

Selection-coordination theory

Sam Tilsen¹

1. Introduction

Speech utterances are comprised of hierarchically organized units, such as features/gestures, segments, moras, and syllables. Why these units are organized hierarchically is not well understood; phonological theories vary widely in their treatment of them, and models of speech production ascribe them diverse roles. This paper presents a *selection-coordination theory* of phonological structure, which attempts to unify our understanding of phonological units and explain why they are organized hierarchically. The theory holds that hierarchical organization emerges from a recurring trend in speech development whereby children acquire coordinative regimes of control over motor plans that were previously competitively selected, with transitions from competitive to coordinative control being made possible by the internalization of sensorimotor feedback. In this framework, segments, moras, and syllables are understood as differently-sized instantiations of the same type of motor planning that have differing roles in the course of development; diverse phonological patterns are given a cognitively grounded account that derives from distinguishing competitive and coordinative regimes of articulatory control. Evidence for the theory is drawn from research in motor control, phonological development, and phonological and phonetic patterns in adult speech.

To begin with an analogy, consider how notes on a piano can be played: piano keys can be pressed one after another, such that the notes are produced in a sequence. Alternatively, several keys can be pressed together as a chord, potentially with leading or trailing grace notes. The sequence of notes corresponds to a mode of competitive control in which actions are planned in parallel but produced sequentially; the chord corresponds to a mode of coordinative control, where actions are selected together and produced with precisely controlled relative timing. These two control regimes are fundamental to motor control generally, and distinguishing between them provides insights into the structure of speech utterances and phonological patterns. The selection-coordination theory presented here holds that there are two distinct cognitive mechanisms involved in control of speech articulation—selection and coordination—and that the interaction between these mechanisms results in two prototypical regimes of control: competitive selection and co-selection/coordination.

Young children are not equally competent in utilizing competitive and coordinative control. Pursuing the piano analogy, children initially learning to play tend to press keys one at a time, relying on competitive control. Subsequently they develop coordinative control: they learn to play sets of notes and patterns in which relative timing is more exactly governed. Competitive control developmentally precedes coordinative control. In the domain of speech, the consequences of this are of profound importance to our understanding of phonological structure: over the course of development children acquire coordinative control over speech gestures that were previously competitively controlled, thereby learning more complex and highly coordinated control structures. The hierarchical structure we observe emerges from this developmental progression. An important consequence of the theory is that phonological units, such as features/gestures, segments, moras, and syllables, are ultimately not so different from each other: all of these are

¹ I thank Abby Cohn for insightful feedback on this manuscript, along with Draga Zec, John Hale, Hosung Nam, Elliot Saltzman, and audiences at Haskins Laboratories and the Cornell Phonetics Lab.

sets of gestures, and what distinguishes them is when in the course of development coordinative control over those gestures is acquired.

In this paper we elaborate the selection-coordination theory, drawing evidence from research in motor control, phonological development, and phonological and phonetic patterns in adult speech. Section 1 presents the distinction between selection and coordination of movement plans and examines how these concepts have been incorporated into models of speech planning and production. Section 2 reviews literature on speech development and identifies a recurring pattern in which coordinative control is acquired over articulatory gestures that were previously competitively selected. Sections 3 and 4 examine phonetic and phonological variation associated with segments and moras from a motoric and developmental perspective. Section 5 further expands upon the relation between development changes in control, hierarchical structure, and phonological patterns.

1.1 Dissociation of Selection and Coordination

A fundamental issue in speech motor control is how the sequencing and timing of movements is accomplished. Here we distinguish between two mechanisms: *selection* and *coordination*. Selection involves the choice of which movements to make and control over the order in which to execute them. Coordination involves finer control over when movements are initiated relative to one another. Related distinctions have been drawn by numerous researchers using various terminology: 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 holds that two prototypical regimes of control arise from the dissociation of selection and coordination: *competitive control*, in which movement plans may be selected in a mutually exclusive manner and executed sequentially, and *coordinative control*, in which plans may be selected together—*co-selected*—and executed with precisely coordinated relative timing (see also Tilsen, 2013, 2014). The role of sensory feedback is key in distinguishing competitive and coordinative control and in understanding how the two are developmentally related. Competitive control is feedback-driven: sensory feedback regarding the attainment of a movement target induces suppression and deselection of the currently selected movement plan, allowing other plans to compete for selection. Coordinative control on the other hand is feedforward: plans are co-selected and phasing mechanisms determine their precise timing.

The distinction can be usefully understood with the analogy between articulatory timing and the timing of notes played on a piano. Piano keys can be pressed one after another, such that the notes are produced in a sequence, without substantial overlap, as in Fig. 1A. This corresponds to competitive control: each note is selected after the preceding one achieves its target (the notes are *competitively selected*). Alternatively, two or more keys can be pressed together as a chord or as a principle note co-produced with grace notes, various examples of which occur in Fig. 1B-E. The notes that are combined as such are coordinatively controlled: they constitute a group of actions which are selected together (*co-selected*) and produced in an overlapping manner with precisely controlled relative timing (*coordinated*).

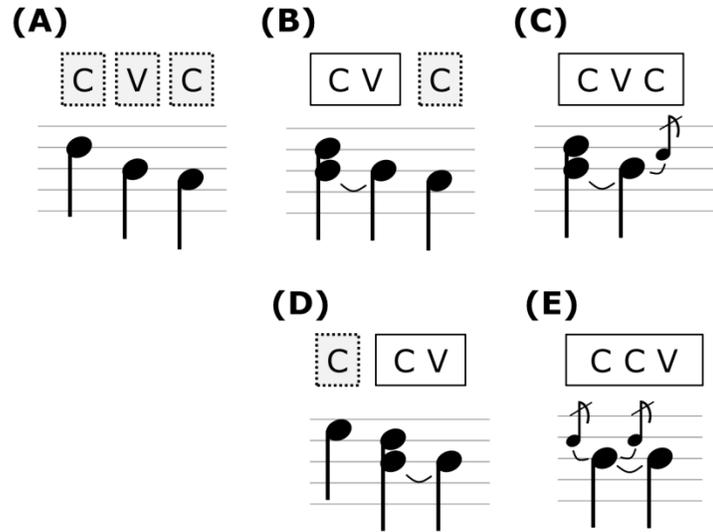


Fig. 1. Analogy between articulatory timing and timing of notes played on a piano. Shaded boxes represent articulatory gestures which are competitively controlled, unshaded boxes represent coordinatively controlled gestures.

Coordinative control can develop from competitive control through the internalization of feedback. Internalization refers to the development of an internal cognitive model that predicts the sensory consequences of motor commands, allowing them to be anticipated. Much recent research has emphasized the role of internal models in motor control and motor learning (Desmurget & Grafton, 2003; Todorov & Jordan, 2002; Wolpert, Ghahramani, & Flanagan, 2001; Wolpert & Kawato, 1998). Fig. 2 schematizes the relations between feedforward control, external feedback, and internal feedback. Feedforward control involves a mapping from a sensory target to a motor plan via an inverse model. External feedback control utilizes a correction signal derived from comparing afferent sensory feedback with sensory targets. Internal feedback utilizes a correction signal derived from comparing the predicted consequences of current motor commands with sensory targets. This prediction involves mapping an efference copy of the motor commands to a sensory representation through an internal, forward model.

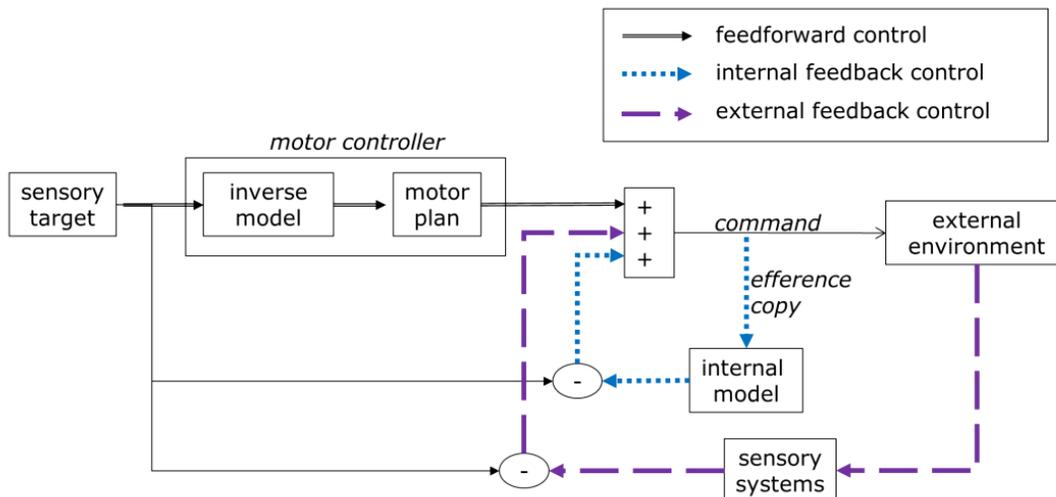


Fig. 2. Schematic illustration of feedforward control, external feedback control, and internal feedback control. The internal model maps an efference copy of the motor commands to a sensory representation, which can be compared with a target to generate an error correction signal. For simplicity of illustration different modalities of feedback are not distinguished.

1.2 Competitive Selection: a Mechanism for Sequencing

Competitive selection is a dynamical mechanism for controlling the sequencing of motor plans activated in parallel. Early models of sequencing held that movement plans are activated sequentially through an associative reflex chain. However, on the basis of anticipatory and perseveratory error patterns, free combinatoriality of movements, and other observations, Lashley (1951) proposed instead that motor plans are active in parallel and a controller selects them in the appropriate order. Some competitive selection models are based on computer memory buffer metaphors (Shaffer, 1976; Sternberg et al., 1988), but here we emphasize a dynamical approach called *competitive queuing* (Bullock & Rhodes, 2002; Grossberg, 1978, 1987), schematized in Fig. 3. Motor plans are associated with activation variables and the plans are initially activated to varying degrees. The activation levels of the plans grow as they compete to reach a selection threshold, and the first plan to reach the threshold is selected for execution. A gating mechanism ensures that only one plan at a time “wins” the competition. The selected plan is subsequently suppressed; this induces a new round of competition, and the process iterates. Hence in this framework the relative activation of motor plans determines the order in which they are selected: sequencing is controlled through relative activation.

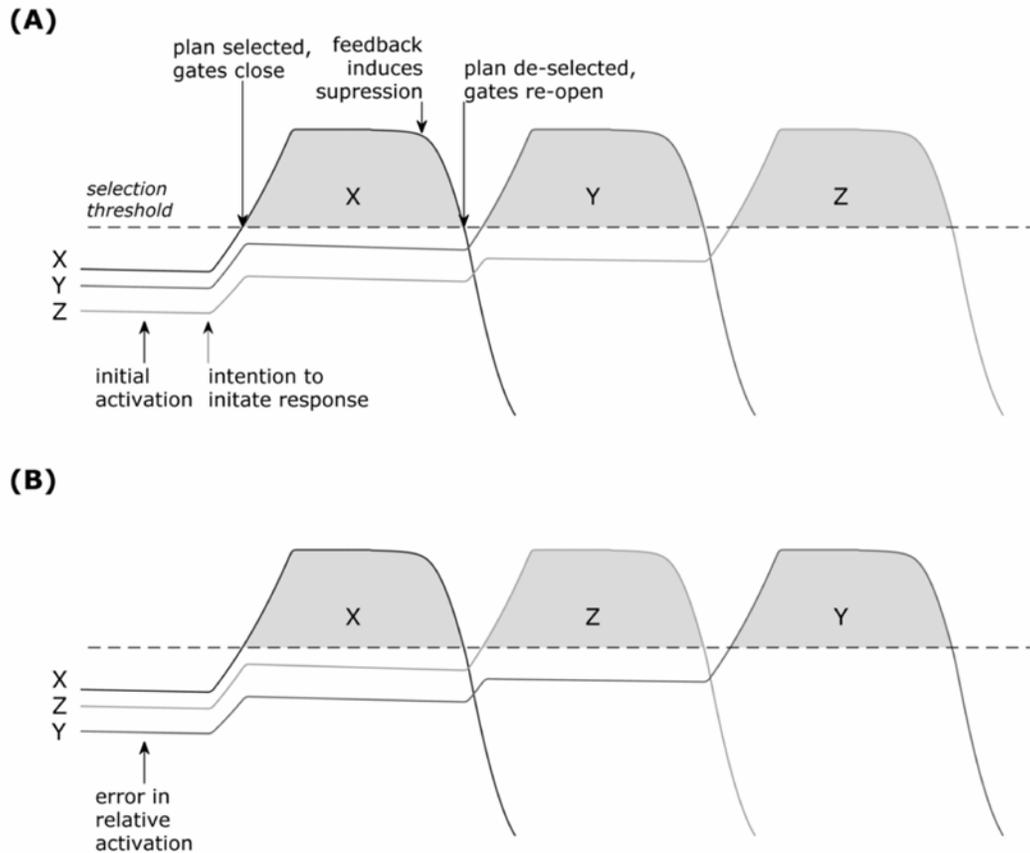


Fig. 3. (A) Illustration of competitive selection dynamics governing the sequencing of three movements. When the first plan exceeds a threshold it is selected. Competitors are temporarily gated until feedback induces suppression and deselection of the selected plan. The initial activation gradient determines the order in which plans are selected. (B) Error in initial relative activation results in errorful sequencing.

Competitive selection models are an innovation over simply stipulating memory for serial order, because the conceptual device for representing motor plans—activation—is intrinsically integrated with the mechanism that sequences movements. Competitive selection models can account for response initiation latencies and error patterns (Bullock & Rhodes, 2002; Bullock, 2004). Specifically, when more plans are co-active, inhibitory interactions between plans or activation normalization diminishes their individual activation levels and increase latencies to selection. This accounts for the finding that reaction time to initiate a spoken or typed 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). Common anticipatory and perseveratory error patterns can arise when noise or external influences alter the relative activations of plans and cause selection to proceed in an errorful manner (e.g. Fig. 3B). Notably, neural recordings in primates trained to perform sequential actions provide support for competitive selection dynamics (Averbeck, Chafee, Crowe, & Georgopoulos, 2002; Averbeck, Crowe, Chafee, & Georgopoulos, 2003; Cisek & Kalaska, 2005).

Feedback is responsible for suppression and deselection motor plans in competitive selection models. Grossberg (1978) cited two potential sources of suppression: inhibition from external

feedback or recurrent inhibition, the latter of which can be associated with internal feedback control. Most models implementing competitive selection incorporate some mechanism akin to internal feedback for the purpose of deselecting plans, such as an a deactivation time parameter (Rumelhart & Norman, 1982), “postselection negative feedback” (G. Dell, 1986), or a “a non-specific response suppression signal” (Bohland, Bullock, & Guenther, 2010).

Competitive selection dynamics can generate a pattern in which a previously selected movement is re-selected at the expense of another (Fig. 4B), or in which a plan is intentionally reselected (Fig. 4C). Reselection of one plan at the expense of another occurs when the first plan is not fully suppressed after deselection and/or the second is weakly activated, which may be due to immature representations of movement plans. In these circumstances the previously selected plan wins a second round of competition. Reselection may also arise as an intentional strategy for prolonging the period of time that a target posture is maintained.

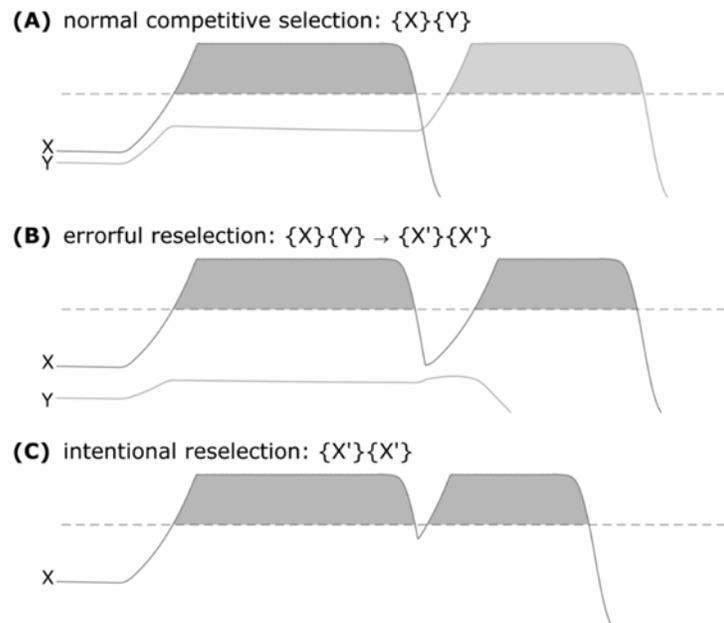


Fig. 4. Illustration of reselection in a competitive selection model. Activation variables of motor plans X and Y are shown; X' indicates identity between plans. (A) normal competitive selection; (B) errorfull reselection due to deficient activation of a competitor; (C) intentional reselection.

