetic mechanisms and phonological phenomena co-emerge. Because such interactions are recapitulated in development for each speaker, their effects can be understood as an important source of bias in the diachronic evolution of the language. Coordinative control regimes originate from developmental transitions involving feedback internalization. This process must be influenced by the input that children receive from their language environment, which in turn depends on various stochastic and historical factors that future work should investigate.

6. Conclusion

Selection-coordination theory addresses a variety of phenomena that alternative theories and models do not satisfactorily explain. Hierarchical production models fail to account for developmental and cross-linguistic phonological/phonetic patterns because they assume a static organization of control. The explanatory capacity of articulatory phonology is unduly restricted because it lacks mechanisms for organizing coordinatively controlled gestures into competitively controlled sets. In contrast, the success of selection-coordination theory is ultimately attributable to its integration of two independently motivated cognitive mechanisms—selection and coordination—and to the concept of internalization as a mechanism for developmental transitions from competitive to coordinative control. The notion that the organization of control is dynamic makes selection-coordination theory flexible enough to explain a wide range of phenomena.

Selection-coordination theory has profound importance for our understanding of the emergence of hierarchical phonological structure, insisting that such structure is the result of a developmental variation in the organization of control, rather than an *a priori* design feature of phonological systems. Over the course of development children acquire coordinative control over articulatory gestures that were previously competitively controlled, thereby learning more complex sets of co-selected, coordinated gestures. Hierarchical structure may thus be viewed as a linguistic formalization of patterns which emerge in different stages of development. An important corollary is that phonological units such as segments, moras, and syllables are ultimately not so different from each other: all of these units can be conceptualized co-selection sets, i.e. sets of coordinated gestures. The units are differentiated not by essential characteristics but by when in the course of development, and in which phonological environments, they describe the organization of gestural selection.

New paths of research are illuminated in the selection-coordination framework. We have only explored a handful of the phonological phenomena which may be related to competitive and coordinative control. Revisiting the theoretical developments of phonological representations at all levels of the prosodic hierarchy, as well as the empirical data motivating those developments, should provide substantial insights. Sources of variation in control regimes must be extensively investigated to better understand and predict variation in phonological patterning. Along these lines, more sophisticated kinematic and acoustic assessments must be developed to distinguish between internalized feedback control and coordinative control, and there is a pressing need for more frequently sampled longitudinal phonetic data.

Another path of future investigation aims to elucidate the neural implementation of selection and coordination. The concept of a co-selection set relies fundamentally on the concept of a gesture, and gestures should be associated with distinct spatiotemporal patterns of neural activity. Selection can be viewed as a phase-transition occurring when the integrated activation of a pre-motor neural population, representing a gesture, reaches a threshold-like critical point. Coordination, in contrast, is conceptualized as relative phase coupling among oscillatory systems. Thus both oscillatory phase-coupling and critical phase transitions in neural ensembles are expected to be identifiable with sufficient temporal and spatial resolution of neural dynamics.

Lastly, our analysis has focused on only the lowest levels of the prosodic hierarchy. Yet the mechanisms and their developmental interplay can be more generally related to the phenomenon of chunking, i.e. the combination of representations into larger representations. This raises the question of whether selection-coordination dynamics apply to higher-level units such as feet and phrases, or even to morphosyntactic structure. A tantalizing possibility, the details of which remain to be worked out, is that parallel competitive and coordinative dynamics provide a fully general explanation for hierarchical structure. For gestural systems, phasing interactions govern relative timing of movement initiation, but for representations of conceptual systems, phasing interactions may serve relational or associative functions, rather than controlling timing directly. If we presume that cognitive representations of more abstract units interact simultaneously through selection and relational/associative coordination, the possibility arises that the serialization of actions can be reconciled with the parallelism of action relations.

