Accepted for publication in *JSLHR* February 15, 2019

Speech production from a developmental perspective

Melissa A. Redford

University of Oregon

Running title: Speech production from a developmental perspective

Keywords: speech motor learning; language acquisition; whole-word approach; child phonology.

Please address correspondence to:

Melissa A. Redford
Linguistics Department
1290 University of Oregon
Eugene, OR 97403

email: redford@uoregon.edu

Abstract

Purpose: Current approaches to speech production aim to explain adult behavior and so make assumptions that, when taken to their logical conclusion, fail to adequately account for development. This failure is problematic if adult behavior can be understood to emerge from the developmental process. This problem motivates the proposal of a developmentally sensitive theory of speech production. The working hypothesis, which structures the theory, is that feedforward representations and processes mature earlier than central feedback control processes in speech production.

Method: Theoretical assumptions that underpin the two major approaches to adult speech production are reviewed. Strengths and weaknesses are evaluated with respect to developmental patterns. A developmental approach is then pursued. The strengths of existing theories are borrowed and the ideas resynthesized under the working hypothesis. The speech production process is then reimagined in developmental stages, with each stage building on the previous one.

Conclusion: The resulting theory proposes that speech production relies on conceptually-linked representations that are information-reduced holistic perceptual and motoric forms, constituting the phonological aspect of a system that is acquired with the lexicon. These forms are referred to as exemplars and schemas, respectively. When a particular exemplar and schema are activated with the selection of a particular lexical concept, their forms are used to define unique trajectories through an endogenous perceptual-motor space that guides implementation. This space is not linguistic, reflecting its origin in the pre-speech period. Central feedback control over production emerges with failures in communication and the development of a self-concept.

Introduction

Speech motor control allows for flexible, fast, and precise coordination of speech articulators to achieve a motor goal. Adult performance in auditory feedback perturbation experiments suggests not only sensitivity to deviations between, say, an intended vowel and the acoustics of the vowel produced, but also an ability to compensate for these deviations with fine motor adjustments that can raise or lower a particular formant frequency by as little as 50 hertz (see, e.g., MacDonald, Goldberg, & Munhall, 2010; Katseff, Houde, & Johnson, 2012). It is perhaps not surprising that this kind of fine-grained spatiotemporal control over articulation develops slowly. Large gains in speech motor skill are made during the first few years of life, but adult-like control is not achieved until mid-adolescence. Evidence for this claim dates back to Kent and Forner (1980), who pointed out that temporal variability in young school-aged children's segmental durations is higher than in adults' speech, and that this remains true until 12 years of age (see, also, B. Smith, 1992; Lee, Potamianos, & Narayanan, 1999). These acoustic findings were later supplemented with kinematic ones, which validated the interpretation of greater temporal variability in children's speech as the result of immature articulatory timing control (Sharkey & Folkin, 1985; Smith & Goffman, 1998; Green, Moore, Higashika, & Steeve, 2000). A. Smith and Zelaznik (2004) followed up on this work with older children and showed that articulatory timing control is not fully mature until mid-adolescence. So, given the protracted development of speech motor control, why can we more or less understand what children are saying when they first begin to use words at about 12 months of age? And, even more strikingly, how is it possible that 3-year-old children seem to never stop talking when their speech motor skills are still so immature? The answer put forward in this paper is that feedforward processes mature earlier than central feedback control processes.

More specifically, the argument developed herein is that speech production relies on conceptually-linked representations that are abstract (i.e., information-reduced) holistic perceptual and motoric forms. These forms constitute the phonological aspect of the lexicon. The perceptual

phonological forms are exogenous representations. They are exemplars that are acquired with lexical concepts beginning around 9 months of age. The motoric phonological forms are endogenous representations. They are schemas that begin to be abstracted around 12 months of age with first word productions. When a particular exemplar and schema are activated with the selection of a particular concept, their forms are used to define unique trajectories through an endogenous perceptual-motor space that guides implementation. This space is not linguistic: it processes are entirely free from conceptual information. The absence of conceptual information reflects the origin of this space in the pre-speech period when infants' vocal explorations create the first linkages between perceptual and motoric trajectories.

By hypothesis, schemas are modified through developmental time as central feedback control is incorporated into the production process. This is because the act of speaking indirectly modifies schemas via the same process used to first abstract them. The onset of high-level predictive feedback control emerges with communication failures. These failures are assumed to significantly increase with vocabulary size due to homophony, motivating a shift in the production system towards exemplar representations around 18 months of age. The shift drives the emergence of an internal loop that matches the (projected) perceptual consequences of self-productions against targeted exemplar representations. Selective attention to auditory feedback develops later during the preschool years with the emergence of self-concept. At this point, the child begins to focus on sound production per se in addition to communication. The latter hypothesis could explain why literacy acquisition becomes possible around age 5 years and why direct intervention for speech sound disorders also becomes effective at this age.

The argument outlined above is in fact a general theory of speech production that is developmentally sensitive. The theory combines those aspects of existing adult-focused theories that best accommodate acquisition to define whole-word production at different stages of development from infancy to childhood on into adulthood. This developmentally-sensitive theory of

speech production is further motivated below. This motivation begins with a review of adult-focused theories. A major point of the review will be that the two major approaches to speech, the ecological dynamics and information processing approach, lead to different emphases regarding the type of feedforward information used in production (motoric versus perceptual) and to different views on the type of feedback control processes engaged during execution (peripheral versus central). I will argue that the holistic motoric representations that drive production in the ecological dynamics approach are consistent with functional approaches to child phonology and better account for young children's speech patterns than the discrete perceptual representations that drive production in the information processing approach. Nonetheless, the information processing assumption of distinct production and perception systems is embraced in the developmentally-sensitive theory of speech production that I put forward because central feedback control is deemed necessary to account for the evolution of children's speech patterns from first words to adult-like forms.