1.3 Coordination: a Mechanism for Control of Relative Timing

Competitive selection does not readily accommodate actions in which movements are executed simultaneously or with a high degree of overlap. Many motor actions involve multiple movements conducted in a precisely timed, overlapping manner. For example, when adults reach to grab an object, they adjust the shape of the fingers while moving the hand to a desired position: the reaching and hand-shaping movements are precisely coordinated (Jeannerod, 1986). In speech, consonantal movements in a syllable onset are co-produced with the movements responsible for attaining the target vocalic posture (Kozhevnikov & Chistovich, 1965; Öhman, 1967; Perkell, 1969), often with nearly synchronous relative timing (Browman & Goldstein, 1988; Krakow, 1989). Overlapping execution of movements and precision control cannot be readily modeled with competitive selection: the initiation of the vocalic movement in a

consonant-vowel syllable is not delayed until the consonantal movement has achieved its target. Hence an alternative mechanism is required to understand simultaneously produced movements.

Bernstein (1967) developed the concept of coordination 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 the individual degrees of freedom. Bernstein proposed instead that groups of muscles become linked in synergies or coordinative structures, thereby reducing the degrees of freedom (cf. 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, thereby reducing degrees of freedom; 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 control of which is governed by general principles.

The task-dynamic model of articulatory phonology holds that coordinative structures called *articulatory gestures* are the basic units of speech (Browman & Goldstein, 1989; Saltzman & Munhall, 1989). Gestural goals are associated with target values of vocal tract variables, which are defined in coordinates of vocal tract geometry. For example, a bilabial closure gesture, associated with the articulatory goal of closing the lips, is associated with a zero or negative target value of lip aperture. Articulatory gestures are hypothesized to be the fundamental units of both information and action in phonological systems (Goldstein, Byrd, & Saltzman, 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/effector positions, the model allows for gestures to reach their targets regardless of initial conditions.

Rhythmic movements have played a fundamental role in extending coordinative principles to control of multiple gestures. Periodic motions can be associated with limit-cycle oscillators, whose nature is determined by elastic 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, oscillations with different intrinsic frequencies will entrain to a compromise frequency. Moreover, there are two preferred ways in which a pair of oscillations can be synchronized: in-phase and out-of-phase, the former being more stable than the latter. In a classic experiment, Haken, Kelso, & Bunz, (1985) showed that people attempting to wag their fingers out of phase will exhibit a spontaneous transition from anti-phase coordination to in-phase coordination when 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 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 in repeating a VC syllable with increasing rate, the timing of the articulatory movements reorganizes to a CV pattern. Tuller & Kelso (1990, 1991) replicated this effect; they argued that the coda gesture is initially anti-phase coordinated with the vocalic one, and when a critical value of repetition rate is reached, the relative phase of the gestures undergoes a phase transition to in-phase coordination (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 critical rate and hysteresis in the phase transition.

However, the movements of spontaneous conversational speech are not rhythmic. In order to apply principles of rhythmic coordination to the relative timing of non-rhythmic speech movements, a conceptual leap is required. Drawing on the notion of intrinsic time (Kelso & Holt, 1980; Prigogine, Stengers, & Toffler, 1984; Richardson & Rosen, 1979), Kelso & 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 coordination between non-rhythmic movements can be defined with a relative phase variable, where the component phases are derived from the virtual cycles. These concepts provided the basis for *phasing* in the gestural score of task dynamics (Browman & Goldstein, 1992; Saltzman & Munhall, 1989). Control of phasing was subsequently associated with stabilized relative phases in systems of coupled planning oscillators (Saltzman & Byrd, 2000).

The coupled oscillators model of gestural phasing provides conceptually appealing characterizations of articulatory timing. One example is the C-center effect, illustrated schematically in Fig. 5C below. The figure relates the relative phase of planning oscillators to the initiation of gestural activation intervals, which are the periods of time when a gesture drives movement. Browman & Goldstein (1988) observed that in CV, CCV, and CCCV syllables the midpoint of the midpoints of consonantal gestures (i.e. the C-center) maintains a relatively constant temporal relation to the vocalic gesture. Browman & Goldstein (2000) reconceptualized this pattern as a result of pairwise coupling interactions between the virtual cycles of the gestures: the consonantal gestures are coupled out of phase with one other but coupled in-phase with the vowel, and the C-center pattern arises from minimization of energy in relative phase potentials. Nam & Saltzman (2003) implemented a dynamical model of a system of three oscillators that could produce this pattern. Replications of the C-center effect in English have been reported in a number of subsequent studies (Byrd, 1995, 1996; Honorof & Browman, 1995; Marin & Pouplier, 2010; Sproat & Fujimura, 1993), and the effect has likewise been observed in word-initial onsets in other languages, such as French (Kuhnert, Hoole, & Mooshammer, 2006), Italian (Hermes, Grice, Mücke, & Niemann, 2008), Georgian (Goldstein, Chitoran, & Selkirk, 2007), and Serbian (Tilsen et al., 2012).

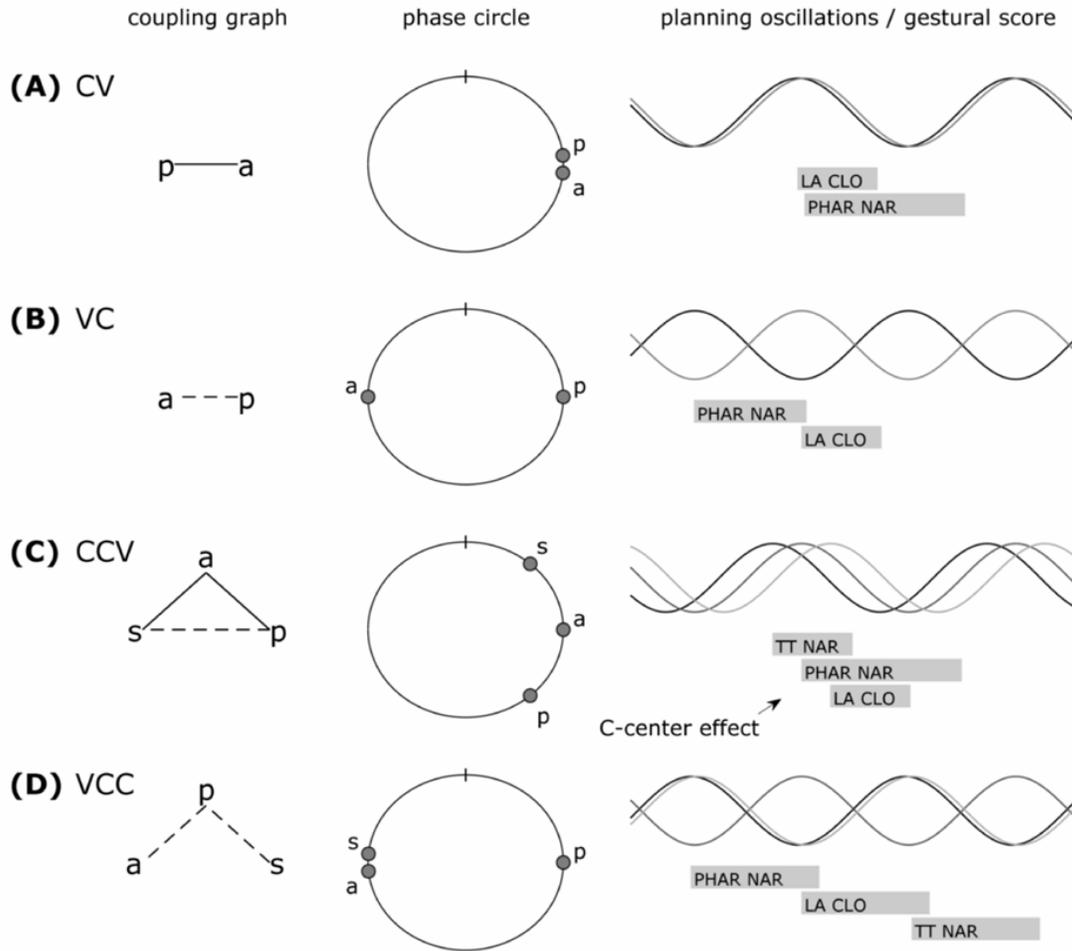


Fig. 5. Illustration of the coupled oscillators model of intergestural coordination for several different syllable shapes. For each syllable a hypothesized coupling graph is shown (solid lines: in-phase coupling; dashed lines: anti-phase coupling), along with a representation of stabilized phases of planning oscillators on a phase circle. Peaks in planning oscillations correspond to the top of the phase circle and determine when gestures are initiated in the gestural score.

No evidence for a C-center effect has been observed in post-vocalic coda gestures. Instead, coda consonantal gestures appear to be timed relative to the attainment of the target of a preceding gesture (Browman & Goldstein, 1988; Marin & Pouplier, 2010). Byrd (1996) observed greater variability in the relative timing of consonantal gestures in codas compared to those in onsets. Articulatory phonology hypothesizes that the planning oscillators associated with coda consonants are anti-phase coupled to the planning oscillators associated with a preceding vocalic or consonantal gesture (Nam, Goldstein, & Saltzman, 2009; Nam & Saltzman, 2003; Saltzman, Nam, Krivokapic, & Goldstein, 2008). Hence the coupled oscillators model utilizes coordinative timing for both onset and coda consonantal gestures.

1.4 Integrating selection and coordination

The coupled oscillators model, or more generally, the concept of coordinative phasing, is successful in accounting for co-production of gestures and precisely controlled relative timing, while competitive selection cannot readily account for such patterns. In contrast, competitive

selection models are successful in accounting for error patterns and effects of sequence complexity on response initiation, while the coupled oscillators model does not address these phenomena. Hence a model that incorporates both mechanisms is desirable. In recent work I have focused on developing and experimentally testing a speech production model that integrates selection and coordination mechanisms (Tilsen, 2013). In the model, sets of gestural plans can be either competitively selected or co-selected, and the relative timing of co-selected plans is controlled through coordinative phasing.

A basic question that arises in the framework is this: for a given utterance, which articulatory gestures are competitively selected and which are co-selected/coordinated? Three possibilities for a CVC syllable are represented in Fig. 6 below. In addition to relative phasing and gestural activation intervals, the figure shows planning activation, which represents the states of higher-level systems hypothesized to determine when gestures are selected. Sets of co-selected gestures—*co-selection sets*—are henceforth denoted with combinations of “C” and “V” symbols in curly brackets, i.e., {C}, {V}, {CV}, and {VC}. The reader should keep in mind that the “C” and “V” symbols do not refer to segments or autosegmental timing slots; rather they refer to oral articulatory gestures associated with consonants and vowels. Moreover, the order of gestures in a bracketed co-selection set does not indicate a sequential ordering of movements.

In a “purely competitive” selection model, i.e., {C}{V}{C}, each oral articulatory gesture is competitively selected relative to the others. This model is akin to the gradient ordering directions into articulatory vectors model (GODIVA: Bohland et al., 2010) and the hierarchical selection model of Dell (1986), where segments are competitively selected. The purely competitive model does not utilize coordinative timing; it therefore cannot account for the coproduction of onset consonants and vowels or generate C-center effects. In a “purely coordinative” model, i.e. {CVC}, all oral articulatory gestures are co-selected and coordinated. The coda and vocalic gestures are anti-phase coordinated, while the onset and vocalic gestures are in-phase coordinated. This model corresponds to what is hypothesized by articulatory phonology, although selection is not explicitly modeled in that framework.

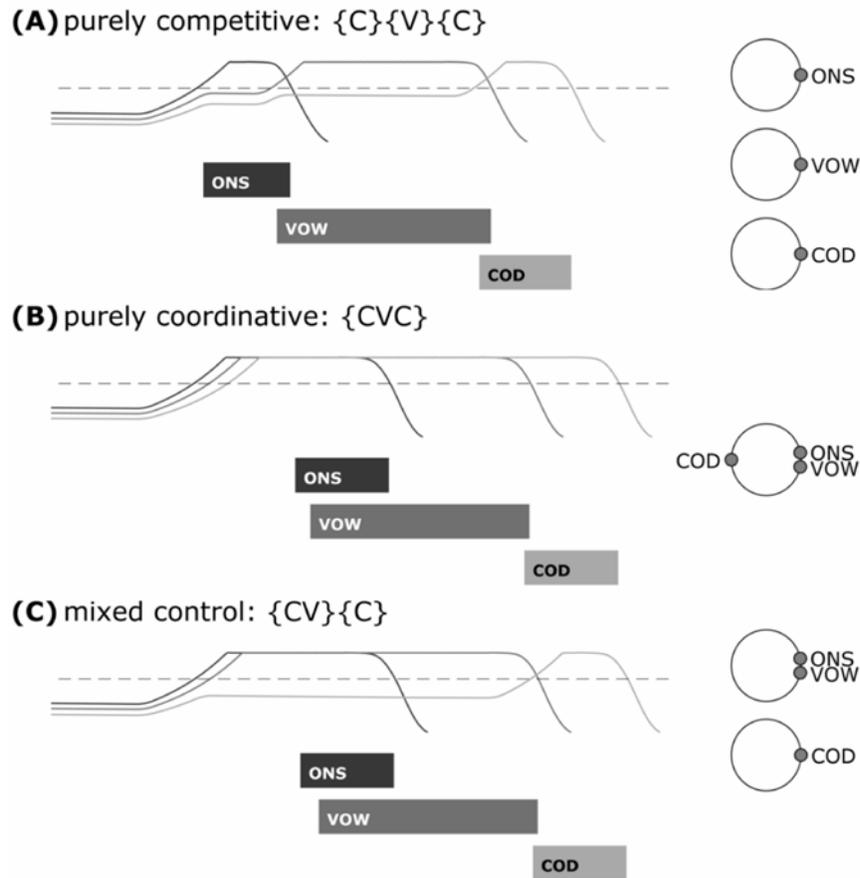


Fig. 6. Alternative control regimes for a CVC syllable. (A) purely competitive control in which all gestures are competitively selected; (B) purely coordinative control in which all gestures are co-selected and coordinated; (C) mixed control: vowel and coda gestures are competitively selected, onset and vowel gestures are co-selected and coordinated.

In the “mixed control” model, i.e. $\{CV\}\{C\}$, the onset and vocalic gestures are co-selected and coordinated, while the coda gesture is competitively selected relative to the vocalic gesture (Tilsen, 2013). In subsequent sections we argue that many developmental and phonological patterns can be more readily interpreted with a mixed control model than with a purely coordinative model, and that task-specific variation in articulatory timing suggests that adult speakers can employ either competitive or coordinative regimes of control over post-vocalic gestures.

Selection-coordination theory holds that coordinative control develops from decreasing reliance on external feedback in conjunction with increasing reliance on internal feedback. This allows for the dissociation of internal feedback for gating and internal feedback for suppression, which enables gestural plans to be co-active. The co-activation of plans in turn facilitates the co-selection of those plans and the development of coordinative control. Fig. 7 shows the consequences of the hypothesized progression from competitive control to coordinative control, with two intermediate stages. In each stage the figure shows the activation variables of two gestures, X and Y, along with a phase circle representing the influence of phase coupling. The dynamics of the activation variables are derived from an extension of the production model in (Tilsen, 2013); increased reliance on internal feedback is assumed to result in a decrease in the

strength of inhibitory gating and a decrease in the relative timing of release from inhibitory gating and feedback regarding target achievement.

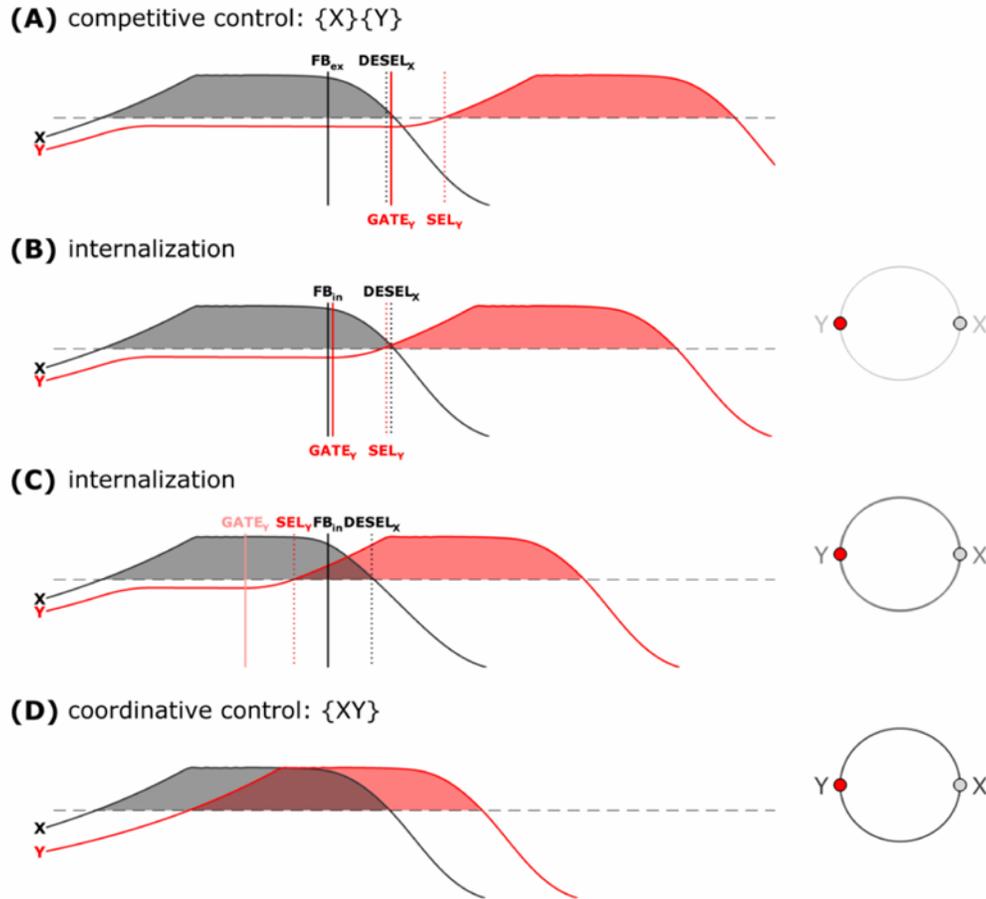


Fig. 7. Schematization of the consequences of feedback internalization on articulatory timing. (A) gestures X and Y are competitively selected with external feedback control. (B) internalized feedback allows Y to be selected when X is deselected. (C) further internalization dissociates gating and suppression, allowing selection of Y prior to suppression of X. (D) loss of inhibitory gating and co-activation of gestures results in coordinative control.