Acknowledgments

The ideas developed in this manuscript have benefitted from discussions with Abby Cohn, John Hale, Louis Goldstein, Hosung Nam, Elliot Saltzman, Rachel Walker, Draga Zec, and from audiences in the Cornell Phonetics Lab and at Haskins Laboratories.

References

- Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, 90(2), 119–161.
- Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences*, 99(20), 13172.
- Averbeck, B. B., Crowe, D. A., Chafee, M. V., & Georgopoulos, A. P. (2003). Neural activity in prefrontal cortex during copying geometrical shapes. *Experimental Brain Research*, 150(2), 142–153.
- Bagemihl, B. (1989). The crossing constraint and “backwards languages.” *Natural Language & Linguistic Theory*, 7(4), 481–549.
- 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. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219–252.
- Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, 45(2–4), 140–155.
- Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6(2), 201–251.
- Browman, C., & Goldstein, L. (1990). Representation and reality: Physical systems and phonological structure. *Journal of Phonetics*, 18(3), 411–424.
- 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.
- 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. (1995). C-centers revisited. *Phonetica*, 52, 285–306.
- Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24(2), 209–244.
- Byrd, D., Tobin, S., Bresch, E., & Narayanan, S. (2009). Timing effects of syllable structure and stress on nasals: A real-time MRI examination. *Journal of Phonetics*, 37(1), 97–110.
- 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.
- Camarata, S. M. (1989). Final consonant repetition: A linguistic perspective. *Journal of Speech and Hearing Disorders*, 54(2), 159.
- Campbell, N., & Beckman, M. (1997). Stress, prominence, and spectral tilt. In A. Botinis, G. Kouroupetroglou, & G. Carayannis (Eds.), *Intonation: Theory, models and applications* (pp. 67–70).
- Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. *Laboratory Phonology*, 7, 419–447.
- Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron*, 45(5), 801–814.
- Cohn, A. (2003). Phonological structure and phonetic duration: The role of the Mora. *Working Papers of the Cornell Phonetics Laboratory*, 15, 69–100.
- Davis, S. (1990). Italian onset structure and the distribution of il and lo. *Linguistics*, 28(1), 43–56.
- De Jong, K. (2001). Rate-induced resyllabification revisited. *Language and Speech*, 44(2), 197–216.
- Dell, F., & Elmehraoui, M. (1985). Syllabic consonants and syllabification in Imdlawn Tashlykt Berber. *Journal of African Languages and Linguistics*, 7, 105–130.
- Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283–321.
- Demuth, K. (1995). Markedness and the development of prosodic structure. *Proceedings of NELS*, Vol. 25, 13–26.
- 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.
- Demuth, K., & Fee, E. J. (1995). *Minimal words in early phonological development*. Ms., Brown University and Dalhousie University.
- Desmurget, M., & Grafton, S. (2003). Feedback or feedforward control: End of a dichotomy. In S. Johnson-Frey (Ed.), *Taking action: Cognitive neuroscience perspectives on intentional acts* (pp. 289–338).
- Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology*, 10(6), 732–739, [http://dx.doi.org/10.1016/S0959-4388\(00\)00153-7](http://dx.doi.org/10.1016/S0959-4388(00)00153-7).
- Doyon, J., & Benali, H. (2005). Reorganization and plasticity in the adult brain during learning of motor skills. *Current Opinion in Neurobiology*, 15(2), 161–167.
- Duanmu, S. (1994). Syllabic weight and syllabic duration: A correlation between phonology and phonetics. *Phonology*, 11(1), 1–24.
- Faber, A. (1992). Phonemic segmentation as epiphenomenon. In P. Downing, S. D. Lima, & M. Noonan (Eds.), *The linguistics of literacy* (pp. 83–109). Amsterdam: John Benjamins.
- Fee, J., & Ingram, D. (1982). Reduplication as a strategy of phonological development. *Journal of Child Language*, 9(01), 41–54.