Adult-focused Theories of Speech Production

Adult-focused theories of speech production assume the activation of an abstract phonological plan that is then rendered in sufficient phonetic detail for the sensorimotor system to activate speech movements (e.g., Dell, 1986; Garret, 1988; Browman & Goldstein, 1992; Guenther, 1995; Roelofs, 1999; Keating & Shattuck-Hufnagel, 2002; Goldrick, 2006; Goldstein, Byrd, & Saltzman, 2006; Turk & Shattuck-Hufnagel, 2014; *inter alia*). The detailed phonetic plan is known as a speech plan. It contains or directly activates linguistic representations that provide relevant feedforward information for implementation. The representations and type of feedback control processes used in production differ according to the theoretical approach taken. Here, the two main approaches to speech production are reviewed: the ecological dynamics approach and the information processing approach (see Figure 1). These approaches represent an amalgam of different theories, hence the generic labels. The different sets of theories emerge from two fundamentally different approaches

to human cognition—an ecological-embodied approach versus a representation-based information processing approach, which are briefly described next.

Richardson and colleagues (Richardson, Shockley, Fajen, Riley, & Turvey, 2009) outline the tenets of an ecological-embodied approach in contrast to the assumptions of an information processing approach as follows. In an ecological-embodied approach, behavior is emergent and self-organized, which is to say behavior is not planned or controlled (pp. 170-173). Perception and action are viewed as continuous and cyclic and thus functionally united (pp. 173-175). In particular, the concept of affordances assumes that the objects of perception provide information about action possibilities (pp. 178-182). The theory of direct perception assumes that these useful objects are wholly conveyed by sensory input (pp. 176-178). This means that knowledge is simply extracted from the environment within which the individual lives and moves (pp. 167-170).

The ecological-embodied view of knowledge contrasts with the information processing view where knowledge emerges from learned associations, which give rise to mediating representations. These representations *are* knowledge in the information process approach This view of knowledge follows from other assumptions: individuals are separate from their environment; the mind is separate from the body; and, action is separate from perception. Overall, representational and computational processes are "lifted away from the organism-environment system and … studied on their own, permitting cognitive scientists to proceed whereas other specialists work to understand the body and environment of the knower (Richardson et al. 2009: 161-162)." This approach to human cognition is likely more familiar to readers than the ecological-embodied approach because it has provided the philosophical foundation for much of mainstream cognitive sciences in North America, including linguistics and psychology, since the "cognitive revolution" in the 1950s (see Mandler, 2007:Ch.10). The assumptions of this approach are detailed in Newell and Herbert's (1972) classic book, *Human Problem Solving*.

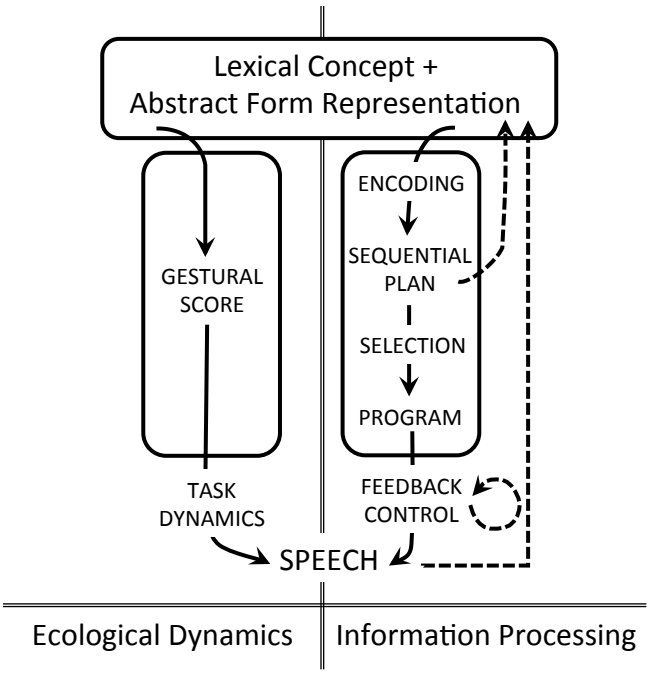

Figure 1. The ecological dynamics and information processing approaches to speech production both assume 3 major levels of analysis: a phonological level where abstract form representations are associated with conceptual meaning; a speech plan level where abstract forms are elaborated for implementation; and an implementation level where articulatory action is formulated and adjusted in real-time to achieve the plan. The two approaches otherwise adopt very different fundamental assumptions, resulting in different theories of representation, sequencing, and control. In particular, the ecological dynamics approach emphasizes speech as action, assumes gestalt articulatory representations, emergent sequential structure, and self-organized articulation. In contrast, the information processing approach emphasizes the importance of discrete elements, assumes executive control over sequencing and implementation, thus promoting a strong role for perception in production while assuming that the two processes are distinct. Solid lines with arrows represent feedforward processes; dotted lines with arrows represent feedback processes.

The information processing approach has resulted in the modular study of language (e.g., syntax versus phonology) and in a sharp division of expertise between those who study language and those who are interested in speech production (e.g., phonology versus phonetics). Among the latter, those who adhere closely to the approach often focus on the translation problem that follows from their computational view; for example, the problem of how discrete phonological elements are transformed into continuous speech action (Wickelgran, 1969; MacNeilage, 1970; Bladon & Al-Bamerni, 1976; Blumstein & Stevens, 1981; Recasens, 1989; Keating, 1990; *inter alia*). This focus also structures psycholinguistic models of production that posit multiple processing stages to generate production units (e.g., Dell, 1986; Garret, 1988; Levelt, 1989; Roelofs, 1999; Goldrick,

2006), a generic version of which is presented in the right-hand panel of Figure 1. Models of speech motor control that have discrete elements as goals emphasize feedback control to ensure accurate implementation of these elements in speech movement (e.g., Lindblom, Lubker, & Gay, 1972; Abbs, Gracco, & Cole, 1984; Perkell et al., 1997; Houde & Nagarajan, 2011; Tourville & Guenther, 2011; Hickok, 2012; Niziolek, Nagarajan, & Houde, 2013).