Fig. 7A shows prototypical competitive selection: release of inhibitory gating is driven by external feedback, hence Y is gated until X has been deselected. Fig. 7B represents a relatively small degree of internalization, where X target attainment is anticipated through an internal model, thereby eliminating the delay between target achievement and release of inhibitory gating. More extensive internalization with weakening of inhibitory gating in Fig. 7C allows Y to be selected prior to deselection of X. This results in a circumstance in which X and Y are contemporaneously active. The co-activation may strengthen coordinative phase coupling and further diminish inhibitory gating, leading to the development of prototypical coordinative control, as shown in Fig. 7D. It should be emphasized that coordinative control does not appear out of thin air in this conceptualization: as gating becomes dissociated from deselection, coordinative phasing may develop in parallel with the internalization of feedback.

2. Selection and Coordination in Development

A review of studies on the acquisition of speech and phonological structure indicates that competitive control precedes coordinative control in development. 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 precedence of competitive control is expected on the basis of the hypothesized model of internalization described above and may apply generally to motor control. 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 control of hand posture until they have transported the hand to the object (Jeannerod, 1986; von Hofsten, 1979). Since developmental transitions from competitive to coordinative control occur in nonspeech domains, it is not surprising that they are pervasive in speech as well.

2.1 Selection and Coordination in Pre-Babble and Babble

Consonant-vowel articulations produced in pre-babble appear to be competitively selected, but in babble children develop the ability to co-select them. The consonantal articulations in "goos" and marginal babble tend to involve first the production of a closure, and then the production of a vocalic posture, or vice versa (Oller, Eilers, Neal, & Schwartz, 1999). Children will combine previously learned elements together, e.g., a raspberry and vocoid, but these are produced in such a way that the transition between elements is too slow to be characteristic of a consonant-vowel syllable in adult speech (Oller, 2000). The same is not so in babbling. Most of consonantal gestures observed in babble are transcribed as singleton onsets of CV syllables and their acoustic transitions into vowels are more rapid than those in pre-babble (Kent & Bauer, 1985; Oller, Wieman, Doyle, & Ross, 1976).

Several forms of evidence suggest that babble begins as a rhythmically organized movement which subsequently becomes more linguistically organized (Iverson, Hall, Nickel, & Wozniak, 2007; Oller, 2000). "Silent babbles" produced by infants indicate that mandibular oscillations are initially independent of the phonation that often accompanies them (Meier, McGarvin, Zakia, & Willerman, 1997). The relatively limited repertoire of consonantal articulations in early babble and subsequent expansion of that repertoire (MacNeilage, Davis, & Matyear, 1997; MacNeilage & Davis, 2000), as well as continuities between late babbling patterns and early word forms—e.g., reduction/omission of codas and clusters, absence of aspiration, word-final devoicing, substitutions, preference for apical over dorsal consonants (Oller, 2000; Vihman, 1996)—evidence a trend for increasing linguistic organization. Several studies have reported a coincidence of the onset of reduplicated babble with non-speech rhythmic activities (Eilers et al., 1993; Iverson et al., 2007; Thelen, 1979). Iverson & Thelen (1999) argued that the mechanism for this coincidence is entrainment through direct coupling between vocal and non-vocal motor systems.

There are two ways in which the oscillatory nature of babble provides a basis for transitioning from competitive to coordinative control. First, in order to co-select a vocalic and consonantal movement, children must internalize feedback control over the movements. The rhythmicity of babbling may facilitate internalization of feedback because the repetition of a sequence enhances the predictability of its acoustic manifestation; indeed, the onset of babbling has been related to the use of auditory awareness to control motor activities (Ejiri, 1999; Kent, 2000). Second, the

presence of a physical oscillation implies the presence of a cognitive one. The cognitive oscillation may facilitate phasing-based control over the consonantal and vocalic articulations.

Adoption of a coordinative regime of control over consonantal and vocalic articulations may also account for developmental progressions in babbling. Babbling is often classified as one of two types: canonical (or reduplicative) babble in which each syllable involves basically the same consonantal and vocalic movements; and variegated babble, in which the vowels and/or consonants change in some cycles. The difference between canonical and variegated babble amounts to gaining control over the selection of differentiated movement plans in each babble cycle through control over activation gradients. Co-selection of consonantal and vocalic gestures simplifies control because fewer competing activations must be maintained.

Hence babbling promotes co-selection by inducing feedback internalization and facilitates coordination by providing a cognitive cycle for phasing-based control. This results in a developmental transition from competitive selection to co-selection and coordination of a consonantal and vocalic gesture, i.e., $\{C\}\{V\} > \{CV\}$. It may also be sensible to identify earlier developmental transitions in the formation of coordinative structures of effectors associated with individual gestures, although this possibility is left for future investigation.

2.2 Competitive Control in Early Words

Non-standard productions in early words indicate that children utilize a $\{CV\}\{C\}$ mixed control model: they competitively select coda consonantal gestures relative to a preceding vocalic gesture, but co-select an onset consonantal gesture with a vocalic gesture. Furthermore, non-standard productions in consonant clusters suggest that children employ $\{C\}\{CV\}$ and $\{V\}\{C\}\{C\}$ models: they competitively select the initial consonant of an onset cluster as well as consonants in coda clusters. Because of this onset/coda control asymmetry, many of the non-standard production patterns differentially involve onsets and codas. Analysis of these patterns indicates that most of them can be best understood to result from the interaction between two factors: overgeneralization of $\{CV\}$ co-selection learned in babble and deficient control over activation/suppression of gestures.

Non-standard productions 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). Many such patterns are likely attributable to immature muscular control, perceptibility, phonotactic predictability, and genetic and environmental influences (Edwards & Beckman, 2008; Hoff & Shatz, 2009), and for these reasons certain sounds—i.e. fricatives, rhotics, and laterals—tend to be acquired later than others (McLeod, 2007; Rose & Inkelas, 2011). More relevant to the current theory are nonstandard sequencing patterns, which differentially involve onset and coda consonants. Typically in their second year children transition from babbling to producing words with simple (C)V and (C)VC syllable shapes, and consonant clusters begin to emerge around 2 y.o. (McLeod, Doorn, & Reed, 2001); children are observed to produce many non-standard sequencing patterns during these stages, which are listed in Table 1 below.

Table 1. Nonstandard sequencing patterns in singletons and clusters

	<i>singleton consonants</i>		<i>consonant clusters</i>	
	<i>onset</i>	<i>coda</i>	<i>onset</i>	<i>coda</i>
omission	/dag/ → [ag]	/dag/ → [da]	/blu:/ → [bu:]	/mæsk/ → [mæk]
substitution	[gag]	[dau]	[bwu:]	[mæfk]
compensatory lengthening	?	[da:]	?	[mæs:]
reduplication	?	[dada]		[mæk:]
V epenthesis	?	[daga]	[bəlu:]	?
C lengthening	?	[dag: ^h]		
V truncation	?	/di:d/ → [did]		
local metathesis	-	-	?	[mæks]
transvowel metathesis	-	-	[lu:b]	?
fusion	-	-	[fu:]	?

? indicates pattern is rare or not attested; - indicates pattern is not applicable

Omission and substitution are common non-standard productions in singletons and clusters. They tend to precede other non-standard patterns in the course of development, and omissions diminish as substitutions and other patterns increase (Elbert & McReynolds, 1979; Greenlee, 1974; Smit, 1993). Omission involves the failure to select a gesture and substitution involves the selection of an alternative gesture instead of the target; these occurrences are expected in both onsets and codas and hence do not distinguish control regimes. Substitution is most commonly observed with difficult segments, and children can be fairly consistent in which consonants are substituted; onset cluster omissions exhibit more regularity in which cluster member is retained than coda cluster omissions (Bernhardt & Stemberger, 1998; Locke & Studdert-Kennedy, 1983; Smit, 1993). While this may be attributable to sonority (Gierut, 1993) or language-specific phonotactic frequency, it can also be attributed to co-selectional control learned in babbling: stops and nasals are more prevalent in babble than fricatives or liquids, and thus more likely to be retained in clusters.

One of the most striking aspects of nonstandard sequencing patterns is that many of the patterns which involve singleton codas do not occur with onsets. 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).

The onset/coda asymmetry follows naturally from errorful reselection in a {CV}{C} mixed control model. Errorful reselection is attributable to deficient activation and/or suppression of gestures, and many nonstandard productions additionally involve overgeneralization of previously acquired co-selectional control. For example, compensatory vowel lengthening is hypothesized to involve a reselection of the vocalic gesture that occurs in combination with the failure to select the coda consonantal articulation, i.e. {CV}{C} → {CV'}{V'}. This pattern can arise if the vocalic gesture is not sufficiently suppressed, or if the coda consonantal gesture is not

sufficiently active to outcompete the vocalic gesture for selection (see Fig. 4). The reduplication pattern is closely related, except that {CV} co-selection is overgeneralized, resulting in reselection of both the consonantal and vocalic gesture, i.e. {CV}{C} → {C'V'}{C'V'}. Overgeneralization amounts to deficient suppression of both the onset and vocalic gestures. Similarly, vowel epenthesis and reduplication arise from overgeneralization of {CV} co-selection, i.e. {CV}{C} → {CV'}{CV'} and {CV}{C} → {CV}{C'}, the main difference between these being whether the preceding vocalic gesture is reselected with the coda or whether an alternative vocalic gesture is co-selected with the coda.

Truncation of a preceding long vowel or diphthong is hypothesized to arise from a deficiency in co-excitation of multiple plans. Long vowels and diphthongs are assumed to involve two selection events: reselection for the long vowel, i.e. {V'}{V'}, and selection of a second vocalic target in the diphthong, i.e. {V}{V}. Hence the shortening pattern can be understood to arise from selection of the coda consonant at the expense of the second vocalic selection, i.e. {V}{V}{C} → {V}[V]{C}.

Compensatory vowel lengthening and vowel truncation patterns have been taken to indicate that children have a bimoraic minimal word template because the productions appear to result in a binary foot (Demuth & Fee, 1995; Demuth, 1995). Further evidence in support of this notion is a common truncation of trisyllabic forms (/banana/ as [nana]) and the observation that long vowels and codas are acquired contemporaneously (Kehoe & Stoel-Gammon, 2001). For these reasons, the patterns have been interpreted in an optimality theoretic framework to arise because children have a highly-ranked constraint to produce a bimoraic word (C. Levelt, Schiller, & Levelt, 2000; C. Levelt & Van de Vijver, 2004).

However, the bimoraic template account is contradicted by observations of subminimal CV productions and by compensatory lengthening for omitted codas in both CVC and CVVC syllables. Song & Demuth (2008) observed that compensatory lengthening occurs with both short and long vowels, and concluded that it must be due to a timing slot associated with the coda, rather than adherence to a bimoraic template. Moreover, the consonant lengthening pattern constitutes a violation of a bimoraic constraint; this can be understood as errorful reselection of a coda gesture.

Although the evidence does not favor a strict constraint on the number of selection events, there are tendencies for coda consonants to be acquired earlier and produced more frequently after short or lax vowels than after long tense vowels or diphthongs (Demuth et al., 2006; Kehoe & Stoel-Gammon, 2001). This is a consequence of the number of selection events required for a word: {V}{V}{C} and {V'}{V'}{C} targets involve the three selection events, whereas {V}{C} structures involve only two. Co-exciting and maintaining an appropriate activation gradient among three sets of gestures is relatively difficult presumably because children have more limited control over relative activation, restricting the number of elements they can sequence.

Asymmetries in nonstandard productions of target clusters are more complicated and less extensively studied, but the patterns are nonetheless amenable to understanding with {C}{CV} and {V}{C}{C} mixed control models. Epenthesis and fusion are more commonly associated with onset clusters than coda clusters (Bernhardt & Stemberger, 1998). The {C}{CV} model predicts that vowel epenthesis may occur after the first consonant in an onset cluster due to overgeneralization of {CV} co-selection. Vowel epenthesis in coda clusters is also predicted to occur, but no examples were identified in the literature. It may be that once children have learned competitive control over singleton codas, their productions of post-vocalic consonants are highly biased against overgeneralized {CV} co-selection. Fusions, where an onset cluster appears to be reduced to a single consonant that combines features of both segments, can be viewed as attempted co-selection of onset consonants but with omission of some gestures or non-canonical coordination. The hypothesized {C}{CV} and {V}{C}{C} control models account for why fusion is restricted to onset clusters: in early words children employ co-selection only for onset

consonants and vowels, hence their attempts at co-selection should occur primarily with onset clusters.

In contrast to vowel epenthesis and fusion, compensatory lengthening and metathesis occur in coda clusters but not within onset clusters (Bernhardt & Stemberger, 1998; Fikkert, 1994; Kirk & Demuth, 2005). The asymmetry follows naturally from the $\{C\}\{CV\}$ and $\{V\}\{C\}\{C\}$ mixed control models: because the prevocalic consonant is co-selected with the vowel, metathesis of onset consonants would violate the integrity of the $\{CV\}$ selection set. Instead, the model predicts rare transvowel metathesis pattern $\{C_1\}\{C_2V\} \rightarrow \{C_2V\}\{C_1\}$ (Jaeger, 2013; Menn, 1976), which always involves transposition of the initial consonant rather than the pre-vocalic one. In coda clusters metathesis transposes selection sets, i.e. $\{V\}\{C_1\}\{C_2\} \rightarrow \{V\}\{C_2\}\{C_1\}$, and hence can occur. Along these line, compensatory lengthening of a prevocalic consonant does not occur because there is no preceding gesture in the word to be reselected, but is possible for either consonant in a coda cluster.

The onset/coda asymmetries observed in non-standard production patterns cannot be readily accounted for by a purely coordinative model, which posits $\{CCV\}$, $\{VC\}$, and $\{VCC\}$ control regimes. The purely coordinative model must associate them with anti-phase coupling in some way. Yet the reason why anti-phase coupling results in these patterns is unclear. Specifically, it does not provide a mechanism for inclusion of a gesture that was not otherwise present in the target, nor does it provide an account of why a variety of patterns tend to involve the inclusion of one gesture at the expense of another. By incorporating selectional and coordinative aspects of control, mixed control models provide a more direct and comprehensive account in which deficient activation and suppression of gestural plans, along with overgeneralization of $\{CV\}$ co-selection, are implicated. The compensatory lengthening, reduplication, vowel epenthesis, and consonant lengthening patterns are thus more straightforwardly conceptualized to result from errorful competitive selection of sets of gestures.

2.3 The Development of Coordinative Control in Early Words

Non-standard production patterns in early words suggest that there is a developmental stage in which children employ $\{C\}\{CV\}$ and $\{V\}\{C\}\{C\}$ control regimes, as well as $\{V\}\{V\}$ and reselectional $\{V'\}\{V'\}$ regimes for diphthongs and long vowels, respectively. Later in development children may, depending on language-specific factors, acquire coordinative regimes such as $\{VV\}$, $\{VC\}$, $\{CCV\}$, and $\{VCC\}$. In other words, children may develop coordinative control over gestures associated with mora- or syllable-sized domains. Progressions from competitive to coordinative control are expected to be accompanied by increased gestural overlap and hence predict decreased segmental durations and more extensive coarticulation.

A number of studies have indeed found that young children produce longer and more variable segmental durations than adults, and segmental duration and variability have in general been found to continue decreasing to adult norms as late as 10-12 years of age (Kent & Forner, 1980; Kent, 1976; Smith, Kenney, & Hussain, 1996; Smith, 1992; Vihman & Velleman, 2000; Weismer, 1984). The reductions in duration and variability that occur over the broad timespan from the onset of word production to adulthood are usually attributed to non-specific “maturation” or improved motor control.

In contrast, selection-coordination theory advocates a more nuanced view in which internalization of feedback is a relatively abrupt and gesture-specific development, with later changes arising from prolonged refinement of coordination. Hence this view elaborates on the observation that “young children show an underlying tendency to produce speech segment by segment and...this tendency decreases with age” (Katz, Kripke, & Tallal, 1991). Co-selection regimes can be learned fairly early and are specific to certain combinations of gestures, yet the period in which coordinative control is refined may be fairly extended. Because some $\{CV\}$ control structures are acquired in babbling, the durations of onset consonants in those structures

should approximate adult norms earlier than those of vowels, coda consonants, or consonant clusters. Tests of this prediction are complicated by the fact that young children are not amenable to the relatively invasive methods required to obtain articulatory data. Moreover, there are relatively few longitudinal acoustic studies, and most of them are either sparsely sampled in time or conflate onsets and codas as well as different types of gestures.