- Ferguson, C. A. (1983). Reduplication in child phonology. *Journal of Child Language*, 10(2), 239–243.
- Fikkert, P. (1994). *On the acquisition of prosodic structure*. Leiden and Amsterdam: Holland Institute of Generative Linguistics.
- Forssberg, H. (1999). Neural control of human motor development. *Current Opinion in Neurobiology*, 9(6), 676–682.
- Forssberg, 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.
- Frisch, S. (1997). *Similarity and frequency in phonology*. Northwestern University.

- Frisch, S., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, 30(2), 139–162.
- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 27–52.
- Gao, M. (2008). *Tonal alignment in Mandarin Chinese: An articulatory phonology account* (Doctoral dissertation). New Haven, CT: Yale University.
- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52(1), 45.
- Gick, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. *Papers in Laboratory Phonology VI: Constraints on Phonetic Interpretation*, 222–236.
- Gick, B., Campbell, F., Oh, S., & Tamburri-Watt, L. (2006). Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*, 34(1), 49–72.
- Goldrick, M., Baker, H. R., Murphy, A., & Baese-Berk, M. (2011). Interaction and representational integration: Evidence from speech errors. *Cognition*, 121(1), 58–72.
- Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, 21(6), 649–683.
- Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 215–249). Cambridge: Cambridge University Press.
- Goldstein, L., Chitoran, I., & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber. In *Proceedings of the XVIth international congress of phonetic sciences* (pp. 241–244).
- 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.
- Gordon, M. (2006). *Syllable weight: Phonetics, phonology, typology*. New York: Routledge.
- Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. London: Erlbaum.
- 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.
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3), 280–301.
- Gurney, K., Prescott, T. J., & Redgrave, P. (2001). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biological Cybernetics*, 84(6), 401–410, <http://dx.doi.org/10.1007/PL00007984>.
- 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*. New York: Routledge.
- Hawkins, S. (1984). On the development of motor control in speech: Evidence from studies of temporal coordination. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 11).
- Hayes, B. (1980). *A metrical theory of stress rules*. Massachusetts Institute of Technology.
- Hayes, B. (1982). Extrametricality and English stress. *Linguistic Inquiry*, 13(2), 227–276.
- Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20(2), 253–306.
- Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.
- Henry, F. M., & Rogers, D. E. (1960). Increased response latency for complicated movements and a "memory drum" theory of neuromotor reaction. *Research Quarterly of the American Association for Health, Physical Education, & Recreation*, 31, 448–458.
- 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.
- Hertlich, I., & Ackermann, H. (2000). Lip-jaw and tongue-jaw coordination during rate-controlled syllable repetitions. *The Journal of the Acoustical Society of America*, 107, 2236.
- Honorof, D. N., & Brownman, C. (1995). The center or edge: How are consonant clusters organized with respect to the vowel. In *Proceedings of the XIIth international congress of phonetic sciences* (Vol. 3, pp. 552–555).
- Hoole, P. (2006). Experimental studies of laryngeal articulation. In *Habilitationsschrift*. Germany: University of Munich.
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216, <http://dx.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.
- Hurme, P., & Sonninen, A. (1985). Development of durational patterns in Finnish CVCV and CVCCV words. *Papers on Speech Research*, 6, 1–14.
- Hyman, L. M. (1984). On the weightlessness of syllable onsets. In *Proceedings of the annual meeting of the berkeley linguistics society* (Vol. 10).
- Hyman, L. M. (1985). *A theory of phonological weight*. Dordrecht: Foris Publications.
- Imamizu, H., Miyazaki, S., Tamada, T., Sasaki, Y., Takino, R., Pütz, B., & Kawato, M. (2000). Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature*, 403 (6766), 192–195, <http://dx.doi.org/10.1038/35003194>.
- Jaeger, J. J. (2005). *Kids' slips: What young children's slips of the tongue reveal about language development*. Mahwah, NJ: Lawrence Erlbaum.