In contrast to the information processing approach, the ecological-embodied approach has been mainly applied to the study of speech (Fowler, 1986; Kelso, Saltzman, & Tuller, 1986; Saltzman & Kelso, 1987; Saltzman & Munhall, 1989; Browman & Goldstein, 1992; Best, 1995; Goldstein & Fowler, 2003; Galantucci, Fowler, & Turvey, 2006). The assumption of separate language and speech systems is thus preserved by default, but all aspects of the speech process are fully consistent with the tenets of an ecological-embodied approach, which entails no translation between higher level speech sound representations and lower level speech movement. According to the ecological-embodied approach, phonological forms are objects of both action and perception. These forms become increasingly elaborated when activated through self-organization rather than through planning. Thus, the flow from high to low is better conceived of as the emergence of speech form, which is mediated only by a linearized version of a nonlinear representation (i.e., a gestural score; see Figure 1, left). The specific assumptions of each approach to speech production are elaborated further below, beginning with the action-focused ecological dynamics approach.

The Ecological Dynamics Approach

The ecological dynamics approach to speech production is best represented by Articulatory Phonology (Browman & Goldstein, 1992, and subsequent), a task-dynamic approach to articulation (Kelso, Saltzman, & Tuller, 1986; Saltzman & Kelso, 1987; Saltzman & Munhall, 1989), as well as ecological theories of speech perception (Fowler, 1986; Best, 1995; Goldstein & Fowler, 2003; Galantucci, Fowler, & Turvey, 2006) and speech sound acquisition (Best, 1995; Best, Goldstein, Pam, & Tyler, 2016). The fundamental unit of analysis is a vocal tract constriction that serves as an

articulatory attractor. This unit is known as a gesture. Gestures are linguistic primitives, similar to distinctive features in generative theory, that emerge during development under the assumption that infants acquire "a relation between actions of distinct (articulatory) organs and lexical units very early in the process of developing language." (Goldstein & Fowler, 2003:35; see also Best et al., 2016). Gestures are defined as "events that unfold during speech production and whose consequences can be observed in the movements of the speech articulators (Browman & Goldstein, 1992:156)." More specifically, they are abstract representations of "the formation and release of constrictions in the vocal tract (*ibid*)," which are realized dynamically thus giving them an event-like status. This status in turn confers intrinsic timing; that is, once activated, gestures take time to achieve a target vocal tract constriction and then time to move away from the constriction.

The assumption of intrinsic timing has a number of interesting theoretical consequences, several of which are compatible with a developmental perspective on speech production. Perhaps the most important of these consequences is in the representation of sequential articulation (see, e.g., Fowler, 1980; Kelso et al., 1986; Saltzman & Munhall, 1989; Browman & Goldstein, 1992; Fowler & Saltzman, 1993). Gestures, like their distinctive feature counterparts in generative phonology, are always realized as part of a larger whole (i.e., a "molecule"). But, unlike distinctive features, the wholes are not bundled up into individual phonemes that must be sequenced during the production process. Instead, gestures participate in an articulatory gestalt that is, minimally, syllable-sized. Moreover, all relevant gestures associated with a lexical entry are co-activated when that entry is selected for production (Browman & Goldstein, 1989; 1992; Goldstein, Byrd, & Saltzman, 2006). Put another way, the Articulatory Phonology view of lexical form representations is that these are holistic and motorically-based. The developmentally-sensitive theory I propose shares this view of lexical representation; I also argue for holistic, perceptually-based form representations.

Under the ecological-embodied assumption of cyclic action, appropriate sequencing within a word is emergent. To understand emergent sequencing, consider, for example, the coordination of a single consonantal and vocalic gesture. Consonantal gestures are intrinsically shorter than vocalic gestures. They are also phased relative to one another: if the cyclic gestures are coordinated without a phase difference, a consonant-vowel (CV) syllable emerges; if they are 180 degrees out of phase, a VC syllable emerges (Browman & Goldstein, 1988; Goldstein, Byrd, & Saltzman, 2006; Nam, Goldstein, & Saltzman, 2009). These in-phase and anti-phase relations are stable coordination patterns in motor systems (Haken, Kelso, & Bunz, 1985; Turvey, 1990). Of course, languages allow for consonant or vowel sequences that complicate stable coordination dynamics (e.g., consider the English word "sixths" among many, many others). Thus, gestural timing associated with individual words may be learned during speech acquisition and incorporated into a coupling graph, which is the lexical form representation in Articulatory Phonology (Fowler & Goldstein, 2003; Goldstein et al., 2006; Nam et al., 2009).

Note that the ecological dynamics conception of coordination also has implications for a theory of coarticulation, which is understood within this approach to speech production as coproduction (see Fowler, 1980). In contrast to information processing approaches to coarticulation, dynamic formant trajectories and distributed spectral effects of rounding and nasalization and so on emerge directly from the representation; they are never due to a central executive that "looks ahead" to the next sound(s) while preparing the current one. This view of coarticulation appears to be more compatible with developmental findings on coarticulation than the information processing view, a point to which I return later.

When words are selected for production, their coupling graphs give rise to linearized gestural scores (Goldstein et al., 2006; inter alia). These scores meet the generic definitions of both a speech plan and a motor program. They are plans in that they specify, abstractly, the relative timing and duration of specific speech actions. They are programs in that they drive these actions directly via

task dynamics (Saltzman & Munhall, 1989). The dynamic transformation from coupling graph to gestural score means that there is no speech *planning* in the ecological dynamic approach to speech production, only speech plans that serve also as phonological representations. I make a similar assumption in the developmentally-sensitive theory proposed herein.