Studies of coarticulation, on the other hand, show that adult-like CV coarticulation appears to emerge earlier in development than adult-like VC coarticulation. Several studies have found that CV coarticulation is similar to adults in children 3 years old or younger (Katz & Bharadwaj, 2001; Repp, 1986; Sussman, Duder, Dalston, & Cacciato, 1999; Turnbaugh, Hoffman, Daniloff, & Absher, 1985). Some studies have in fact found evidence for greater coarticulation of CV in younger children (Goodell & Studdert-Kennedy, 1993; Nittrouer, Studdert-Kennedy, & McGowan, 1989; Nittrouer, Studdert-Kennedy, & Neely, 1996); this is consistent with the notion that developmental changes subsequent to the acquisition of {CV} co-selection primarily involve refinement of coordination. In contrast, development of adult-like VC coarticulation and timing emerges relatively late. Kent (1983) found little evidence for anticipatory coarticulation in VC syllables in 4-year-old children, and Hawkins (1984) observed that while most children attain adult-like accuracy in codas by 3 or 4 y.o., the articulations are not timed correctly. Along these same lines, Goodell & Studdert-Kennedy (1993) observed a greater degree of CV coarticulation than VC coarticulation in əCV sequences. The developmental discrepancy between CV and VC coarticulation supports the notion that acquisition of {CV} co-selection precedes {VC} co-selection; since children eventually do acquire adult-like VC coarticulation we can infer that feedback control over VC gestures is internalized later on.

Another form of evidence for a progression from competitive to coordinative control involves a common developmental pattern associated with long vowels, geminate consonants, and diphthongs. Early on the durations of long vowels, for example, are highly variable but 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 shorten through subsequent refinement (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). A key observation regarding long segments and diphthongs is that once a length contrast is correctly produced, the duration ratio between long and short segments may reduce over a prolonged period of development; this is precisely what the internalization model predicts.

In sum, the evidence considered in this section indicates that over the course of development children acquire more complex competitive control regimes and may develop co-selectional/coordinative control. In babbling children learn to co-select and coordinate C and V gestures. Non-standard production patterns argue for a developmental stage in which codas and non-prevocalic consonants are competitively selected, but subsequent reductions in segmental durations and increases in coarticulation support the notion that children internalize feedback and may learn regimes of coordinative control over these structures. A consequence of the general developmental trend to co-select larger sets of gestures is that phonological units such as segments, moras, and syllables can be reinterpreted as differently-sized instantiations of a more general type of motor planning unit. Below we explore this idea further.

3. Segments as Co-Selection Sets

Selection-coordination theory invites us to reconceptualize the theoretical construct of the “segment”. Instead of corresponding to a distinct level of motor organization, segments are viewed as instantiations of a more general type of motor unit, a *co-selection set*. Other sorts of

units such as moras and syllables are likewise viewed as co-selection sets. In support of this, several arguments are presented in this section. First, the sets of gestures commonly referred to as segments are a typologically diverse group of entities from a motoric and developmental perspective. Second, gestures associated with segments may decohere, particularly in contexts where a mixed control predicts them to. Third, although segment-sized sequencing errors have traditionally provided the primary argument for associating segments with a level of motor organization, such errors are not unique to segments and hence suggest that they arise from a more general mechanism that affects co-selection sets that are both smaller and larger than those associated with segments.

3.1 Segments as Sets of Co-Selected Gestures

A co-selection set is a collection of gestures which are selected at the same time due to an absence of competitive gating. There are several important characteristics of co-selection sets. First, in theory they have no maximal size, although there must be functional restrictions on their sizes due to cognitive limitations. In early development children acquire {CV} co-selection sets, later on children may acquire larger sets such as {CVC}, {CCV}, {CVCC}, etc., depending on external factors associated with their language environment. "Size" here refers to cardinality, the number of gestures that are co-selected. Second, the minimal size of a co-selection set is one gesture; there is no apparent motivation for positing an empty co-selection set. Third, there can be "internal structure" within a co-selection set, i.e. there are subsets associated with gradient variation in parameterized associations (coupling) among gestures. Speakers learn these parameters from statistical patterns in the linguistic environment and the patterns are constrained by perceptual and motoric factors. In adult speech segments can thus be reconceptualized as co-selection subsets, i.e., segments may correspond to relatively strongly associated co-selected gestures, but they are not necessarily the maximal domain of co-selection. Fourth, co-selection sets characterize control regimes employed in specific utterances; when we refer to co-selection sets outside of a particular speech act, we are referring in an indirect manner to behaviors that result from learned patterns of interaction between cognitive representations of gestures.

Most current phonological theories view segments as a level of organization in a hierarchical structure which dissociates the featural content of segments, segments, and timing slots. Whereas Chomsky & Halle (1968: 164-165) viewed segments as bundles of features, more sophisticated representations subsequently proposed a hierarchical organization of features; in these representations there is topmost level of the feature hierarchy called the *root node* such that root nodes correspond to segments and are arranged on a root node tier (Clements & Keyser, 1983; Clements, 1985; Mohanan, 1982). The root node has the property of "dominating all features... [and] expresses the coherence of the 'melodic' segment as a phonological unit" (Clements & Hume, 1995). Hierarchically organized feature representations were developed in parallel with models proposing an autosegmental level of representation, where root nodes link directly to C and V timing slots (McCarthy, 1979, 1988) or more generic X timing slots (Levin, 1985; Lowenstamm & Kaye, 1986). Around the same time, an alternative class of representations was developed in which segments are associated with moras (Hyman, 1984, 1985; McCarthy & Prince, 1986), which are also viewed as timing slots.

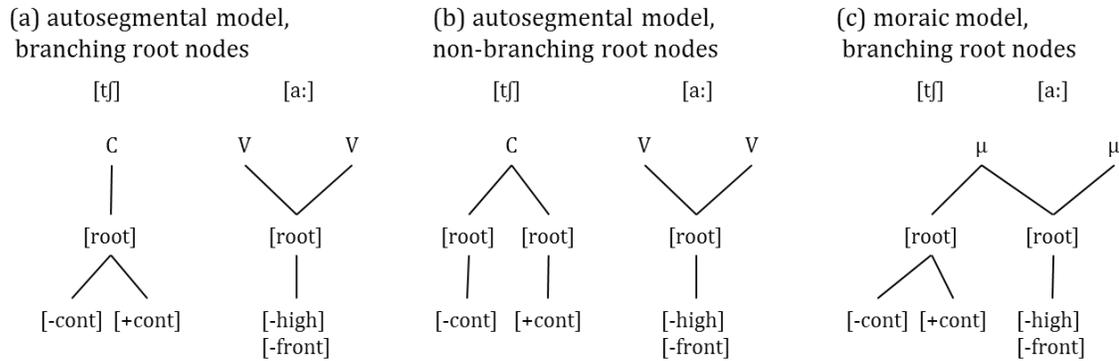


Fig. 8. Timing slot representations of phonological structure. (a) a branching root node associates two conflicting feature values with a single segment, and a root node is associated with two autosegments. (b) two root nodes are associated with a single autosegment. (c) two root nodes are associated with one mora and a root node is associated with two moras.

Feature geometry and autosegmental/moraic timing slot representations dissociate segmental content and timing from segments, thereby allowing for various non-isomorphisms, as shown in Fig. 8. One non-isomorphism is where multiple features are associated with a single root node, which corresponds to contour segments such as affricates or pre-/post-nasalized stops (Lombardi, 1990; Sagey, 1986). Alternatively, contour segments have been analyzed as an non-isomorphism between root nodes and timing slots (Clements, 1987; Piggott, 1988; Rosenthal, 1992). Another non-isomorphism is where a single root node is associated with multiple timing slots (long vowels in both autosegmental and moraic representations). A third is where multiple root nodes are associated with a single timing slot (the onset and vowel in the moraic representation).

The dissociation of featural content and timing slots from segments creates a dilemma for the concept of the segment. Length and contourality cannot be defined solely with the segmental tier; rather, length and contourality are configurations in which a root node is associated with multiple timing slots or conflicting features. Although phonological theories describe root nodes as the representational manifestation of segments, the structures required to classify segments (as contour, complex, long, etc.) require a representation of more than just the root nodes themselves. This results in a conflict between theory and practice: in theory there is no single level of structure that allows segments to be classified, yet most researchers treat segments as coherent structural entities and most production models have imbued them with a special status in motor planning (e.g. Bohland et al., 2010; Dell, 1986; Garrett, 1980; W. Levelt, 1993).

Articulatory phonology circumvents this problem by reconceptualizing what the features are and how they interact with each other. Gestures supplant features and segments are viewed as constellations of gestures. Moreover, constellations may be comprised of fewer or more gestures than are associated with an individual segment. As Browman & Goldstein (1990) state: “we would 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 “in general, however, the interdependencies among gestures are not restricted to those that constitute single segments in traditional approaches. Rather, 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” (Browman & Goldstein, 1986: 225). Descriptions of articulatory phonology have not made explicit what the maximal size of a constellation is: often the assertion is made that an utterances are constellations of interacting gestures without any additional structure. If one insists on a notion of a segment or mora in this framework, these would be subparts of constellations in which gestures may interact relatively strongly. Yet the spirit of articulatory phonology has never been to hierarchicalize the structure.

If the segment does not play a special role in speech production it remains possible that the segment functions as a *perceptual* unit. But there is no unequivocal evidence for segments in perceptual studies. It has been observed that non-linguistic auditory stimuli are perceived more accurately when occurring at segment boundaries than when occurring within segments (Fodor & Bever, 1965), but this suggests that the transitions between segments are attended to more closely than steady-states within them. Indeed, many studies indicate that units larger than segments or the transitions between segments play an most important role in speech perception (Klatt, 1979; Stevens, 1981). Ohala (1992) provided a unique perspective, arguing that segments emerge from historical processes that coordinate gestures in order to maximize modulations of the acoustic signal, yet the units that convey information are generally transitions between steady states; only the transitions require coordination, not the steady states. The segment thus emerges as a unit for purposes of information transmission. Hence for Ohala the segment is motivated primarily by considering the necessity of mapping between gestures and perceptual representations of acoustic information.

A more controversial perspective holds that the concept of the segment is an artifact of literacy or of a cultural affinity for theories that employ formal, computational logic. Port & Leary (2005) suggest that the tendency of linguists to posit discrete symbolic units in analyzing language reflects a motivation to understand language from the perspective of discrete computations. Alternatively, literacy and alphabetic writing have been argued to create a bias for using segments in linguistic theory (Faber, 1992; Linell, 2004). In support of this it is observed that children do not acquire phonemic awareness until they are taught to read and write (Goswami & Bryant, 1990), illiterate adults tend to lack phonemic awareness (Lukatela, Carello, Shankweiler, & Liberman, 1995), and segment-based language games are attested only in languages with alphabetic writing systems (Bagemihl, 1989). Indeed, children and illiterate adults exhibit more robust awareness of syllables than segments (Ziegler & Goswami, 2005), and the earliest writing systems were syllabaries, rather than alphabets. Along these lines, Ladefoged (2001) suggested that “consonants and vowels are largely figments of our good scientific imaginations” (2001: 170), and referred to “the phonemic conspiracy” (Ladefoged, 1984), which is to say that because we *can* use segmental units, we are compelled to do so.

Rather than viewing the segment as an artifact of literacy or scientific preconceptions, here we take a more nuanced and constructive perspective: (some of) the things that we call segments emerge from the interaction of motor control and perception in the course of development as cognitively real categories. However, segments are not unique in this regard, and it is not clear that they differ essentially from other units, nor that they have any special status in adult speech. Moreover, as argued below, a theory of production should avoid lumping together all of the entities which have been called “segments”.

3.2 Segmental Diversity

Segments are a remarkably diverse group of entities from a motoric and developmental perspective. Many of the units commonly identified as “segments” in adult speech are sets of several co-selected, coordinated gestures. However, these co-selection sets differ in several important respects: when in the course of development co-selection of component gestures is acquired; the degree of phase asynchrony between the component gestures; and the control regimes that may be employed in their production. Fig. 9 presents a classification on the basis of these qualities; many less common types of segments (e.g. clicks, doubly-articulated stops) have not been studied extensively from articulatory and developmental perspectives and are not included in the classification. Phrasing is classified on the basis of behavior in syllable onsets, because (as we discuss below) the component gestures of some segments tend to decohere post-vocally.

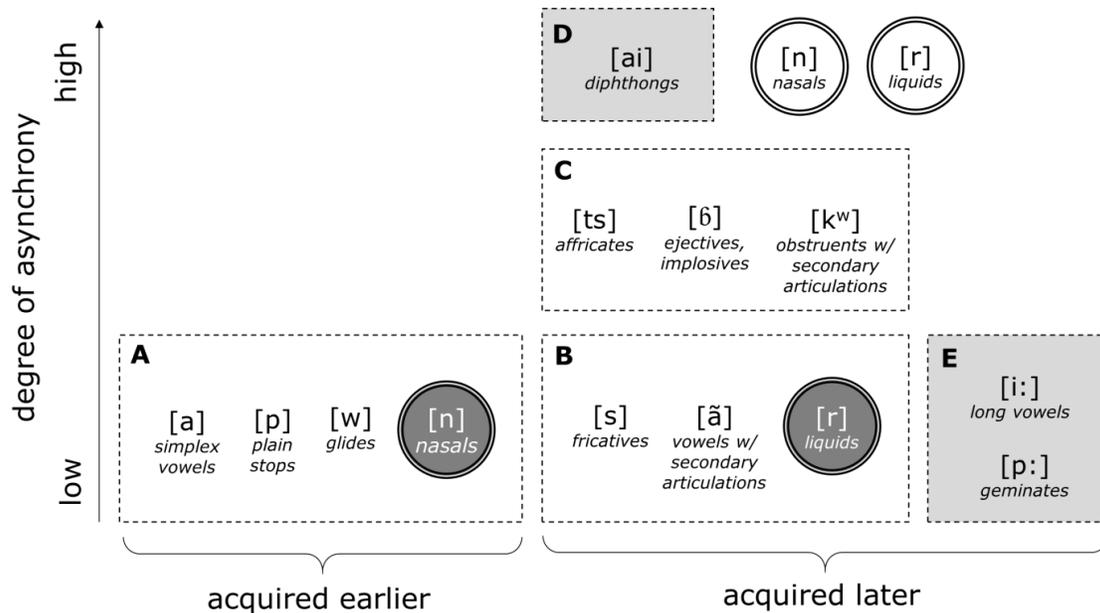


Fig. 9. Grouping of segments according to relative timing of acquisition and degree of asynchrony among component articulations. Nasals and liquids are highlighted because of their propensity to decohere post-vocally. Long segments and diphthongs are shaded because of their similarity in phonological patterning.

Class A segments (i.e., simple vowels, plain stops, nasals, and glides) are acquired early and involve highly synchronous coordinated gestures. These segments appear frequently in transcriptions of babbling, which suggests that co-selection of their component gestures is acquired relatively early. Subsequent refinement of coordinative control leads to a high degree of synchrony in adult speech. For example, in plain stops the oral occlusion and a glottal abduction gestures are initiated closely in time (Hoole, 2006; Löfqvist & Yoshioka, 1984). Similarly highly synchronous timing is observed for velar opening and oral closure in nasals (Byrd, Tobin, Bresch, & Narayanan, 2009; Krakow, 1989) and for lip rounding and tongue body retraction in labiovelar glides (Gick, 2003).

Class B segments (e.g. fricatives, liquids, vowels with secondary feature) are acquired subsequent to babbling but like class A segments exhibit highly synchronous phasing of gestures. The relatively late acquisition may be attributable to the incorporation of a greater number of oral gestures or finer control requirements on individual effectors. For example, most fricatives involve multiple gestures, such as labial rounding or retraction gestures for /s/ and /ʃ/, and require a precisely controlled constriction aperture to produce turbulent airflow. Liquids and rhotics involve tongue tip and tongue body gestures (and lip rounding for [ɹ]) that are highly synchronous in onsets (Browman & Goldstein, 1995; Gick, Campbell, Oh, & Tamburri-Watt, 2006; Proctor, 2009).

Class C segments (e.g. affricates, ejectives, implosives, consonants with secondary articulations) are acquired after babbling and involve moderately asynchronous gestures. The asynchrony serves the purpose of achieving aerodynamic and/or acoustic effects. For example, affricates are acquired subsequent to babble and involve a fully occluded constriction gesture followed by a narrow constriction gesture. These gestures cannot be too synchronous otherwise the closure would not be perceptually recoverable. Nonetheless, adults do not appear to produce errors in which the component gestures of affricates are split (Fromkin, 1971), which suggests that these gestures are co-selected. Consonants with secondary articulations (e.g.

labialized/palatalized/pre-nasalized stops, ejectives, and implosives) probably also belong to this class, although articulatory and developmental data are needed to confirm this.

Class D and E segments (i.e. long vowels, diphthongs, geminates) are particularly special because it is ambiguous whether speakers employ competitive or coordinative control over their component gestures. Developmental evidence for a competitive regime was seen in nonstandard production patterns such as compensatory lengthening, epenthesis, truncation, etc. Certain phonological patterns discussed below provide further evidence for competitive control. However, as discussed above, durational characteristics of these segments may be reflective of some degree of feedback internalization and possibly coordination. Below we argue that cross-linguistic variation in phonological behavior of these segments can be attributed to differences between competitive and coordinative control regimes.