- Jeanerod, M. (1986). The formation of finger grip during prehension: A cortically mediated visuomotor pattern. *Behavioural Brain Research*, 19(2), 99–116.
- Jordan, M. I. (1997). Serial order: A parallel distributed processing approach. *Advances in Psychology*, 121, 471–495.
- Katz, W. F., & Bharadwaj, S. (2001). Coarticulation in rictative-vowel syllables produced by children and adults: A preliminary report. *Clinical Linguistics & Phonetics*, 15(1–2), 139–143.
- Kawato, M., Furukawa, K., & Suzuki, R. (1987). A hierarchical neural-network model for control and learning of voluntary movement. *Biological Cybernetics*, 57(3), 169–185.
- Kay, B. A., Kelso, J. A., Saltzman, E. L., & Schöner, G. (1987). Space-time behavior of single and bimanual rhythmical movements: Data and limit cycle model. *Journal of Experimental Psychology: Human Perception and Performance*, 13(2), 178.
- Kaye, J. (1992). Do you believe in magic? The story of s+C sequences. *SOAS Working Papers in Linguistics and Phonetics*, 2, 293–313.
- Kehoe, M., & Lleó, C. (2003). The acquisition of nuclei: A longitudinal analysis of phonological vowel length in three German-speaking children. *Journal of Child Language*, 30(03), 527–556.
- Kehoe, M., & Stoel-Gammon, C. (1997). Truncation patterns in English-speaking children's word productions. *Journal of Speech, Language and Hearing Research*, 40(3), 526.
- Kelso, J. (1982). *Human motor behavior: An introduction*. New York: Routledge.
- Kelso, J., & Holt, K. G. (1980). Exploring a vibratory systems analysis of human movement production. *Journal of Neurophysiology*, 43(5), 1183–1196.
- Kelso, J., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, 14, 29–59.
- Kelso, J., & Tuller, B. (1987). Intrinsic time in speech production: Theory, methodology, and preliminary observations. In E. Keller, & M. Gopnik (Eds.), *Sensory and motor processes in language*, Vol. 203 (p. 222). Hillsdale, NJ: Erlbaum.
- Kent, R. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, 19(3), 421–447.
- Kent, R. (1983). The segmental organization of speech. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 57–89). New York: Springer.
- Kent, R., & Foner, L. L. (1980). Speech segment duration in sentence recitations by children and adults. *Journal of Phonetics*, 8, 157–168.
- Kirk, C., & Demuth, K. (2005). Asymmetries in the acquisition of word-initial and word-final consonant clusters. *Journal of Child Language*, 32(4), 709.
- Clapp, S. T. (1995). Motor response programming during simple choice reaction time: The role of practice. *Journal of Experimental Psychology: Human Perception and Performance*, 21(5), 1015.
- Kozhevnikov, V., & Chistovich, L. (1965). *Speech: Articulation and perception*, Vol. 30, p. 543. Clearing House for Federal, Scientific, and Technical Information, JPRS.
- Krakauer, J. W., Ghilardi, M.-F., & Ghez, C. (1999). Independent learning of internal models for kinematic and dynamic control of reaching. *Nature Neuroscience*, 2(11), 1026–1031.
- Krakow, R. A. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures* (Doctoral dissertation). New Haven, CT: Yale University.
- Kuhnhert, B., Hoole, P., & Moschammer, C. (2006). Gestural overlap and C-center in selected French consonant clusters. In *Proceedings of the 7th international seminar on speech production* (pp. 327–334).
- Kunnari, S., Nakai, S., & Viham, M. M. (2001). Cross-linguistic evidence for the acquisition of geminates. *Psychology of Language and Communication*, 5(2), 13–24.
- Ladefoged, P. (1984). Out of chaos comes order: Physical, biological, and structural patterns in phonetics. In *Proceedings of the tenth international congress of phonetic sciences* (pp. 83–95).
- Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of the world*. Blackwell Publications.
- 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.
- Lavoie, L., & Cohn, A. (1999). Sesquisyllables of English: The structure of vowel-liquid syllables. In *Proceedings of the XVIth international congress of phonetic sciences* (pp. 109–112).
- Lebrun, Y., & Van Borsel, J. (1990). Final sound repetitions. *Journal of Fluency Disorders*, 15(2), 107–113.
- Levelt, C., Schiller, N. O., & Levelt, W. J. (2000). The acquisition of syllable types. *Language Acquisition*, 8(3), 237–264.
- Levelt, C., & Van de Vijver, R. (2004). Syllable types in cross-linguistic and developmental grammars. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in phonological acquisition* (pp. 204–218). Cambridge: MA: Cambridge University Press.

- Levelt, W. (1993). *Speaking: From intention to articulation*. The MIT Press.
- Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8(2), 249–336.
- Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2), 185–199.
- Linell, P. (2005). *The written language bias in linguistics: Its nature, origins and transformations*. London: Routledge.
- Löfqvist, A., & Yoshioka, H. (1984). Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production. *Speech Communication*, 3(4), 279–289.
- Logan, G. D. (1985). Executive control of thought and action. *Acta Psychologica*, 60(2), 193–210.
- Lombardi, L. (1990). The nonlinear organization of the affricate. *Natural Language & Linguistic Theory*, 8(3), 375–425.
- Lukatela, K., Carello, C., Shankweiler, D., & Liberman, I. Y. (1995). Phonological awareness in illiterates: Observations from Serbo-Croatian. *Applied Psycholinguistics*, 16, 463–488.
- MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, 8(3), 323–335.
- MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review*, 89(5), 483.
- 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.
- Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., & Wallace, M. E. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. *Contemporary Issues in Communication Science and Disorders*, 31, 105–122.
- McCarthy, J., & Prince, A. (1986). *Prosodic morphology I: Constraint interaction and satisfaction* (p. 1996)Rutgers University for Cognitive Science 1996 Technical Report 32.
- McLeod, S., Doorn, J. van, & Reed, V. A. (2001). Normal acquisition of consonant clusters. *American Journal of Speech-Language Pathology*, 10(2), 99.
- Menn, L. (1976). *Pattern, control, and contrast in beginning speech: A case study in the development of word form and word function*. University of Illinois at Urbana-Champaign.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81.
- Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology*, 50(4), 381–425, [http://dx.doi.org/10.1016/S0301-0082\(96\)00042-1](http://dx.doi.org/10.1016/S0301-0082(96)00042-1).
- Mucke, D., Nam, H., Hermes, A., & Goldstein, L. (2012). Coupling of tone and constriction gestures in pitch accents. In P. Hoole (Ed.), *Consonant clusters and structural complexity* (pp. 205–230). Berlin: Mouton de Gruyter.
- Munhall, K. G., Löfqvist, A., & Kelso, J. 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.
- Naeser, M. A. (1970). *Influence of initial and final consonants on vowel duration in CVC syllables*. Technical report no. 130. Wisconsin Research and Development Center for Cognitive Learning.
- 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.
- 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.
- Noteboom, S. G. (1973). The tongue slips into patterns. *Speech errors as linguistic evidence* (pp. 144–156).
- Ohala, J. (1975). The temporal regulation of speech. In G. Fant, & M. A.A. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 431–453). New York: Academic Press.
- Öhman, S. E. G. (1967). Numerical model of coarticulation. *The Journal of the Acoustical Society of America*, 41(2), 310–320.
- Ota, M. (2001). Phonological theory and the development of prosodic structure: Evidence from child Japanese. *Annual Review of Language Acquisition*, 1(1), 65–118.
- Perkell, J. S. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Cambridge, MA: MIT Press.
- Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28(1), 73–193.
- Port, R. F., & Leary, A. P. (2005). Against formal phonology. *Language*, 81, 927–964.