During the implementation stage of the production process, gestures represent motor goals (Löfqvist, 1990; Fowler & Saltzman, 1993). Articulators self-organize to effect these goals. Self-organization is based in large part on functional synergies that stabilize over developmental time to become part of the motor control system (see, e.g., A. Smith & Zelaznik, 2004). In other words, gestures give rise to a type of functional motor unit of coordination (i.e., a "coordinative structure"). Peripheral perceptual feedback provides relevant context information to subcortical structures and the peripheral nervous system for goal achievement (see, e.g., Saltzman & Munhall, 1989:48) and to automatically compensate for perturbations (see, e.g., Abbs & Gracco, 1984). In this way, there is no real control over production in the sense of cortically-mediated adjustments to movement direction and velocity. Whereas this view of implementation and its development can account for infant vocalizations and early speech attempts as well as for the overall slow development of speech motor skills, I argue below that the strong evidence from adult speech for cortically-mediated control over production must be incorporated into a developmentally sensitive theory of speech production to account for phonological change through developmental time.

In sum, an ecological dynamics approach to speech production assumes an entirely feedforward process. Motor goals are articulatory and event-like and phased relative to one another in articulatory gestalt representations that are linked to conceptual information in the lexicon. Sequential structure and coarticulatory overlap emerge from gestural dynamics. Production itself is a self-organized process. Thus, the approach eschews the concept of central control over speech production based on first principles.

The Information Processing Approach

The information processing approach to speech production is best represented by mainstream psycholinguistic theories of language production (e.g., Dell, 1986; Garret, 1988; Roelofs, 1999; Goldrick, 2006), phonetically-informed theories of implementation (e.g., Guenther, 1995; Keating & Shattuck-Hufnagel, 2002; Guenther & Perkell, 2006; Turk & Shattuck-Hufnagel, 2014) and prediction-based models of speech motor control (e.g., Houde & Nagarajan, 2011; Tourville & Guenther, 2011; Hickok, 2012; Niziolek, Nagarajan, & Houde, 2013). In this approach, phonological representations mediate between perception and production. They are abstract *and* symbolic.

The phoneme—a categorical and discrete element—is often the fundamental unit of analysis in this approach. The emphasis on phonemes is due to a modeling focus on speech errors (e.g., Dell, 1986; Garret, 1988; Levelt, 1989; Roelofs, 1999; Bock & Levelt, 2002), which are best described with reference to segmental structure (see also MacKay, 1970; Shattuck-Hufnagel & Klatt, 1979). These modeling efforts have led to the psycholinguistic assumption that segment sequencing is an active process during production (Dell, 1986; Garret, 1988; Levelt, 1989; Roelofs, 1999; Bock & Levelt, 2002, *inter alia*). This process has come to be known as phonological encoding (see Figure 1, right). Theories diverge on how encoding happens, but once encoded, all theories recognize that the phonemic string must be further specified before it can be used as a plan for output. In Levelt's (1989) highly influential model, the string is metrically chunked for output, allowing for specification of positional information via allophone selection; for example, the aspirated variant of the voiceless alveolar stop is chosen for *tab* (i.e., [t^hæb]), the unreleased variant is selected for *bat* (i.e., [bæt̚]), and the stop is replaced by a flap in *batter* (i.e., [bærɚ]). From a developmental perspective, the mainstream assumption of phonological and phonetic encoding complexifies speech acquisition since it predicts that infants must learn a symbolic system as well as the computational steps necessary to translate symbolic representations into action plans.

Once a phonological string has been phonetically encoded, it can be implemented. Implementation can mean the appropriate selection of a syllable-sized motor programs from a mental syllabary (e.g., Levelt, 1989; Guenther, Ghosh, & Tourville, 2006; Bohland, Bullock, & Guenther, 2010) or careful specification of articulatory timing information (e.g., Keating, 1990; Turk & Shattuck-Hufnagel, 2014). Either way, discrete phones remain high-level motor goals during execution. These goals are conceived of specifically as speech sound categories (e.g., Lindblom, Lubker, & Gay, 1972; Lindblom, 1990; Johnson, Flemming, & Wright, 1993; Guenther, 1995; Hickok & Poeppel, 2000; *inter alia*) or more generally as perceptual categories (e.g., Perkell, Matthies, & Svirsky, 1994; Savariaux, Perrier, & Orliaguet, 1995; Schwartz, Boë, Vallée, & Abry, 1997; *inter alia*). Importantly, the goals remain non-overlapping even in high-frequency combinations when, through repeated practice, they may be stored together as part of a larger chunk (see, e.g., Bohland et al., 2010:1505). This view that stands very much in contrast to the ecological-dynamics view where chunks are articulatory gestalts comprised of overlapping gestures/articulatory events. The assumption of discrete goals also requires computationally intensive accounts of coarticulation, especially long-distance coarticulation, which is explained in the information processing approach to result either from feature spreading at an early stage of encoding (e.g., Daniloff & Harmmarberg, 1973; Bladon & Al-Bamerni, 1976; Recasens, 1989) or from planning for the articulation of individual phones within a well-defined window during a later stage of encoding (e.g., Keating, 1990; Guenther, 1995). These accounts wrongly predict the slow development of coarticulation (see below).

Although discrete perceptual speech motor goals are problematic from a development perspective, they are posited in the information processing approach to explain "the exquisite control of vocal performance that speakers/singers retain for even the highest frequency syllables (Bohland et al., 2010:1509)". Exquisite control of vocal performance requires the coordination of multiple independent speech articulators through time, each of which also has many degrees of

movement freedom—another developmentally-unfriendly computational problem. The coordination problem is solved in the information processing approach by assuming central perceptual feedback control over articulatory movements. An assumption for which there is now abundant evidence.