Although not considered segments in the conventional sense, tones and pitch accents can be analyzed within the same classificatory framework. Whereas simple tones and monotonal pitch accents are akin to class B segments—they are acquired relatively early and are synchronized with vocalic gestures—contour tones and bitonal pitch accents are akin to class D segments: they are acquired relatively late and involve a high degree of asynchrony between their component gestures.

Tones and pitch accents can be conceptualized as manifestations of pitch gestures. Browman & Goldstein (1989) suggested that additional gestures for pitch control could be incorporated in to the task dynamic model. Recently, lexical tone pitch gestures in Mandarin were found to exhibit a C-center effect in conjunction with vocalic and consonantal pitch gestures (Gao, 2008), and intonational pitch gestures were found to be coordinated with vocalic gestures in German and Italian (Mucke, Nam, Hermes, & Goldstein, 2012; Niemann, Mücke, Nam, Goldstein, & Grice, 2011). The gestural conceptualization of tones and pitch accents is made possible by the autosegmental-metrical theory view that tones and pitch accents are associated with abstract target levels (Ladd, 2008; Pierrehumbert, 1980): hence pitch gestures are associated with F0 targets (which may, however, be modulated by a variety of factors). Contour tones and bitonal pitch accents are associated with two targets in the autosegmental-metrical approach and hence two pitch gestures, analogous to diphthongs. Like class D segments, contour tones have been found to be acquired later than simple ones (Li & Thompson, 1977; Tse, 1978) and phonological patterns considered below speak to an ambiguity in whether their component gestures are competitively selected or co-selected and coordinated.

The ambiguity regarding control over class D segments and contour tones/bitonal pitch accents is of particular theoretical import: it shows that from a developmental and motoric perspective, segments are a fairly diverse group of entities. The moderate degree of gestural phasing asynchrony observed in class C segments likewise further diversifies the group of entities which are conceptualized as segments. Ambiguity in control and phasing asynchrony are of course related to the non-isomorphisms between root nodes and features/timing slots discussed earlier. The diversity of segments, from a control perspective, begs the question of whether they are associated with a unique level of motor organization.

3.3 Decoherence of Gestures

Another problem with viewing segments as a unique level of motor organization involves decoherence of gestures. Decoherence refers to the propensity for a gesture that is typically highly synchronous with other co-segmental gestures to become substantially more asynchronous in some contexts. There are many examples of decoherence but a coarse analysis suggests that they fall into two main classes, according to whether they result from $\{V\}\{C\} \rightarrow \{VC\}$ or $\{C\}\{C\} \rightarrow \{CC\}$ coordinative developments. In both cases, production patterns suggest that different degrees of feedback internalization can apply to the gestural components of a co-selection set.

Decoherence of post-vocalic gestures is commonly observed in nasals and liquids. For example, in 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 decoherence has been 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). The same holds for rhotics, which involve a pharyngeal constriction gesture with a tongue tip retroflexion or bunching gesture (Gick et al., 2006; Proctor, 2009).

The decoherence patterns contrast with highly synchronous phasing in syllable onsets. In a competitive regime of control (over nasals, for example) anticipation of velum lowering associated with a coda nasal is not expected, since both gestures should not be selected until the preceding vocalic target is achieved. Internalization of feedback may allow for differential gating of the oral closure and velum lowering, resulting in velum lowering being phased earlier in the syllable. Crucially, this decoherence of coda gestures should not occur in earlier stages of development where a {V}{C} mixed control regime predominates. Indeed, the dissociated phasing of the velar lowering and oral constriction appears to be learned relatively late in development: 3-6 year old children do not exhibit substantial nasal airflow during a preceding vowel (Thompson & Hixon, 1978), indicating that the velum lowering gesture in a nasal remains coherent with the oral articulatory gesture.

Exactly why some segments but not others exhibit post-vocalic decoherence is a topic for another study. In fact not all languages exhibit the decoherence pattern for liquids and nasals. As Gick et al. (2006) suggest, biomechanical, perceptual, and motoric factors can likely explain cross-linguistic variation. Certain gestures are likely too detrimental to perception of the vocalic gesture to be phased early in the vowel. Decoherence also potentially obscures the anticipated gesture when precise timing is required for acoustic or aerodynamic targets to be obtained.

An even more widespread decoherence phenomenon relates to phonological rules in which features spread or are constrained not to differ between adjacent consonants. In many languages onset or coda consonant clusters are observed to agree in features such as place, voicing, and/or manner (Baković, 2007; Steriade, 2000, 2001). These patterns may be understood to arise from {C}{C} → {CC} developments. Although a detailed analysis of these phenomena is deferred for future studies, it is sensible to understand such assimilations to result from decoherent phasing in conjunction with this omission of one of a pair of conflicting gestures. This is consistent with the explanation of such phenomena provided by articulatory phonology, which views assimilations as the result of gestural overlap or blending of active gestures (Browman & Goldstein, 1990; Gafos, 1999, 2002). The concept of decoherence elaborates on this account by providing a more specific mechanism for how overlap can be augmented in the course of development.

Decoherence phenomena are important for understanding segmental organization because they show that the domain over which gestural phasing is controlled is often larger than a single segment. Gestures which decohere must be co-selected with gestures associated with other segments, and this ultimately raises doubt regarding whether motor planning is organized around segmental units.

3.4 Error Patterns Involving Co-Selection Sets

The most commonly cited form of evidence for positing a segmental level of organization is based on substitution, omission, addition, and transposition errors in adult speech. Many such errors appear to involve segment-sized sets of gestures, and accordingly, many production models explicitly incorporate a segmental planning level. The reasoning behind this is that if errors can involve segment-sized units, then such units must constitute a level of organization in the production mechanism. Indeed, the same reasoning has been applied subsegmentally to features, and suprasegmentally to prosodic constituents such as syllable onsets, nuclei, and rhymes.

However, the alternative perspective advanced here is that variation in the size of units affected by errors indicates that the different types of units are differently sized instantiations of the same generic type of unit, a co-selection set.

Studies of speech error patterns have found that errors involving segments, i.e. single vowels or consonants, are the most common type of error, and that errors involving only individual segments can occur within onset and coda clusters (Fromkin, 1971; W. Levelt, 1993; Shattuck-Hufnagel, 1983). These observations have been widely used as a justification for the reality of segmental units, 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). Examples of errors discussed in this section are shown in Table 3 below, which draws from examples in Fromkin (1971).

	<i>domain</i>	<i>example</i>
a)	segment	cup of coffee → cu[f] of coff <u>ee</u>
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>st</u> orm
g)	onset-vowel	pussy cat → [kæ]ssy [pɜ]t
h)	vowel-coda	a heap of junk → a h[unk] of j[ip]
i)	syllable	tremend <u>ously</u> → tremenly;
j)	foot	butter <u>fly</u> and cater <u>pillar</u> → butterpillar and caterfly

Table 3. Examples of sequencing errors affecting units of varying sizes

Sequencing errors in which individual features or gestures are unambiguously involved are not uncommon. For example, in (c) a velum lowering gesture has transposed, and in (d) oral gestures have exchanged but laryngeal ones have not. Indeed, many of the errors that are commonly identified as segmental can be alternatively interpreted as errors involving one or more individual gestures or features (Fromkin, 1971). For example, (Stemberger, 1983) analyzed "benefit *spall*—small businesses" as a segmental anticipation, but this could alternatively be viewed as omission of the velar gesture, i.e. failure to select the velar opening gesture. Detailed analyses of errors have found that the majority of "segmental" substitution errors involve just one feature, and that only a small percentage of segmental errors involve all distinctive features of the segments involved (MacKay, 1970; Nooteboom, 1973). These findings imply that many segmental errors may be more appropriately conceptualized as errors involving sets of one to several gestures.

Sequencing errors also quite frequently involve larger sets of gestures. Errors involving entire clusters are not uncommon (e, f), and they can also involve larger units incorporating a vowel as well (g, h). Such errors have been interpreted to involve groups of segments or to imply that syllable constituents are domains of control. Errors involving syllabic and suprasyllabic size units (i, j) have been interpreted to imply a syllabic domain of control.

Many models of speech production ascribe a fundamental role to the segment. Shattuck-Hufnagel (1979) used error patterns to argue for a slots and fillers model, where segments are fillers, onsets/nuclei/codas are slots, and errors arise from inaccurate mappings between the two. Levelt (1993) presented a similar view of phonetic planning as the filling of timing slots by segmental material, with the phonetic plan itself being specified as a sequence of phones (1993: 284). A phonological encoding module groups the phones into syllable-sized units and an articulatory module implements gestures with timing patterns determined by their syllabification.

Dell (1986) modeled morphemes, syllables, onsets, rimes, segments, and features as nodes as a network through which activation spreads. Bohland et al. (2010) modeled syllable positions as slots which are filled with competitively selected segments. In all of these approaches segments play a fundamental role in organizing motor planning.

One important characteristic of selection errors is that the involved co-selection sets are almost always associated with the same type of constituent. Most segmental slips obey a structural law with regard to syllable place, whereby segments associated with a particular syllable sub-constituent can interact only with others associated with the same type of sub-constituent (Fromkin, 1971; MacKay, 1970; Nooteboom, 1973). This characteristic has led to the postulation of *unit-similarity constraint*: “the intruding element is of the same level of representation and category as the target element” (Levelt 1993: 330). Pierrehumbert & Nair, (1995) showed that errors in infixation games are likewise sensitive to prosodic positions.

The unit-similarity constraint suggests that the selection mechanism is sensitive to learned coordination patterns. A gesture in onset position is “an onset gesture” by virtue of being co-selected with a vocalic gesture and phased with a particular phasing relation learned in the course of development. A gesture in a coda may likewise be co-selected with a vowel, but with a substantially different phasing relation. The observation that competitive processes responsible for errorful selection between gestures associated with different words are sensitive to these differences suggests that in adult control selection is sensitive to learned patterns of co-selection. Further evidence for this assertion involves another noteworthy constraint on errors, which is that they do not generally result in phonotactically prohibited sequences or phones that are not in the speaker's phonological inventory (Fromkin, 1971; Shattuck-Hufnagel & Klatt, 1979). For example, *play the victor* to *flay the pictor* can occur, but not *vlay the pictor*. In the latter, both the voicing feature and manner are transposed and this would result in a phonotactically prohibited sequence, [vl]. In other words, when gestures are errorfully selected or omitted, the resulting co-selection set is one that a speaker has already learned.

The errors that are used to rationalize a segmental level of motor organization are generally errors of selection, but errors may also involve nonstandard coordinative control. Most analyses of errors have relied purely on transcription and thus are biased to identify discrete, selectional errors (Frisch & Wright, 2002). Several studies have in fact found evidence for gradient sub-segmental errors. Frisch & Wright (2002) observed gradient errors in voicing of [s] and [z] in a tongue twister error elicitation paradigm. Some of these gradient errors may result from noncanonical coordination, rather than errorful selection. Indeed, gradient gestural intrusions in a repetition task suggest that errors can arise in the implementation of gestures (Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; Pouplier, Chen, Goldstein, & Byrd, 1999). Errors may also be associated with activation of individual muscles (Mowrey & MacKay, 1990).

The categorical error patterns reviewed above have been interpreted as evidence for hierarchically structured “types” of constituents, including a segmental level of organization. However, many of the errors associated with segments involve just one or two features, and analogous types of errors (i.e. substitution, omission, transposition) occur with sets of gestures larger than a segment. Selection-coordination theory offers an alternative interpretation of the error patterns: similar errors affect gestures, segments, syllable sub-constituents, and syllables because these units are in essence the same type of motor planning unit, a co-selection set. The occurrence of errors involving segment-sized co-selection sets does not entail that segments are the maximal domain of selection or even a domain around which planning is typically organized.

4. Moras as Co-Selection Sets

Selection-coordination theory offers a more explanatory understanding of phonetic and phonological patterns associated with moras by viewing them as co-selection sets. As with segments, this does not necessarily imply that the mora is the maximum domain of co-selection or

even the most common control strategy in adult speech. However, there exist numerous phonological patterns which are best described with moras, and moraic organization corresponds to $\{V\}\{C\}$ and $\{V\}\{V\}$ control regimes hypothesized to emerge in development. The resemblance between the mixed control regimes and moraic representation suggests that moraic phonological patterns may be associated with a developmental stage in which control regimes are organized predominantly around mora-sized co-selection sets.

The concept of the mora was adopted from metrics into phonological theory as an alternative to autosegmental representations (Hayes, 1989; Hyman, 1984, 1985; McCarthy & Prince, 1986). Moras have been used to account for cross-linguistic patterns in stress assignment, using the concept of syllable weight: light syllables consist of one mora, heavy syllables two, superheavy syllables three, as shown in Fig. 10. A key distinction is drawn between languages in which coda consonants are moraic (c) and those in which codas are non-moraic (b). In the former (e.g. Hindi, Latin) a syllable with a coda is phonologically heavy and attracts stress; in the latter (e.g. Malayalam, Mongolian) a syllable with a coda is phonologically light and does not attract stress because the coda shares a mora with a preceding vowel. Moraic representations have been employed to account for a variety of other phonological patterns: prosodic minimality and maximality constraints, compensatory lengthening, and constraints on contour tones. Below we argue that these diverse phenomena can be best understood to arise from competitive control over mora-sized co-selection sets, and that the distinction between moraic and non-moraic codas is associated with the distinction between competitive $\{V\}\{C\}$ and coordination $\{VC\}$ control.

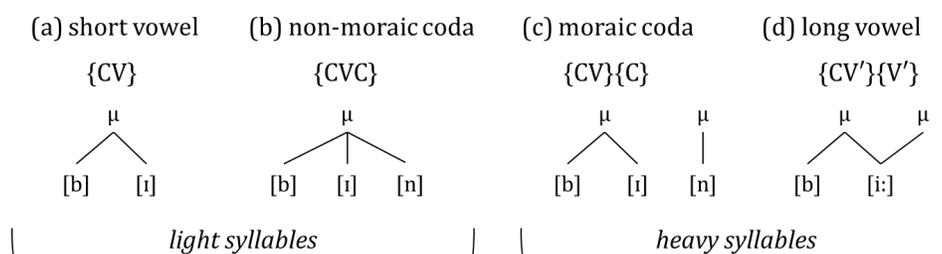


Fig. 10. Moraic representations of various syllables. (a) a short vowel is associated with one mora. (b) a short vowel and non-moraic coda are associated with one mora; the coda is co-selected with the vocalic gesture. (c) a short vowel and moraic coda are associated with two moras; the coda is competitively selected relative to the vocalic gesture. (d) a long vowel is a vowel associated with two moras; the vocalic gesture is reselected.

4.1 Phonetic Evidence for Moras as Co-Selection Sets

The distinction between moraic and non-moraic codas (Fig. 10 b vs. c) has a straightforward interpretation: moraic codas correspond to a $\{V\}\{C\}$ mixed control regime, while non-moraic codas correspond to $\{VC\}$ co-selection regime. Hence “sharing” a mora with a vowel amounts to being co-selected with a vowel. The presence or absence of a competitively selected coda should have a relatively small effect on the duration of a preceding vocalic gesture. This is because the time-courses of selection and deselection of the preceding vocalic gesture will not be greatly affected by any co-selection sets that are subsequently competitively selected. In contrast, the $\{VC\}$ co-selection model predicts a reduction in the acoustic duration of the vocalic gesture because the coda can be selected prior to the deselection of the vocalic gesture and hence the two gestures will overlap. The period of time during which the vocalic gesture dominates control over movement is truncated in this coordinative regime.

Phonetic evidence indeed supports distinguishing between moraic and non-moraic codas (cf. Cohn, 2003 for a review). Comparison of vowel and consonant durations in Hindi and Malayalam shows that the durations of mora-sharing vowels in Malayalam are significantly shorter than

vowels that do not share a mora, whereas no similar effects were observed in Hindi, where the coda consonant is associated with its own mora (Broselow, Chen, & Huffman, 1997). Syllables with weight-bearing codas are longer in Mandarin Chinese than in Shanghai Chinese, where they are phonologically weightless (Duanmu, 1994). Weight distinctions also apply to geminate codas, where a consonantal gesture may share a mora with a preceding vowel or may be associated with its own mora. For instance, the geminate-to-singleton duration ratio is relatively small in Madurese and Bernese (≈ 1.5), but relatively large in Levantine and Hungarian (≈ 2.0) (Ham, 2001). In the languages with a relatively small ratio, coda gestures are co-selected with a preceding vocalic gesture and subsequently re-selected; in the languages with a larger ratio, the coda gesture is competitively selected relative to the preceding vowel and is subsequently reselected.

An important consequence of viewing moras as a representational expression of competitively selected co-selection sets is that they cannot be viewed as "timing units" and are not predicted to be isochronous. The time-courses of selection and deselection processes will differ from mora to mora (i.e. from co-selection set to co-selection set) because feedback processes may differ according to the gestural composition of mora-sized co-selection sets. Selection processes do not govern timing directly, rather they govern the ordering and choice of which gestures to produce.