- Posner, M. L., & Snyder, C. (1975). Attention and cognitive control. In D. A. Balota, & E. J. Marsh (Eds.), *Information processing and cognition: The Loyola symposium* (pp. 55–85).
- Pouplier, M., & Beňuš, Š. (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. *Laboratory Phonology*, 2(2), 243–273.
- Prigogine, I., & Stengers, I. (1984). *Order out of chaos: Man's new dialogue with nature*. New York: Bantam Books.
- Proctor, M. I. (2009). *Gestural characterization of a phonological class: The liquids*. Yale University.
- Purcell, D. W., & Munhall, K. G. (2006). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, 119(4), 2288–2297.
- Raphael, L. J., Dorman, M. F., & Geffner, D. (1980). Voicing-conditioned durational differences in vowels and consonants in the speech of three-and four-year old children. *Journal of Phonetics*, 8, 335–341.
- Repp, B. H. (1986). Some observations on the development of anticipatory coarticulation. *The Journal of the Acoustical Society of America*, 79, 1616–1619.
- Richardson, I. W., & Rosen, R. (1979). Aging and the metrics of time. *Journal of Theoretical Biology*, 79(4), 415–423.
- Rosenbaum, D. A., Cohen, R. G., Jax, S. A., Weiss, D. J., & van der Wel, R. (2007). The problem of serial order in behavior: Lashley's legacy. *Human Movement Science*, 26(4), 525–554, <http://dx.doi.org/10.1016/j.humov.2007.04.001>.
- Rose, Y., & Inkelas, S. (2011). The interpretation of phonological patterns in first language acquisition. *The Blackwell Companion to Phonology*, 3, 2414–2438.
- Ross, J. R. (1969). *A reanalysis of English word stress*. Massachusetts Institute of Technology.
- Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science*, 6(1), 1–36.
- Sagey, E. C. (1986). *The representation of features and relations in non-linear phonology*. Massachusetts Institute of Technology.
- Sakai, K., Hikosaka, O., Takino, R., Miyauchi, S., Nielsen, M., & Tamada, T. (2000). What and when: Parallel and convergent processing in motor control. *The Journal of Neuroscience*, 20 (7), 2691–2700.
- Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, 19(4), 499–526.
- 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., & Loevenbruck, H. (2006). Multistable syllables as enacted percepts: A source of an asymmetric bias in the verbal transformation effect. *Perception & Psychophysics*, 68(3), 458–474.
- Schmidt, R. A., & Lee, T. (1999). *Motor control and learning*. Champaign, IL: Human kinetics.
- Schmidt, R. C., Bienvénu, M., Fitzpatrick, P. A., & Amazeen, P. G. (1998). A comparison of intra-and interpersonal interlimb coordination: Coordination breakdowns and coupling strength. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 884–900.
- Schmidt, R. C., & Turvey, M. T. (1994). Phase-entrainment dynamics of visually coupled rhythmic movements. *Biological Cybernetics*, 70(4), 369–376.
- Schöner, G., Jiang, W. Y., & Kelso, J. S. (1990). A synergistic theory of quadrupedal gaits and gait transitions. *Journal of Theoretical Biology*, 142(3), 359–391.
- Schöner, G., & Santos, C. (2001). Control of movement time and sequential action through attractor dynamics: A simulation study demonstrating object interception and coordination. *Proceedings of the 9th international symposium on intelligent robotic systems*.
- 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.
- Selkirk, E. O. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, 11(3), 563–605.
- Shaffer, L. H. (1976). Intention and performance. *Psychological Review*, 83(5), 375–393.
- Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In *The production of speech* (pp. 109–136). New York: Springer.
- Shaw, J., Gafos, A., 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.
- Simko, J., & Cummins, F. (2011). Sequencing and optimization within an embodied task dynamic model. *Cognitive Science*, 35(3), 527–562.