Central feedback control means cortically-mediated adjustments to articulation made with reference to perceptual goals in order to achieve on-target sound production. Of course, slow central processing of perceptual feedback presents a problem for perceptual feedback during real-time speech production (see e.g., MacNeilage, 1970; Lindblom, Lubker, & Gay, 1979). Lindblom and colleagues (1979:160) were the first to propose a viable solution to this problem. Specifically, they proposed that motor control does not rely on processing perceptual feedback per se, but instead references the simulated perceptual results of planned action while execution unfolds. Lindblom and colleagues called this proposal *predictive encoding* and with it they foreshadowed the emphasis in current models of speech motor control where a copy of the output signal (= efference copy) is used to predict sensory outcomes (e.g., Houde & Nagarajan, 2011; Tourville & Guenther, 2011; Hickok, 2012; Niziolek, Nagarajan, & Houde, 2013) for error correction purposes (e.g., Tourville & Guenther, 2011) or real-time speech motor control (see, e.g., Niziolek et al., 2013). The proposal is supported by speakers' remarkable ability to correctly produce target sounds when normal articulation is disrupted.

Lindblom and colleagues (1979) proposed predictive encoding to account for their speakers' near instantaneous adaption to different bite-block manipulations during vowel production. Since then, many sophisticated perturbation experiments have been conducted (e.g., Savariaux, Perrier, & Orliaguet, 1995; MacDonald et al., 2010; Katseff et al., 2012; Lametti, Nasir, & Ostry, 2012; *inter alia*). These experiments provide strong evidence in favor of perceptual goals and for the role of central feedback control in speech production. Consider, for example, a study by Lametti and colleagues (2012), which investigated the effects of different types of perceptual feedback

perturbations on the repetition of a target word, *head*. Somatosensory feedback was disrupted by a robot arm, which tugged randomly at the speakers' lower jaw, thereby disrupting the normal articulatory path for the target /ɛ/ vowel. Auditory feedback was perturbed by altering the speaker's own F1 upward in the direction of an /æ/ vowel. This real time alteration was sent to the speaker via headphones. The results indicated that speakers counteracted the effects of perturbation through compensation to maintain the target, *head*, production. While the majority of speakers compensated more for auditory perturbations than somatosensory perturbations, some speakers showed the opposite effect and many adapted to both types of perturbations.

It has been argued that whereas perturbation experiments provide evidence for error correction based on perceptual feedback, conclusions about real-time speech motor control are more dubious since the experimental findings require manipulations that create very unnatural speaking conditions (see, e.g., Guenther et al., 2006:288). Yet, the basic behavior observed in perturbation experiments—speaker adjustments based on incoming perceptual information—is also observed in phonetic imitation experiments, which are significantly more natural. Instead of participants hearing their own perturbed speech, they simply repeat words that others have produced (e.g., Goldinger, 1998; Shockley, Sabadini, & Fowler, 2004; Nielsen, 2011; Babel, 2012). Just as in the perturbation paradigm, participants are found to make fine-tuned adjustments to their own speech in the direction of the input; for example, participants' production of voice onset time in stop production is measurably changed when shadowing exposure to stop-initial words with substantially different voice onset time (VOT) values than their own (Shockley et al., 2004). Moreover, behavior in these laboratory experiments also corresponds to the real-world language phenomenon of convergence (Giles & Powesland, 1997), where interlocutors begin to sound like one another over the course of an exchange. When speakers subconsciously "converge" on a set of phonetic features during an interaction they are demonstrating that perceptual input informs on-line spoken language production (see, e.g., Babel, 2012). Thus, speakers behavior in contrived and

natural speaking conditions provides strong evidence for the importance of perceptual feedback during speech production. The developmentally sensitive theory proposed herein is meant to accommodate this evidence.

In sum, the information processing approach emphasizes the importance of discrete elements, and so assumes executive control over sequencing and implementation. This assumption entails a role for perception in production. The evidence for on-line vocal-motor adjustments based on self and other generated auditory information is especially strong, and consistent with the hypothesis of central perceptual feedback control over speech production.

Implications of Adult-focused Theories for the Development of Speech Production

From a developmental perspective, the different approaches to speech production each have strengths and important limitations that were alluded to above. The main strength of the ecological dynamics approach is the central hypothesis that temporal relations between articulators are preserved as part of an articulatory gestalt lexical representations. This hypothesis, consistent with whole-word approaches to child phonology, provides a framework for understanding children's speech patterns. The strength of the information processing approach is in recognizing the importance of perceptual feedback for tuning speech production. This emphasis is not only consistent with adult behavior, it also provides a powerful mechanism for learning, and thus the ability to explain change over developmental time. These points are elaborated below with a focus on explaining children's speech patterns and developmental change.

Children's Speech Patterns

Child phonology is often viewed from the adult perspective, hence the description of children's speech as fronted, harmonized, simplified, and so on. Implicit is the idea of transformed adult-like representations. As long as the transformation results in a string of phonemes readied for output, speech acquisition can be handled by an information processing approach and construed as phonemic acquisition (see Vihman, 2017, for a review and critique of this view). When construed in

this way, the learning problem is restricted to the mapping of phoneme-related speech sounds to articulatory movement. The DIVA model (Guenther, 1995; Guenther et al., 2006) instantiates this view of speech acquisition and production. The following discussion focuses on the shortcomings of this model to convey a general, developmental critique of the information processing approach. This focus is a testament to DIVA's influence on the field and to its status as the most complete and explicit statement of an information processing theory of speech production. Also, the original DIVA model (Guenther, 1995), though ultimately adult-focused, was at least constructed to reflect the knowledge that adult behavior emerges over developmental time. This further increases the relevance of DIVA to the present discussion.