Studies of rhythmic typology have nonetheless suggested that there are languages in which timing is regulated by a tendency to isochronize moras, Japanese being the parade example. But investigations of mora duration have not found solid evidence for isochrony of moras (Beckman, 1982; Hoequist, 1983; Warner & Arai, 2001), instead finding substantial variation in mora duration with segmental composition being the strongest predictor. For example, Hoequist (1983) observed that CVN syllables are significantly longer than CVV syllables in Japanese. A number of studies have found that the count of moras in a word has a linear relation with the duration of the word (Hubbard, 1995; Nagano-Madsen, 1992; Port, Dalby, & O'Dell, 1987). This sort of finding seems to provide evidence for isochrony, but in actuality is more likely a statistical artifact: the durations of processes involved in the production of mora-sized competitively selected co-selection sets may tend toward a common value, particularly when they represent units with similar gestural compositions. To wit, in the Port et al. (1987) study, the most linear relations demonstrated between mora count and word duration were derived from words with only open syllables; more substantial departures from the linear relation were associated with comparisons between words with and without codas.

Assessing the mora-timing hypothesis by examining correlations among segment durations within and between moras also suffers from methodological issues. In a CVC syllable where the coda is moraic, the timing unit view predicts negatively correlated onset and vowel durations, and the absence of correlation between the coda and vowel. Such correlation patterns have been observed in some studies of Japanese (Campbell & Sagisaka, 1991; Sagisaka & Tohkura, 1984). However, Beckman (1982) observed a negative CV correlation only in some pairs of CV, and other studies have found negative correlations where they are not predicted, between a moraic nasal and the preceding vowel (Campbell, 1999; Y. Sato, 1993). In any case, there are serious methodological issues with such analyses (Warner & Arai, 2001): inaccuracy in segmentation will induce spurious negative correlations between adjacent intervals, and the inaccuracy is amplified by the fact that acoustic intervals provide only an indirect representation of articulatory timing.

4.2 Constraints on Contour Tones

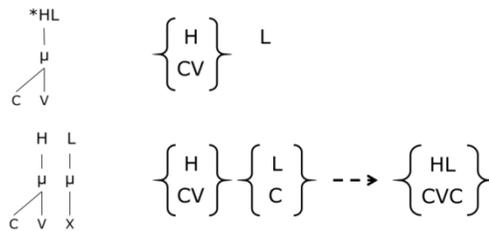
Contour tones are often constrained to occur only in heavy syllables, or only in syllables with long vowels or sonorant codas. Hyman (1985) proposed that individual tones associate with moras, and hence that restrictions on contour tones follow from the availability of moras for tones

to associate with. An alternative, perceptual explanation for contour tone restrictions is based on the notion that a longer and louder interval of voicing promotes the perception of tonal information (Gordon, 2002). Hence long vowels (VV), vowel-sonorant (VR) sequences, vowel-obstruent (VO) sequences, and short vowels (CV) provide decreasingly suitable environments for perception of F0 changes. In support of this, Gordon (2002) found an implicational hierarchy of contour tone restriction: VV >> VR >> VO ~ V, from least to most restrictive.

A big problem with the perceptibility account is that it does not explain why sonorant onset consonants fail to influence tone-bearing capacity. If the phonetic motivation for constraints on contour tones derives primarily from the duration and salience of F0 information, then sonorant onsets should contribute to systematic constraints on contour tones as well. Furthermore, there are languages which do not conform to the perceptibility hierarchy by virtue of restricting contour tones on VO but not V (e.g. Cantonese, Vietnamese, Maru), as well as languages that restrict contour tones on V but not VO (e.g. Hausa, Musey, and Luganda)—differential restrictions between V and VO are not expected on the basis of perceptibility.

The resolution of these issues in a selection-coordination framework follows from viewing lexical tones as gestures. In section 3.2 we cited evidence that contour tones consist of two (or sometimes three) tonal gestures which are coordinated with vocalic gestures. Restrictions on contour tones can be interpreted to arise from a developmental bias attributable to co-selection of tonal gestures with other articulatory gestures. In a syllable produced with two co-selection sets, the first and second gestures of a contour tone may be selected independently with the first and second co-selection sets, respectively; the tonal gestures need not be coordinated with one another in this case. In contrast, in a syllable produced with only one co-selection set, the tonal gestures must be co-selected and coordinated. These explanations are schematized in Fig. 11A. Given the developmental precedence of competitive selection over co-selection, production of contour tones should be more difficult to learn in syllables with only one co-selection set.

(A) contour tone restrictions



(B) quantity-sensitive stress

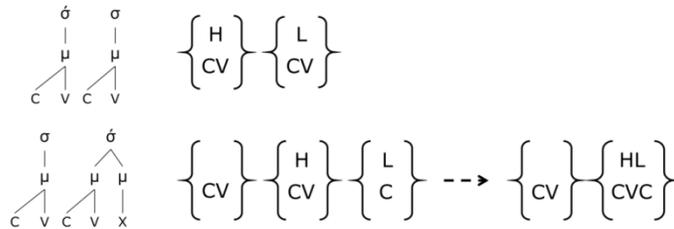


Fig. 11. Interpretation of contour tone restrictions (A) and quantity sensitivity (B) in the selection-coordination framework. H and L represent tone/pitch accent gestures, which must be co-selected with consonantal or vocalic gestures.

This account predicts exceptions to the perceptibility-based hierarchy and explains why onsets play little role in constraining contour tones: unlike coda consonantal gestures, onset consonantal gestures are co-selected with vocalic gestures, and hence do not provide an additional

co-selection set for a tonal gesture. The account does not preclude a co-selectional regime of control, however: contour tones may be acquired first in the contexts where selection is facilitated due to the presence of multiple selection sets, and their pitch gestures may subsequently become co-selected and coordinated through internalization.

4.3 Quantity-Sensitive Stress and Extraprosodicity

In quantity-sensitive stress systems, syllable weight plays a role in determining stress patterns (cf. Zec, 2003, 2007). In English, for example, the basic stress pattern in nouns locates primary stress on the penultimate syllable (e.g. A.la.'ba.ma, ba.'na.na). However, when the final syllable has a long vowel or coda, it can receive primary stress (e.g. ka.'zoo, ca.'det). The standard account of this pattern involves associating the final coda with its own mora, and assigning stress to the rightmost heavy syllable. The phonetic explanation offered for why heavy syllables attract stress usually resorts to perception; for example, Gordon (2002) associates weight with "total perceptual energy," which is the integration of loudness over time.

One problem with the perceptual account is that there are no cross-linguistically robust phonetic correlates of stress; the only common feature of stressed syllables is their potential to bear a pitch accent. Increased duration, intensity, and pitch excursion size are often erroneously cited as correlates of stress, yet these associations are indirect and language-specific (Hayes, 1995). Many studies have conflated phonetic effects of pitch accent with those of stress (Fry, 1955), but in the absence of pitch accents, stressed syllables do not seem to have additional duration or intensity (Campbell & Beckman, 1997). Pitch accents are generally understood as independent from stress (Beckman & Edwards, 1994; Bolinger, 1985; Gussenhoven, 2004) and stress has therefore been interpreted as a manifestation of a structural configuration rather than a feature associated with articulatory/perceptual content (Lieberman & Prince, 1977; Selkirk, 1980). The one thing that stressed syllables do have in common across languages is a *potential* to be associated with intonational pitch accents. Indeed, three of the four diagnostics Hayes (1995) proposed for identifying a stressed syllable reduce to the potential to bear a pitch accent.

Parallel to the selection-coordination account of contour tone restrictions, quantity sensitivity can be understood to arise because selection of bitonal pitch accent gestures is facilitated in syllables with two co-selection sets. The presence of an additional selection event associated with a long vowel or coda consonant enables the selection of the second pitch gesture of a bitonal accent, as schematized in Fig. 11B. In other words, heavy syllables "attract" stress because it is easier to select each of two pitch gestures independently with each of two co-selection sets. Quantity sensitivity can thus be understood to result from facilitated selection of pitch gestures in {V}{V} and {V}{C} control regimes.

In some languages word-final consonants are analyzed as extraprosodic, i.e. they appear to be "outside" of the prosodic structure of syllables in that they do not contribute to weight or influence stress patterns. For example, if the final consonants in the words "syrup", "buttock", "vomit", and "tempest" were associated with a mora, quantity sensitivity would result in word-final stress. Hence these consonants are treated as non-moraic in standard phonological representations. In some languages extraprosodicity is fairly uniform and predictable; in others such as English, etymological, syntactic, and articulatory factors are involved and extraprosodicity is ultimately lexically idiosyncratic (Hayes, 1980, 1982; Ross, 1969). Extraprosodicity may arise when a coordinative control strategy for word-final gestures dominates over a competitive one, or when a coordinative strategy is learned relatively early in development. In these circumstances, only one co-selection set is present in a word-final syllable and selection of the second pitch gesture in bitonal pitch accent is not facilitated.

4.4 Conservation and Enumeration of Selection Events

Compensatory lengthening and prosodic minimality/maximality constraints reflect the predisposition of speakers to conserve or count the number of selection events in a word. Compensatory lengthening is the lengthening of one segment occurring in conjunction with the deletion or shortening of another. Crucially, the pattern never results from the omission of an onset consonant (Hayes, 1989; Hyman, 1984). Even in languages where both onset and coda consonants undergo deletion, the pattern is only observed with codas (e.g., [v] in Turkish: Sezer, 1986). Furthermore, compensatory lengthening is only observed in languages with a vowel length distinction (De Chene & Anderson, 1979). Hayes (1989) discusses several different types of compensatory lengthening patterns. One is an omission-lengthening pattern, where a coda consonant is omitted and the preceding vowel is lengthened, as in Latin *kasnus > ka:nus (Ingria, 1980), or where one coda in a cluster is lost and the other geminates. Another type of compensatory lengthening which Hayes (1989) calls "double flop" involves the lengthening of a vowel that precedes a deleted consonant with another consonant intervening, e.g. Greek *odwos > o:dos (Steriade, 1982). A related type of compensatory lengthening involves the lengthening of a vowel when a following vowel, usually word-final, is deleted, as in Middle English *tale > ta:l.

Although omission-lengthening and double-flop patterns are superficially somewhat different, the basic result of their manifestation is the same: the number of selection sets is conserved. In the omission-lengthening pattern, a coda co-selection set fails to be selected but a co-active gesture—often the preceding vowel—is reselected (cf. Fig. 4). In the double-flop pattern, a vocalic gesture is reselected and what was formerly the coda consonantal gesture is integrated with a subsequent co-selection set, rather than omitted.

Compensatory lengthening patterns indicate a predisposition to conserve the number of selection events associated with a word. Moreover, the predisposition should be especially strong during the developmental stage in which {V}{C} competitive control predominates. Hence the origin of compensatory lengthening may be attributable to conservation of a control regime that can be developmentally transient. The account also explains why omission of onset consonants does not result in compensatory lengthening: onset omission does not affect the number of co-selection sets and hence no further reorganization of control is necessary to conserve them. It also suggests an explanation for why a vowel length distinction is a prerequisite for the emergence of the pattern: in languages where children do not learn a vocalic reselection regime (i.e. long vowels), they will not adopt reselection to conserve selection events.

Prosodic minimality and maximality constraints in which words or morphemes conform to a minimal or maximal number of co-selection sets are common. The prosodic morphology hypothesis (McCarthy & Prince, 1990) holds that reduplicative and templatic patterns are defined in terms of units of prosody: moras, syllables, feet, and words. Many languages are observed to prohibit content words that are too "small" or syllables that are too "large", and many minimality and maximality constraints are formalized as prohibitions against too few or too many moras or syllables (McCarthy & Prince, 1990). Fijian and Mohawk are examples of languages which enforce a bimoraic minimality constraint (Michelson, 1988). English also prohibits monomoraic content words: the lexicon does not contain monosyllabic content words without a long vowels or coda. When the lexicons of languages with such constraints do contain subminimal forms, various repairs are enacted to bring them up to minimal size. There are also patterns in which maximality constraints appear to restrict the number of co-selection sets: examples include Arabic noun stems (Broselow, 1992), Kinande verbal reduplication (Downing, 2000), and Maori passives (De Lacy, 2001).

Compensatory lengthening and minimality/maximality constraints are closely related in the selection-coordination framework. Whereas compensatory lengthening conserves a single selection event, typically by reselection of another gesture, minimality/maximality constraints enforce restrictions on the number of selection events associated with a prosodic or morphological domain. We need not assume that constraint-satisfaction is the cognitive mechanism through which these restrictions are enforced; rather, speakers can learn to associate

morpho-lexical domains with some minimal or maximal number of competitively selected co-selection sets. However, the cognitive mechanisms involved in these associations are unclear and hence require further investigation.

Importantly, a coordinative {VC} model cannot straightforwardly account for any of the above phonological patterns. Coda gestures are co-selected and anti-phase coordinated with vocalic gestures in the standard articulatory phonology model, but vowel length contrasts are implemented with independent parameters influencing the duration of activation intervals in the gestural score. There is no underlying unity between long vowels and vowel-coda sequences in this approach, and hence no obvious way in which the similarity in their phonological behavior can be understood. Moreover, since only one control option is available for codas in a purely coordinative model, there is no straightforward way to account for extraprosodicity. Phenomena involving conservation, enumeration, and facilitation of pitch gestures cannot be readily derived from anti-phase coupling specifications: complex onsets also have anti-phase specifications, yet these do not participate in the relevant phonological patterns.

In contrast, the {CV}{C} and {CV}{V} mixed control models provide an intuitively sensible account for why onset consonants do not participate in moraic phonological patterns: they do not contribute an additional selection event to a word. Coda consonants do contribute an additional selection event, at least during some stage of development, and hence can be associated with restrictions on contour tones, quantity sensitivity, compensatory lengthening, and minimality/maximality constraints. In general the qualitative aspects of the patterns can be represented with moraic structure, but selection-coordination theory provides a more detailed motoric and developmental account of why moras are associated with such patterns.

5. Variation in Control and Hierarchical Structure

Selection-coordination theory conceptualizes developmental change in speech motor planning as an agglomerative process whereby co-selection sets are combined into larger co-selection sets. This behavior can 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 become larger over a developmental timescale through internalization of feedback.

The hierarchical appearance of adult control structures is attributable to the protracted time-course of the agglomeration: co-selectional control is learned over an extended period of time, with larger agglomerations contingent on smaller ones. Smaller co-selection sets are acquired earlier and require less sophisticated control, hence they tend to cohere more strongly than larger sets. This “coherence” is hypothesized to be manifested through propensities for co-selection and phase synchrony, which have been modeled as gradiently parameterized strengths of phase coupling and activation coupling (Tilsen, 2013). The existence of these patterns of coupling makes it possible to identify co-selection sets within co-selection sets, i.e. within a maximal co-selection set there are subsets of gestures that interact more strongly with each other than with other gestures in the maximal set. Co-selection sets can therefore be conceptualized as hierarchically organized, as shown in examples of competitive-coordinative control agglomeration in Fig. 12. below. The hierarchical organization reflects the developmental precedence and stronger coherence of smaller co-selection sets. Because the component gestures of the smaller sets interact more strongly, gestures in larger, subsequently developed sets do not interact uniformly. Thus developmentally recurring feedback internalization, the mechanism through which coordinative control develops from competitive control, gives rise to hierarchical structure.

5.1 Factors in Control Regime Variation

Because hierarchical structure originates from developmental processes, and because these processes must depend to some extent on the input children receive from their language environment which in turn depends on stochastic historical factors, selection-coordination theory predicts that there should be language-, speaker-, and word-specific variation in the control regimes employed by adults. Fig. 12. illustrates several control structures arising from competitive-coordinative developmental transitions, all but one of which appear to be optional. The obligatory one is shown in Fig. 12(a), which involves CV coordination acquired in babble. In all languages children appear to learn this coordinative regime very early on.