- Smith, A., & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language and Hearing Research*, 41(1), 18.
- Smith, B. L. (1992). Relationships between duration and temporal variability in children's speech. *The Journal of the Acoustical Society of America*, 91, 2165–2174.
- Smith, B. L., Kenney, M. K., & Hussain, S. (1996). A longitudinal investigation of duration and temporal variability in children's speech production. *The Journal of the Acoustical Society of America*, 99, 2344–2349.
- 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.
- 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. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152).
- Stetson, R. H. (1951). *Motor phonetics: A study of speech movements in action*. Amsterdam: North-Holland Publishing Company.
- Stokes, S. F., & Wong, I. M. (2002). Vowel and diphthong development in Cantonese-speaking children. *Clinical Linguistics & Phonetics*, 16(8), 597–617.
- 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.
- Thompson, A. E., & Hixon, T. J. (1978). *Nasal air flow during normal speech production*. University of Arizona.
- Thoroughman, K. A., & Shadmehr, R. (1999). Electromyographic correlates of learning an internal model of reaching movements. *The Journal of Neuroscience*, 19(19), 8573–8588.
- Tilsen, S. (2009a). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science*, 33(5), 839–879.
- Tilsen, S. (2009b). Toward a dynamical interpretation of hierarchical linguistic structure. *UC Berkeley Phonology Lab Annual Report*, 462–512.
- Tilsen, S. (2011). Metrical regularity facilitates speech planning and production. *Laboratory Phonology*, 2(1), 185–218.
- Tilsen, S. (2013). A dynamical model of hierarchical selection and coordination in speech planning. *PloS One*, 8(4), e62800.
- Tilsen, S. (2014a). Selection and coordination of articulatory gestures in temporally constrained production. *Journal of Phonetics*, 44, 26–46.
- Tilsen, S. (2014b). Selection-coordination theory. *Cornell Working Papers in Phonetics and Phonology*.
- 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., & Ricciardi, E. (2014). Syllable count judgments and durations of liquid rimes in English. *Cornell Working Papers in Phonetics and Phonology*.
- 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.
- Todorov, E., & Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(11), 1226–1235.
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, 39(3), 1429–1443.
- Tse, J. K.-P. (1978). Tone acquisition in Cantonese: A longitudinal case study. *Journal of Child Language*, 5(02), 191–204.
- Tuller, B., & Kelso, J. (1990). Phase transitions in speech production and their perceptual consequences. *Attention and Performance*, 13, 429–452.
- Tuller, B., & Kelso, J. (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. (1990). Coordination. *American Psychologist*, 45(8), 938–953.
- Turvey, M. T. (2007). Action and perception at the level of synergies. *Human Movement Science*, 26(4), 657–697.
- Vihman, M., & Velleman, S. L. (2000). The construction of a first phonology. *Phonetica*, 57(2–4), 255–266.
- 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.
- Weismer, G. (1984). Acoustic analysis strategies for the refinement of phonological analysis. *ASHA Monographs*, 22, 30–52.
- Westling, G., & Johansson, R. S. (1984). Factors influencing the force control during precision grip. *Experimental Brain Research*, 53(2), 277–284.
- Wolpert, D. M., Ghahramani, Z., & Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences*, 5(11), 487–494.
- Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7), 1317–1329.
- Yi, H., & Tilsen, S. (2015). Gestural timing in Mandarin tone sandhi. *Proceedings of Meetings on Acoustics*, 22(1), 060003, <http://dx.doi.org/10.1121/2.0000029>.
- Zec, D. (2003). Prosodic weight. In C. Féry, & R. van de Vijver (Eds.), *The syllable in optimality theory* (pp. 123–143).
- Zec, D. (2007). The syllable. In P. De Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 161–194).
- Zhang, J. (2004). The role of contrast-specific and language-specific phonetics in contour tone distribution. *Phonetically Based Phonology*, 157–190.