In DIVA, speech motor targets are specified as coordinates in an orosensory space. The coordinates correspond to vocal tract shapes. Speech motor goals are acoustically defined and reside in the speech sound map of the model. Linkages between the speech sound map and orosensory space are acquired during babbling. An orosensory to articulation map is established during the first phase of babbling via random articulatory movements. The speech sound map is then acquired during a second phase that relies on overt perceptual feedback to register regions in the orosensory space associated with known (i.e., perceptually acquired) language-specific sounds. Once linkages between discrete sounds and articulation have been established via orosensory space, speech production can be driven by phoneme strings that sequentially activate cells within the speech sound map.

The ease with which the DIVA model can learn to produce language-specific sequences highlights a limitation of the information processing approach to the development of speech production: it does not take seriously the slow development of speech motor skills. Production proceeds just as in the adult once the phoneme-to-sound and sound-to-articulation mappings have been established. For example, "after babbling, the (DIVA) model can produce arbitrary phoneme strings using a set of 29 English phonemes in any combination" (Guenther, 1995:598). In this way,

DIVA's behavior is obviously at odds with real development. Child phonological patterns such as gliding (*leg* → *weg*, *bread* → *bwead*), stopping (*feet* → *peet*, *house* → *hout*) epenthesis (*sleep* → *se-leep*, *green* → *ge-reen*), cluster simplification (*clean* → *keen*, *stop* → *top*) often persist until the school-age years (Stoel-Gammon & Dunn, 1985:43-46).

Although child phonological patterns can be explained within the information processing approach by positing grammatical rules that constrain sequencing (see, e.g., Kager, Pater, & Zonneveld, 2004, and the contributions therein), the assumption that children learn via perceptual feedback to produce discrete perceptual goals in sequence incorrectly predicts that young children produce speech that is *less* coarticulated than adult speech (see, e.g., Kent, 1983; Guenther, 1995; Tilsen, 2014). Guenther (1995:617) cites Thompson and Hixon's (1979) study on anticipatory nasal coarticulation in support of this prediction. However, the vowel midpoint measure used in that study assumes static phonemic targets that are achieved at the middle of an acoustic interval rather than the dynamic specification of movement. Flege (1988) took a different approach and measured the duration of nasalization across the entire vowel in child and adult speech. His results showed that children and adults both open "the (velar-pharyngeal port) long before the lingual constriction for word-final /n/ (p. 533)." Moreover, when vowel duration was controlled, Flege found no significant differences in the degree to which children and adults engaged in anticipatory behavior.

Guenther (1995) also cites Kent's (1983) chapter to argue that children's speech is more segmental than adults. This was Kent's contention, but it was not rigorously demonstrated. Instead, Kent made a qualitative comparison of F2 trajectories in 4-year-old children's and adults' production of spoken phrases. He discussed the F2 patterns in the spectrograms provided and noted that children's vowel productions appeared to be less influenced by adjacent consonantal articulations than adults' vowel productions. I found something similar in an acoustic investigation of unstressed vowels produced by 5-year-olds, 8-year-olds, and adults (Redford, 2018), but other

findings were that anticipatory V-to-C effects on F1 were stronger in children's speech than in adults' speech.

In fact, findings from recent ultrasound studies on coarticulation in children's and adults' speech strongly suggest that children's speech is *more* coarticulated than adults' speech (Zharkova, Hewlett, & Hardcastle, 2011; 2012; Noiray, Menard, & Iskarous, 2013; Noiray, Abakarova, Rubertus, Kruger, & Tiede, 2018; but see Barbier, 2016, for an alternative view). For instance, Zharkova and colleagues' (2011) used ultrasound to investigate C-to-V coarticulation in school-aged children's and adults' production of /ʃV/ syllables in the frame sentence "It's a __ Pam." They found that children's production of the palatal-alveolar fricative was more influenced by the following vowel than adults' productions (see also Zharkova, Hewlett, & Harcastle, 2008). Noiray and colleagues (2018) studied coarticulation degree across a wider age range and more consonantal and vocalic contrasts. Their results showed that coarticulation degree becomes weaker with age. In particular, they found that preschool children's articulation of labial, alveolar, and velar stop consonants were all more influenced by the following vowel than school-aged children's articulation of these consonants, and that coarticulation degree was stronger in school-aged children's productions than in adults' productions. These and other similar results are opposite the prediction from the information processing hypothesis that phonemes provide a basis for speech acquisition and production.

In contrast to the information processing approach, the ecological dynamics approach to speech production predicts that children's speech is more coarticulated than adults' (Nittrouer, Studdert-Kennedy, & McGowan, 1989; Nittrouer, 1993; 1995; Nittrouer, Neeley, & Studdert-Kennedy, 1996; see also Noiray et al., 2013; Noiray et al., 2018). For example, Nittrouer (1995) hypothesized that children's early word productions are articulatory gestalts, and that "the emergence of mature production skills involves two processes: differentiation and tuning of individual gestures, and improvement in coordination among gestures that compose a word" (p. 521). The hypothesis aligns

well with a functional approach to child phonology, which emphasizes the communicative intent behind spoken language production and so argues for word-based analyses of children's speech sound patterns (e.g., Waterson, 1971; Ferguson & Farwell, 1975; Menn, 1983; Stoel-Gammon, 1983; Vihman, et al., 1985; Vihman & Croft, 2007; Vihman, 2017). In fact, Nittrouer and colleagues (1989:120-121) explicitly motivated their prediction that children's speech is more coarticulated than adults' with reference to two of the papers that first introduced the idea of that child phonology should take the word as its principle unit of analysis (see "setting papers" in Vihman & Keren-Portnoy, 2013). Following Ferguson and Farwell (1975), they suggested that a child's failure to appropriately generalize correct phonetic forms (e.g., [n] and [m]) from one word to another (e.g., "no" is [noʊ], but "night" is [mɑɪt] whereas "moo" is [bu:]) indicated that whole words, rather than phonemes, were the targets of acquisition and also the units of production. Nittrouer and colleagues also referred to Ferguson and Farwell's observation of children's variable word realizations to argue for an account of word form representation as a "collection of gestures" that were inappropriately timed and so genuinely more gestalt-like than segment-like. Finally, they cited Menn's (1983) analysis of consonant harmony in her son's first words to make a point about the existence of "articulatory routines" for word production.