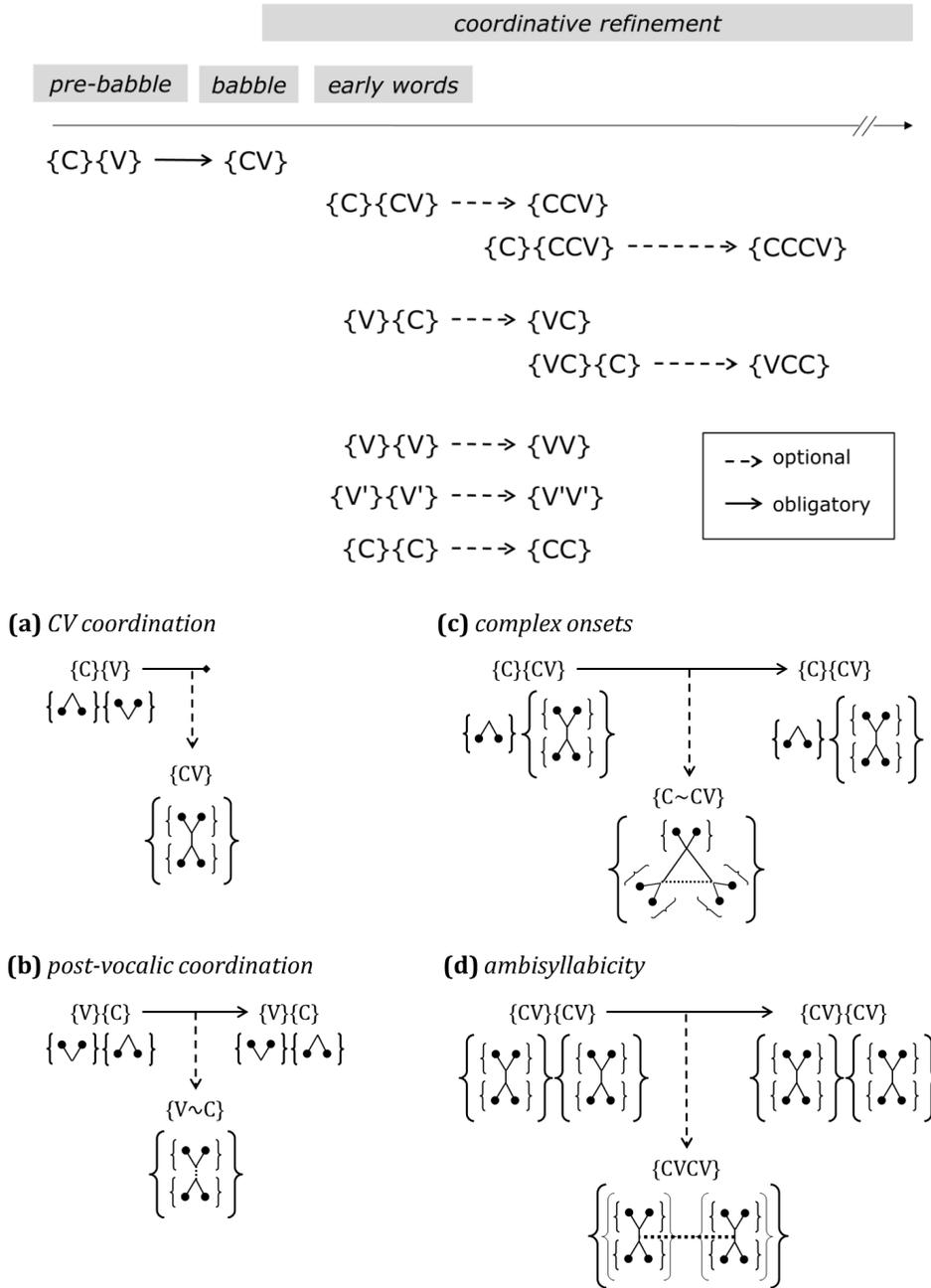


Fig. 12. (Top) Hypothesized developmental progressions in selection-coordination theory. C and V symbols represent generic consonantal and articulatory gestures, curly brackets represent co-

selection sets. All transitions except for $\{C\}\{V\} \rightarrow \{CV\}$ are hypothesized to be optional. (Bottom) Competitive-to-coordinative control transitions. Sets of relatively strongly interacting co-selected gestures are enclosed in curly brackets; in the course of development these sets are agglomerated, resulting in larger co-selection sets which nonetheless retain internal structure through gradient differences in coupling interactions.

In contrast, post-vocalic coordination (Fig. 12b) is evidently a language- and word-specific development. As observed above, there is cross-linguistic variation in moraicity of coda consonants: some languages exhibit no phonological evidence that coda consonantal gestures contribute weight, whereas in others coda consonants do; the distinction is hypothesized to correspond to whether speakers employ a coordinative or competitive regime of control. Evidence from studies of acoustic durations corroborates this interpretation (Broselow et al., 1997; Ham, 2001). Extraprosodicity phenomena, in which word-final consonantal gestures may fail to contribute weight, indicate that control regimes can be word- or gesture-specific. Hence the developmental progression to post-vocalic coordinative control in Fig. 12b appears to be optional.

Post-vocalic coordination may also be speaker-specific, particularly for more complex structures in which a coda gesture interacts with a reselected vocalic gesture or multiple vocalic gestures. Lavoie & Cohn (1999) investigated syllabicity judgments and durations of words with liquid rimes for speakers of Northeastern American English. Speakers generally agree in syllable-count judgments, but for this particular class of words, where liquids follow some tense vowels and diphthongs (e.g. *pool*, *peel*, *pail*, *pile*, *pear*, *pyre*), syllable-count intuitions vary across speakers. They observed that speakers differed in whether they judged such words to be comprised of one or two syllables. The process through which speakers reach these judgments likely involves a subvocal rehearsal of the words, so speakers are likely to base intuitions upon their own control regimes. One possible explanation for the variation is that monosyllabic judgments are associated with a coordinative control regime in which the liquid tongue tip gesture is co-selected and phased with a preceding vocalic gesture, whereas disyllabic judgments are associated with a competitive regime of control over the tongue tip gesture. This account is corroborated by the finding that the presence of an /l/ coda after a diphthong results in a substantial increase in syllable duration, whereas no such increase occurs after a lax vowel (Cohn, 2003; Lavoie & Cohn, 1999).

Onset cluster consonantal gestures provide more examples of language- and gesture-specific variation, as shown in Fig. 12(c). In section 1.2 we noted that a number of languages exhibit a C-center effect in word-initial consonants, which is indicative of co-selectional/coordinative control. Yet the C-center effect is not observed in all languages: in Tashlhiyt Berber, Moroccan Arabic, Slovak, and Hebrew only the immediately pre-vocalic gesture is coordinated with the vocalic gesture (Goldstein, Chitoran, et al., 2007; Pouplier & Beňuš, 2011; Shaw, Gafos, Hoole, & Zeroual, 2009, 2011; Tilsen et al., 2012). This suggests that the initial consonantal gesture in a cluster is competitively selected, i.e. a $\{C\}\{CV\}$ control regime. Phonological treatments of syllabification in Tashylhiyt Berber parallel this account (F. Dell & Elmedlaoui, 1985). Speakers of these languages learn to competitively select an additional consonantal gesture before $\{CV\}$ but do not appear to integrate this gesture into a $\{CCV\}$ co-selection set. Gesture-specific variation in onset cluster control has been observed in Italian, where obstruent-sonorant onsets exhibit a C-center effect but /s/-consonant onsets do not (Hermes et al., 2008): the /s/ gesture is competitively selected relative to the pre-vocalic gesture.

Intervocalic consonants provide yet another example of variation in control, shown in Fig. 12(d). Phonological patterns and syllabification judgments suggest that in some languages intervocalic consonants can be “ambisyllabic,” which means that they are associated with both the preceding and following vowels (Gussenhoven, 1986; Kahn, 1976; Rubach, 1996; van der Hulst, 1985). For example, speakers may vary in whether they feel the [p] in “apple” is associated with the first or second syllable of the word. Articulatory and acoustic studies provide further

evidence for the distinction (Barry, Klein, & Köser, 1999; Gick, 2003). In a selection-coordination framework there are two possible interpretations of ambisyllabicity: one is that the intervocalic consonantal gesture decoheres from its co-selection set and is incorporated into the preceding selection set; the other is that coordinative control develops over a foot-sized domain, in which case a word like "apple" is comprised of just one co-selection set. Although the resolution of these alternatives must be deferred to future investigation, in both cases there is substantial evidence that the pattern varies across languages, speakers, and gestures.

5.2 Task-Induced Variation and Conservation of Competitive Control

Comparisons of studies of articulatory timing in different experimental paradigms indicate that speakers can engage competitive or coordinative control in a task-dependent manner. Section 1.2 explained why CV → VC phase transitions observed in syllable repetition tasks have been taken as evidence that VC structures are coordinated and have provided the basis for the articulatory phonology hypothesis that coda consonantal gestures are anti-phase coordinated with a preceding vocalic gesture. Crucially, this evidence for coordination of coda gestures is observed in a task that imposes a periodic rhythm on selection processes, so speakers may be biased to engage a coordinative control regime. Indeed, recall that rhythmic production was hypothesized to facilitate co-selection of onset consonantal and vocalic gestures in babble. Additional evidence that rhythms facilitate coordinative timing has been reported in (Tilsen, 2009), where rhythmically simpler relations between feet and phrases resulted in diminished variability in articulatory coordination of an onset cluster. Relatedly, Tilsen (2011) observed that regularity of metrical structure resulted in diminished segmental durations. Hence repetition tasks may create a bias for coordinative control of coda gestures.

However, under different task conditions speakers may engage in competitive control over post-vocalic gestures. Tilsen (2014) developed a cued-insertion task to dissociate the selection and coordination of gestures. In this task a context syllable is repeated, and a randomly-timed insertion signal cues the speaker to insert a gesture in the upcoming syllable. The relative timing of articulatory gestures was observed to compress as a function of insertion signal timing in this experiment, and crucially, compression of articulatory timing intervals was more substantial in onset-syllabified insertions than coda-syllabified ones. This compressibility difference is predicted by a {CV}{C} mixed control model: coordinative processes are expected to exhibit variation as a function of the amount of time available for coordinative planning, whereas the time-courses of selection and deselection processes are relatively more fixed. When the time for articulatory planning is restricted, intergestural intervals associated with coordinated gestures should compress more than intervals associated with competitively selected gestures. Hence the externally imposed time pressure on execution in the cued insertion task affects articulatory intervals governed by competitive control less than intervals governed by coordinative control.

Even consonantal gestures in {CV} co-selection sets may become competitively controlled in a task-dependent fashion. Tilsen & Goldstein (2012) investigated articulatory timing in a stop-signal task, in which speakers produced CVCV disyllables in response to a go-signal but were occasionally given a stop-signal shortly thereafter, which required them to halt the response as quickly as possible. It was observed that the consonantal closure and release gestures in the first syllable (associated with the consonant and vowel, respectively) were frequently dissociated, depending on the timing of the stop-signal, and always occurred in a contingent manner. If the consonant and vowel were co-selected and precisely coordinated with a high degree of synchrony, this dissociation and contingency would be unexpected. The potential for independent production of the onset and vocalic gestures suggests that speakers may have responded to the task demands by adopting a regime of competitive control over them.

Articulatory and auditory perturbation studies provide further examples of task-driven competitive control. Munhall, Löfqvist, & Kelso (1994) found that when the lower lip was

mechanically perturbed during a bilabial closure gesture in a voiceless onset consonant, laryngeal abduction was delayed and lengthened. This effect is not expected in a coordinative model where phasing governs the relative timing of movement initiation. But if speakers adapted to the perturbation task by engaging competitive control over the labial and laryngeal gestures, the result would be delayed abduction in response to perturbation. More generally, observations of compensation and adaptation in 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 engage in a mode of production where feedback governs the deselection of movements plans, since segments produced in the contexts of auditory perturbations are typically prolonged.

Differences in articulatory timing across experimental tasks suggest that speakers adapt control regimes to facilitate task performance, but these findings do not necessarily inform our understanding of control in spontaneous conversational speech. Unfortunately direct elicitation tasks 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 the vocalic gesture is reached, or nearly at the same time. Yet because the articulatory implementation of a coda gesture is biomechanically coupled to the tongue to some degree (even if indirectly through the jaw), the apparent synchronicity of measured coda gesture initiation and vocalic target attainment can be viewed as an artifact of their biomechanical coupling. The timing pattern can be in fact interpreted in 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 before the target of the first is in actuality achieved.

Because articulatory studies do not provide conclusive evidence regarding the control regimes utilized in normal speech, we might resort to phonological patterns, such as (non)moraicity of coda consonants, to argue for competitive or coordinative control of a post-vocalic gesture. However, for reasons elaborated below, phonological patterns cannot be taken as direct evidence for adult control regimes. In addition to being somewhat circular, such arguments would ignore the possibility that phonological patterns are learned during a developmental stage in which competitive control predominates with control being subsequently internalized to the extent that it becomes coordinative.

5.3 The Relation between Control Regime Development and Phonological Patterning

The presence of a moraic phonological pattern in a language does not necessarily imply that adult speakers employ a competitive selection regime in association with words subject to that pattern. Selection-coordination theory allows for the possibility that the phonological patterns such as quantity sensitivity, contour tone restrictions, compensatory lengthening, and minimality/maximality constraints are associated with control regimes employed in some developmental stage, and that subsequently alternative regimes can be developed which are no longer consistent with the phonological interpretation the pattern. However, allowing for this possibility begs the question of how phonological patterns which are inconsistent with adult control could be productive. That is, how can adults extend a phonological pattern to novel words if they do not utilize the corresponding control regime for existing words that are subject to the pattern? Below we offer two possibilities.

One possibility is that phonological patterns are acquired during a developmental stage in which the appropriate control regime predominates and adult productivity arises from analogical mechanisms. For example, phonological patterns involving moraic codas may be acquired during the {V}{C} mixed control stage of development. Subsequently children may internalize control and develop a coordinative {VC} regime. To extend the phonological patterns in new words or

morphological contexts they have not previously encountered, adults may apply phonological patterns by analogizing between new words and familiar words on the basis of similarity and frequency/probability (Albright & Hayes, 2003; Frisch, 1997; Gentner & Markman, 1997; Pinker & Prince, 1988). However, the dynamical mechanisms through which analogical reasoning interacts with selection and coordination have not been developed.

The other possibility is that adults retain the ability to utilize competitive control after they develop coordinative control, and the retention of competitive control enables phonological patterns to be productive. The preceding section showed that control is task-specific and this indicates that adult speakers can indeed utilize competitive or coordinative control in some contexts. It could be the case that speakers simply vary from utterance to utterance in their utilization of competitive or coordinative regimes, or that speakers specially rely on competitive regimes when extending phonological patterns to unfamiliar items. Exactly what factors may condition these sorts of variation in control requires further investigation, although speech rate and word familiarity are expected to be large contributors.

A final point regarding relations between phonological patterns and control regimes is that causality is not necessarily unidirectional. For example, in relating contour tone restrictions to coordinative and competitive control regimes it was implied that biases associated with articulatory control give rise to the phonological patterns. Yet it is no less likely that children may infer a control regime from a patterns of pitch accentuation experienced in their language environment. Hence the facilitated selection of contour tones and the presence of contour tones in certain phonological contexts may concurrently reinforce one another. The same reasoning applies to all of the phonological patterns discussed above and reinforces the notion that phonetic mechanisms and phonological patterns interact in a reciprocal fashion.

6. Conclusion

In speech development children acquire procedural memories of increasing complexity. Selection-coordination theory provides a specific interpretation of this process by integrating well-motivated mechanisms of selection and coordination. The theory dissociates the “choice” of which gestures to produce (selection) from control over when, exactly, to initiate them (coordination). The developmental progression is thus interpreted to involve learning to choose more and more gestures at the same time, in conjunction with learning more complex patterns of coordinative control over when to initiate those gestures. The distinction between competitive and coordinative control is particularly useful because it provides a more specific understanding of the origins of phonological patterns as well as sources of variation in those patterns.

The outlines of selection-coordination theory presented here suggest many possible avenues of future research. One of these regards the neural implementation of selection and coordination. The concept of a co-selection set relies fundamentally on the notion of gestural primitives; these primitives can in actuality be viewed as spatiotemporal patterns of neural activity that represent speech targets and can induce movement to attain those targets. Selection can be viewed as a highly non-linear, critical process, corresponding to a phase-transition occurring when the integrated activation of a neural population reaches a threshold-like critical point; coordination is conceptualized as relative phase coupling among oscillatory systems—yet exactly how the nervous system implements these mechanisms remains to be determined.

Further, we have only scratched the surface of phonological phenomena which can 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, there is a need to develop more sophisticated kinematic and/or

acoustic assessments to distinguish between internalized feedback control and coordinative control, and a need for more frequently sampled longitudinal phonetic data.

Lastly, the selection-coordination interpretation of hierarchical structure presented here has focused on 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 all the way up the prosodic hierarchy or even to syntactic structures. A tantalizing possibility, the details of which remain to be fully worked out, is that parallel competitive and coordinative dynamics provide a fully general explanation for hierarchical structure. In the articulatory domain, phasing interactions govern relative timing of movement initiation, but for higher-level representations phasing interactions may serve relational or associative functions. If we presume that cognitive representations of higher-level 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.