In sum, children's speech patterns are more compatible with the hypothesis of whole word production than with the hypothesis of phonemic, or segmental, production. In so far as the systematic patterns of child phonology can also be explained to emerge from motoric constraints (see e.g., Locke, 1983; McCune & Vihman, 1987; Davis, MacNeilage, & Matyear, 2002; inter alia), the ecological dynamics emphasis on action-based representations is also more compatible with children's speech patterns than the information processing emphasis on sequencing constraints derived from a child-specific grammar. For this reason, I deem holistic motoric word-form representations fundamental to a developmentally sensitive theory of speech production.

Explaining Phonological Change Over Developmental Time

As in Redford (2015), the specific proposal is that children begin to acquire holistic motoric representations, or schemas, with their attempts at first words. These schemas then provide the basic speech plan for future word productions. This proposal begs the developmental question: how do schema representations change over time as children's speech becomes more and more adult-like? Here, I argue that the information processing assumption of separate perception and production systems is required to account for developmental change. To make this argument, let us first consider development from the ecological dynamic perspective.

In an ecological dynamics approach, learning is an attunement process (Studdert-Kennedy, 1987; Goldstein & Fowler, 2003). Unsuccessful communication destabilizes representations that encode timing relations between gestures forcing a random walk through motor space until the word-specific timing patterns have been discovered (see, e.g., Nam et al., 2009). This mode of phonological learning implies that the temporary, but systematic patterns of child phonology represent local minima in the random walk. This implication is consistent with articulatory constraint based explanations for these patterns (e.g., Locke, 1983; McCune & Vihman, 1987; Davis & MacNeilage, 2000; Davis et al., 2002). But, similar to the constraint-based explanations, the assumption of a self-organized system based on dynamic principles predicts a universal pattern of speech development, albeit one that interacts in predictable ways with the target language. This prediction is undermined by the strong individual differences in speech development that are observed within a language (e.g., Ferguson & Farwell, 1975; Macken & Ferguson, 1981; Stoel-Gammon & Cooper, 1984; Vihman, Ferguson, & Elbert, 1986; *inter alia*).

Ferguson & Farwell (1975) were among the first to take individual differences in development seriously and to propose, in effect, that these signal the child's control over the speech production process. The specific suggestion was that children select word forms from the adult language that they are able to produce. Word selection implies a kind of insight into the production process

meted out by an executive controller—an implication that is anathema to the ecological dynamics approach. McCune and Vihman (1987; 2001) better defined the "what" of what children are able to produce when they proposed that children build up a unique set of vocal motor schemes during babbling based on individual preferences for particular patterns. Vihman (1996) then recast the notion of selection with respect to these schemes. She proposed that a scheme acted as a kind of 'articulatory filter' that "selectively enhances motoric recall of phonetically accessible words" (p. 142). Elsewhere, Vihman (2017) refers to resonances between the production and perception systems to explain the selective memory for phonetically accessible words. In this way, Vihman is able to explain individual differences in production and word forms attempted while avoiding the homunculus problem inherent to the concept of an executive controller.

Although the idea of an articulatory filter very much implies interactions between action and perception, the specific theory of perception Vihman adopts is very clearly not a direct realist one; for example, elsewhere Vihman is interested in the role of perceptual saliency in children's development of lexical representations (e.g., Vihman, Nakai, DePaolis, & Hallé, 2004). The notion of perceptual saliency relies on the psychoacoustic theory of speech perception that undergirds the information processing approach of speech production; that is, a theory of perception in which the perceptual primitives are "intrinsically meaningless, simple acoustic features, such as spectral distribution patterns, bursts of band-limited aperiodic noise, … into which the speech signal can be analyzed (Best, 1995:175)." Why does Vihman adopt this theory? Probably because a psychoacoustic theory of speech perception provides targets of acquisition that go beyond a child's immediate abilities and so allow for directed motor learning and change (see also Menn, Schmidt, & Nicholas, 2013). More generally, a psychoacoustic theory of speech perception explains a wider variety of speech-related phenomena than a direct realist theory; for example, it accounts for categorical perception in nonhuman animals and why auditory processing constraints appear to affect the structure of phonological systems (see Diehl, Lotto, Holt, 2004, for a review).

In sum, the observation that individual children take very different paths to acquire the same spoken language suggests a developmental process more compatible with the information processing assumption of distinct perception and production systems than with the ecological dynamics assumption of a unified perception-action system. The developmentally sensitive theory to speech production described below further assumes that distinct production and perception systems entail a role for central perceptual feedback control in speech production.

A Developmental Approach to Speech Production

The developmentally sensitive theory of speech production outlined in this section extends the basic idea, first outlined in Redford (2015), that adult speech production processes and representations are structured by the acquisition of spoken language. The alternative view, implicit in mainstream theory, is that adult speech production processes and representations are the targets of spoken language acquisition. As in Redford (2015), the theory assumes that the fundamental unit of production is a word. This assumption follows from the view that "the child's entry into language is mediated by meaning: and meaning cannot be conveyed by isolated features or phonemes" (Studdert-Kennedy, 1987:51). Similar to an ecological dynamics approach, endogenous representations are assumed to be holistic and action-based. As in Redford (2015), I call these representations schemas, not gestural scores or coupling graphs, to acknowledge borrowing from Vihman and McCune's theoretical work on child phonology (McCune & Vihman, 1987; Vihman & McCune, 1994; McCune & Vihman, 2001) and debts to schema theory in the area of skilled action and motor control (Schmidt, 1975; Norman & Shallice, 1986; Arbib, 1992; Cooper & Shallice, 2006). These acknowledgments also signal the aforementioned embrace of certain information processing assumptions; namely, that production and perception are distinct processes and that adults implicitly predict perceptual outcomes and use perceptual feedback to make articulatory (and whole word) adjustments while speaking.