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.
- Baković, E. (2007). Local assimilation and constraint interaction. Retrieved from <http://works.bepress.com/cgi/viewcontent.cgi?article=1016&context=ebakovic>
- Barry, W., Klein, C., & Köser, S. (1999). Speech production evidence for ambisyllabicity in German. *Phonus*, *4*, 87–102.
- Beckman, M. (1982). Segment duration and the “mora” in Japanese. *Phonetica*, *39*(2-3), 113–135.
- Beckman, M., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*, 7–33.
- Bernhardt, B., & Stemberger, J. (1998). *Handbook of phonological development: From the perspective of constraint-based nonlinear phonology*. San Diego: Academic Press.
- Bernstein, N. A. (1967). *The co-ordination and regulation of movements*. London: Pergamon.
- Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience*, *22*(7), 1504–1529.
- Bolinger, D. (1985). Two views of accent. *Journal of Linguistics*, *21*(01), 79–123.
- Broselow, E. (1992). Parametric variation in Arabic dialect phonology. *Perspectives on Arabic Linguistics IV, John Benjamins, Amsterdam*, 7–46.
- 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., 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. (1999). A study of Japanese speech timing from the syllable perspective. *Journal of the Phonetic Society of Japan*, 3(2), 29–39.
- Campbell, N., & Beckman, M. (1997). Stress, prominence, and spectral tilt. In In: Botinis, A., Kouroutpetroglou, G., Carayiannis, G. (Eds.), *Intonation: Theory, models and applications* (pp. 67–70). Athens.
- Campbell, N., & Sagisaka, Y. (1991). Moraic and syllable-level effects on speech timing. *Committee on Speech Research, Acoustical Society of Japan, SP90-107*, 35–40.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- 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.
- Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook*, 2, 225–252.
- Clements, G. N. (1987). Phonological feature representation and the description of intrusive stops. *CLS*, 23(2), 29–50.
- Clements, G. N., & Hume, E. V. (1995). The internal organization of speech sounds. In *The handbook of phonological theory*, ed. John Goldsmith (pp. 245–306). Oxford: Blackwell.
- Clements, G. N., & Keyser, S. J. (1983). CV phonology. a generative theory of the syllable. *Linguistic Inquiry Monographs Cambridge, Mass.*, (9), 1–191.
- Cohn, A. (2003). Phonological Structure and Phonetic Duration: The Role of the Mora. *Working Papers of the Cornell Phonetics Laboratory*, 15, 69–100.
- De Chene, B., & Anderson, S. R. (1979). Compensatory lengthening. *Language*, 505–535.
- De Jong, K. (2001). Rate-induced resyllabification revisited. *Language and Speech*, 44(2), 197–216.
- De Lacy, P. (2001). Maximal Words and the Maori Passive. In *Proceedings of AFLA VIII* (pp. 53–66). Cambridge, MA: MIT Working Papers in Linguistics.
- Dell, F., & Elmedlaoui, M. (1985). Syllabic consonants and syllabification in Imdlawn Tashlhiyt 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. In *Proceedings of NELS* (Vol. 25, pp. 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. *Taking Action: Cognitive Neuroscience Perspectives on Intentional Acts*, 289–338.
- Downing, L. J. (2000). Morphological and prosodic constraints on Kinande verbal reduplication. *Phonology*, 17(1), 1–38.
- Duanmu, S. (1994). Syllabic weight and syllabic duration: A correlation between phonology and phonetics. *Phonology*, 11(1), 1–24.
- Edwards, J., & Beckman, M. E. (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning and Development*, 4(2), 122–156.
- Eilers, R. E., Oller, D. K., Levine, S., Basinger, D., Lynch, M. P., & Urbano, R. (1993). The role of prematurity and socioeconomic status in the onset of canonical babbling in infants. *Infant Behavior and Development*, 16(3), 297–315.
- Ejiri, K. (1999). Relationship between rhythmic behavior and canonical babbling in infant vocal development. *Phonetica*, 55(4), 226–237.
- Elbert, M., & McReynolds, L. V. (1979). Aspects of phonological acquisition during articulation training. *Journal of Speech and Hearing Disorders*, 44(4), 459.
- 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*. Holland Institute of Generative Linguistics, Leiden and Amsterdam.
- Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, 4(5), 414–420.
- Forsberg, H. (1999). Neural control of human motor development. *Current Opinion in Neurobiology*, 9(6), 676–682.
- Forsberg, H., Kinoshita, H., Eliasson, A. C., Johansson, R. S., Westling, G., & Gordon, A. M. (1992). Development of human precision grip. *Experimental Brain Research*, 90(2), 393–398.
- 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.
- Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, 27, 765.
- Gafos, A. I. (1999). *The articulatory basis of locality in phonology*. Taylor & Francis.
- Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory*, 20(2), 269–337.
- Gao, M. (2008). *Tonal Alignment in Mandarin Chinese: An Articulatory Phonology Account*. Doctoral Dissertation, Yale University, New Haven, CT.
- Garrett, M. F. (1980). Levels of processing in sentence production. *Language Production*, 1, 177–220.

- 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.
- Gierut, J. A. (1993). Phonological Aspects of Emerging Grammars. *Confluence: Linguistics, L2 Acquisition, and Speech Pathology*, (4), 213.
- Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In *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. Tashlihyt Berber. In *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 241–244).
- Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, 103(3), 386–412.
- Goodell, E. W., & Studdert-Kennedy, M. (1993). Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: A longitudinal study. *Journal of Speech, Language and Hearing Research*, 36(4), 707.
- Gordon, M. (2002). A typology of contour tone restrictions. *Studies in Language*, 25(3), 423–462.
- Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. London: Erlbaum.
- Greenlee, M. (1974). Interacting Processes in the Child's Acquisition of Stop-Liquid Clusters. *Papers and Reports on Child Language Development*, 7.
- 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.
- Gussenhoven, C. (1986). English Plosive Allophones and Ambisyllabicity in Niet-lineaire fonologie, een theorie in SPE. *Gramma*, 10(2), 119–141.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, 51(5), 347–356.
- Ham, W. H. (2001). *Phonetic and phonological aspects of geminate timing*. Routledge.
- Hawkins, S. (1984). On the development of motor control in speech: Evidence from studies of temporal coordination. *Speech and Language: Advances in Basic Research and Practice*, 11. Retrieved from <http://www.haskins.yale.edu/Reprints/HL0472.pdf>
- Hayes, B. (1980). *A metrical theory of stress rules*. Massachusetts Institute of Technology. Retrieved from <http://dspace.mit.edu/handle/1721.1/16066>
- 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.
- Hertrich, 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.
- Hoequist, C. (1983). Durational correlates of linguistic rhythm categories. *Phonetica*, *40*(1), 19–31.
- Hoff, E., & Shatz, M. (2007). *Handbook of language development*. Oxford, United Kingdom: Blackwell.
- Hoole, P. (2006). Experimental studies of laryngeal articulation. In *Habilitationschrift*. University of Munich, Germany.
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor Adaptation in Speech Production. *Science*, *279*(5354), 1213–1216. doi: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.
- Hubbard, K. (1995). syllable timing: evidence from Runyambo and Luganda. *Phonology*, *12*, 235–256.
- 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.
- Ingria, R. (1980). Compensatory lengthening as a metrical phenomenon. *Linguistic Inquiry*, *465–495*.
- Iverson, J. M., Hall, A. J., Nickel, L., & Wozniak, R. H. (2007). The relationship between reduplicated babble onset and laterality biases in infant rhythmic arm movements. *Brain and Language*, *101*(3), 198–207.
- Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, *6*(11-12), 11–12.
- Jaeger, J. J. (2005). *Kids' slips: What young children's slips of the tongue reveal about language development*. Mahwah, NJ: Lawrence Erlbaum.
- Jeannerod, 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.
- Kahn, D. (1976). *Syllable-based generalizations in English phonology* (Vol. 156). Indiana University Linguistics Club Bloomington.
- Katz, W. F., & Bharadwaj, S. (2001). Coarticulation in fricative-vowel syllables produced by children and adults: a preliminary report. *Clinical Linguistics & Phonetics*, *15*(1-2), 139–143.
- Katz, W. F., Kripke, C., & Tallal, P. (1991). Anticipatory coarticulation in the speech of adults and young children: acoustic, perceptual, and video data. *Journal of Speech, Language and Hearing Research*, *34*(6), 1222.
- 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.
- 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.

- Kehoe, M., & Stoel-Gammon, C. (2001). Development of syllable structure in English-speaking children with particular reference to rhymes. *Journal of Child Language*, 28(2), 393–432.
- Kelso, J. (1982). *Human motor behavior: An introduction*. Routledge.
- Kelso, J., & Holt, K. G. (1980). Exploring a vibratory systems analysis of human movement production. *J Neurophysiol*, 43(5), 1183–1196.
- Kelso, J., & Tuller, B. (1987). Intrinsic time in speech production: theory, methodology, and preliminary observations. *Sensory and Motor Processes in Language*. Hillsdale, NJ: Erlbaum, 203, 222.
- 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 *The production of speech* (pp. 57–89). New York: Springer.
- Kent, R. (2000). Research on speech motor control and its disorders: a review and prospective. *Journal of Communication Disorders*, 33(5), 391–428.
- Kent, R., & Bauer, H. R. (1985). Vocalizations of one-year-olds. *Journal of Child Language*, 12(03), 491–526.
- Kent, R., & Forner, 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.
- Klapp, 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.
- Klatt, D. H. (1979). Synthesis by rule of segmental durations in English sentences. *Frontiers of Speech Communication Research*, 1, 287–300.
- Kozhevnikov, V., & Chistovich, L. (1965). *Speech: Articulation and perception*. (Vol. 30:543). Clearing House for Federal, Scientific, and Technical Information, JPRS.
- Krakow, R. A. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures*. Doctoral Dissertation, Yale University, New Haven, CT.
- Kuhnert, B., Hoole, P., & Mooshammer, C. (2006). Gestural overlap and C-center in selected French consonant clusters. In *Proc. 7th International Seminar on Speech Production* (pp. 327–334).
- Kunnari, S., Nakai, S., & Vihman, M. M. (2001). Cross-linguistic evidence for the acquisition of geminates. *Psychology of Language and Communication*, 5(2), 13–24.
- Ladd, D. R. (2008). *Intonational phonology*. Cambridge: Cambridge University Press.
- 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 XIVth 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 *Constraints in phonological acquisition* (pp. 204–218). Cambridge, MA: Cambridge University Press.
- Levelt, W. (1993). *Speaking: From intention to articulation*. The MIT Press.
- Levin, J. (1985). *A metrical theory of syllabicity*. Massachusetts Institute of Technology. Retrieved from <http://dspace.mit.edu/handle/1721.1/15321>
- 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.
- Locke, J. L., & Studdert-Kennedy, M. (1983). *Phonological acquisition and change*. Academic Press New York.
- Löfqvist, A., & Yoshioka, H. (1984). Intra-segmental 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.
- Lowenstamm, J., & Kaye, J. (1986). Compensatory lengthening in Tiberian Hebrew. *Studies in Compensatory Lengthening*, 97–132.
- 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), 323–350.
- MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review*, 89(5), 483.
- MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, 288(5465), 527–531.
- MacNeilage, P. F., Davis, B. L., & Matyear, C. L. (1997). Babbling and first words: Phonetic similarities and differences. *Speech Communication*, 22(2), 269–277.
- Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control*, 14(3), 380–407.
- Matthei, E. (1989). Crossing boundaries: More evidence for phonological constraints on early multi-word utterances. *Journal of Child Language*, 16(1), 41–54.
- McCarthy, J. (1979). Formal problems in Semitic morphology and phonology. *Dissertation MIT, Cambridge, MA*.
- McCarthy, J. (1988). Feature geometry and dependency: A review. *Phonetica*, 45, 84–108.
- McCarthy, J., & Prince, A. (1986). Prosodic Morphology I: constraint interaction and satisfaction. *Technical Report #32, Rutgers University for Cognitive Science, 1996*.
- McCarthy, J., & Prince, A. (1990). Prosodic morphology and templatic morphology. Retrieved from http://works.bepress.com/cgi/viewcontent.cgi?article=1067&context=john_j_mccarthy
- McLeod, S. (2007). *The international guide to speech acquisition*. Thomson Delmar Learning Clifton Park, NY.
- McLeod, S., Doorn, J. van, & Reed, V. A. (2001). Normal acquisition of consonant clusters. *American Journal of Speech-Language Pathology*, 10(2), 99.
- Meier, R. P., McGarvin, L., Zakia, R. A., & Willerman, R. (1997). Silent mandibular oscillations in vocal babbling. *Phonetica*, 54(3-4), 153–171.

- 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.
- Michelson, K. (1988). *A comparative study of Lake-Iroquoian accent*. Kluwer Academic.
- 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.
- Mohanan, K. P. (1982). *Lexical phonology*. Massachusetts Institute of Technology. Retrieved from <http://dspace.mit.edu/handle/1721.1/15652>
- Mowrey, R. A., & MacKay, I. R. (1990). Phonological primitives: Electromyographic speech error evidence. *The Journal of the Acoustical Society of America*, 88, 1299.
- Mucke, D., Nam, H., Hermes, A., & Goldstein, L. (2012). Coupling of tone and constriction gestures in pitch accents. In *Consonant Clusters and Structural Complexity* (pp. 205–230). Berlin: Mouton de Gruyter.
- 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*.
- Nagano-Madsen, Y. (1992). *Mora and prosodic coordination: A phonetic study of Japanese, Eskimo and Yoruba*. Lund University Press.
- Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. *Approaches to Phonological Complexity*, 299–328.
- Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th International Conference on Phonetic Sciences* (pp. 2253–2256). Barcelona, Spain.
- 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.
- Nooteboom, S. G. (1973). The tongue slips into patterns. *Speech Errors as Linguistic Evidence*, 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.
- Ohala, J. (1992). The segment: primitive or derived. *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, 166–183.
- Öhman, S. E. G. (1967). Numerical model of coarticulation. *The Journal of the Acoustical Society of America*, 41(2), 310–320.
- Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum.
- Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy: the prediction of speech and language disorders. *Journal of Communication Disorders*, 32(4), 223–245.
- Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language*, 3(1), 1–11.
- 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.

- Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Massachusetts Institute of Technology. Retrieved from <http://dspace.mit.edu/handle/1721.1/16065>
- Pierrehumbert, J., & Nair, R. (1995). Word games and syllable structure. *Language and Speech*, 38(1), 77–114.
- Piggott, G. L. (1988). Prenasalization and feature geometry. In *Proceedings of NELS* (Vol. 19, pp. 345–352).
- 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., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *The Journal of the Acoustical Society of America*, 81, 1574–1585.
- Port, R. F., & Leary, A. P. (2005). Against formal phonology. *Language*, 927–964.
- Posner, M. L., & Snyder, C. (1975). Attention and cognitive control. In *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.
- Pouplier, M., Chen, L., Goldstein, L., & Byrd, D. (1999). Kinematic evidence for the existence of gradient speech errors. *The Journal of the Acoustical Society of America*, 106, 2242.
- 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. Retrieved from <http://www.mproctor.net/docs/diss/front.pdf>
- 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. doi:10.1016/j.humov.2007.04.001
- Rosenthal, S. (1992). Prenasalized stops and feature geometry. In *Phonologica 1988: Proceedings of the 6th International Phonology Meeting, [July 1-4, 1988, in Krems, Austria]* (p. 249).
- 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.
- Rubach, J. (1996). Shortening and ambisyllabicity in English. *Phonology*, 13(02), 197–237.
- 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. Retrieved from <http://dspace.mit.edu/handle/1721.1/15106>
- Sagisaka, Y., & Tohkura, Y. (1984). Phoneme duration control for speech synthesis by rule. *IEICE Trans*, 629–636.
- 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.
- Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th international conference on speech prosody* (pp. 175–184). Brazil: Campinas.
- 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.
- Sato, Y. (1993). The durations of syllable-final nasals and the mora hypothesis in Japanese. *Phonetica, 50*(1), 44–67.
- Schmidt, R. C., Bienvenu, 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 synergetic 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. In *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.
- Sezer, E. (1986). An autosegmental analysis of compensatory lengthening in Turkish. *Studies in Compensatory Lengthening, 23*, 227.
- Shaffer, L. H. (1976). Intention and performance. *Psychological Review, 83*(5), 375–393.
- Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342).
- Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In *The production of speech* (pp. 109–136). New York: Springer.
- Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior, 18*(1), 41–55.
- 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.
- Smit, A. B. (1993). Phonologic error distributions in the Iowa-Nebraska articulation norms project: Word-initial consonant clusters. *Journal of Speech, Language and Hearing Research, 36*(5), 931–947.
- Smith, 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.
- Stemberger, J. P. (1983). *Speech errors and theoretical phonology: A review*. Indiana University Linguistics Club.
- Steriade, D. (1982). *Greek prosodies and the nature of syllabification*. Massachusetts Institute of Technology. Retrieved from <http://dspace.mit.edu/handle/1721.1/15653>
- Steriade, D. (2000). Directional asymmetries in place assimilation: a perceptual account. *Perception in Phonology*.
- Steriade, D. (2001). The phonology of perceptibility effects: The P-map and its consequences for constraint organization. *Ms., UCLA*.
- Sternberg, S., Knoll, R., Monsell, S., & Wright, C. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica, 45*(2-4), 175–197.
- Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. *Information Processing in Motor Control and Learning, 117–152*.
- Stetson, R. H. (1951). *Motor phonetics; a study of speech movements in action*. Amsterdam: North-Holland Publishing Company.
- Stevens, K. N. (1981). Evidence for the role of acoustic boundaries in the perception of speech sounds. *The Journal of the Acoustical Society of America, 69*(S1), S116–S116.
- 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.
- Thelen, E. (1979). Rhythmical stereotypies in normal human infants. *Animal Behaviour, 27*, 699–715.
- Thompson, A. E., & Hixon, T. J. (1978). *Nasal air flow during normal speech production*. University of Arizona.
- Tilsen, S. (2009). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science, 33*(5), 839–879.
- 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. (2014). Selection and coordination of articulatory gestures in temporally constrained production. *Journal of Phonetics, 44*, 26–46.
- Tilsen, S., & Goldstein, L. (2012). Articulatory gestures are individually selected in production. *Journal of Phonetics, 40*(6), 764–779.
- 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.
- Van der Hulst, H. (1985). Ambisyllabicity in dutch. *Linguistics in the Netherlands, 1985*, 195–210.
- Vihman, M. (1996). *Phonological development: The origins of language in the child*. Blackwell Publishing.
- 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.
- Warner, N., & Arai, T. (2001). The role of the mora in the timing of spontaneous Japanese speech. *The Journal of the Acoustical Society of America*, 109, 1144–1156.
- 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.
- Zec, D. (2003). Prosodic weight. *The Syllable in Optimality Theory*, 123–143.
- Zec, D. (2007). The syllable. *The Cambridge Handbook of Phonology*, 161–194.

Cornell University
203 Morrill Hall
Ithaca, NY 14850
USA
tilsen@cornell.edu