In addition to building on these assumptions, the developmentally sensitive theory outlined here emphasizes two distinctions: (1) the distinction between others' productions and self-productions; and (2) the distinction between self-productions for oneself and self-productions for others. Self-productions provide a basis for endogenous representations. When these are for oneself, they are assumed to be exploratory and so free from association with conceptual information. In this way, they provide the basis for the non-linguistic perceptual-motor map that is used to integrate exemplar and schema representations for production. When self-productions are for others, they are assumed to be communicative and associated with conceptual information. In this way, they provide the basis for schemas. In contrast to self-productions, others' productions provide the basis for just one type of representation—an exogenous perceptual representation associated with conceptual information. I will call this representation a perceptual exemplar. This label acknowledges inspiration from a class of phonetically-informed phonological theories that emphasize the importance of detailed, often word-specific, acoustic-phonetic information for production (e.g., Pierrehumbert, 2002; Bybee, 2006; Johnson, 2007). Perceptual exemplars provide production targets. A child cannot even attempt first words without having acquired at least a few of these from the ambient language.

The foundational assumptions enumerated above entail speech plan representations that are different from either the ecological dynamics or information processing approaches to speech production. They also entail a different approach to phonology than the ones alluded to so far. Otherwise, the developmentally sensitive theory proposed here borrows heavily from current models of speech production and motor control. It contributes to the field by accounting for the transition from pre-speech to adult-like speech in a series of steps that correspond to major developmental milestones.

Step 1: The Perceptual-Motor Map

As in an information processing approach to speech production, a developmental approach requires a perceptual-motor map; specifically, a mapping between auditory speech and articulatory movement that is likely mediated by somatosensory information (e.g., Perkell, Matthies, Svirsky, & Jordan, 1995; Guenther, 1995; Guenther et al. 2006). The existence of a perceptual-motor map is supported by neuropsychological findings on sensorimotor integration in different regions along the auditory dorsal stream pathway from primary auditory cortex (= superior temporal gyrus, superior temporal sulcus) to anterior premotor cortex (= inferior frontal gyrus; see Hickok & Poeppel, 2007). It is common to assume that the perceptual-motor map develops during the first year of life as infants engage in vocal exploration (e.g., Guenther, 1995; Davis & MacNeilage, 2000; Kuhl, 2000; Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Menn et al., 2013). Following Oller (2000:165-179), I will assume that this exploration includes all pre-speech vocalizations from cooing to squealing to babbling, and so describes the mapping of continuous acoustic and motor dimensions, with somatosensory information at the intersection of these two. For example, it associates the frequency sweeps of squealing with continuous changes to the length and tension of the vocal folds and the amplitude modulated frication of raspberries with the forcing of air through loosely coupled lips. It also associates static sounds, like silence, to transient actions in the vocal tract, such as a briefly sustained oral or glottal closure. This view of the perceptual-motor map enables the gestural interpretation of acoustic form (*cf*. Best, 1995; see also, Hickok, 2012; 2014), and so can take holistic representations as input.

Although the map develops during the pre-speech period of infant vocalization, it is important to stipulate that it continues to evolve with the acquisition of speech motor skills and across the lifespan with the acquisition of new languages and with conformity to or disengagement from the sociolinguistic environment (see Kuhl, Ramirez, Bosseler, Lin, & Imada, 2014, for a related view). In the context of the current theory, this assumption is required to explain developmental changes

that are traditionally attributed to the phonology; that is, the evolution of word forms from child-like to more adult-like. This is because the perceptual-motor map provides a source for the abstract action-based word form representations that are schemas, as described below.

Step 2: Perceptual Word Forms and Action Schemas

Children's first words mark the onset of speech production. Word production depends on conceptual development, including the insight that adult vocalizations are referential. This insight, which occurs perhaps as early as 7 months of age (Harris, Yeeles, Chasin, & Oakley, 1995; Bergelson & Swingley, 2012), coincides with the acquisition of perceptual word forms—exemplars—from the ambient language. Bergelson and Swingley (2012) provided evidence for this claim when they used eye-tracking to assess 6- to 9-month-old infants' ability to comprehend familiar nouns by discriminating between paired pictures while listening to spoken stimuli, (e.g., "Can you find the *X*?" and "Where's the *X*"). The authors reported that infants as young as 6 months of age were reliably able to discriminate a significant number of the pairs. Note that, by most accounts, perceptual attunement to the native language occurs between 6 and 10 months of age (see Vihman, 2017, for a review). Bergelson and Swingley therefore interpreted the finding to indicate that learning the sounds of a language goes hand-in-hand with learning its vocabulary.

At around 12 months of age, the infant has acquired both a reasonably stable perceptual lexicon and a perceptual-motor map. The production of first words is now possible. This heralds the onset of speech production, which is imagined here as the moment when the infant, motivated to communicate a specific referential meaning, uses her perceptual-motor map to translate an exogenously-derived perceptual exemplar into vocal action. As in Redford (2015), I assume that the motor routines an infant first uses to convey a particular concept are abstracted and associated with that concept when the child has succeeded in communicating the intended meaning. This abstraction is the schema. Similar to gestural scores, schemas encode routine-specific relational information between articulators across time; for example, tongue advancement during jaw

opening. Similar to coupling graphs, they are the action-based word form representations. Put another way, schemas are both the phonological representation and speech plan for a given word/concept, where word is broadly construed as any conventionalized form-meaning association that is part of the child's repertoire (e.g., "uh oh" or "gimme" for "give me"). Figure 2 depicts first word production and schema abstraction.

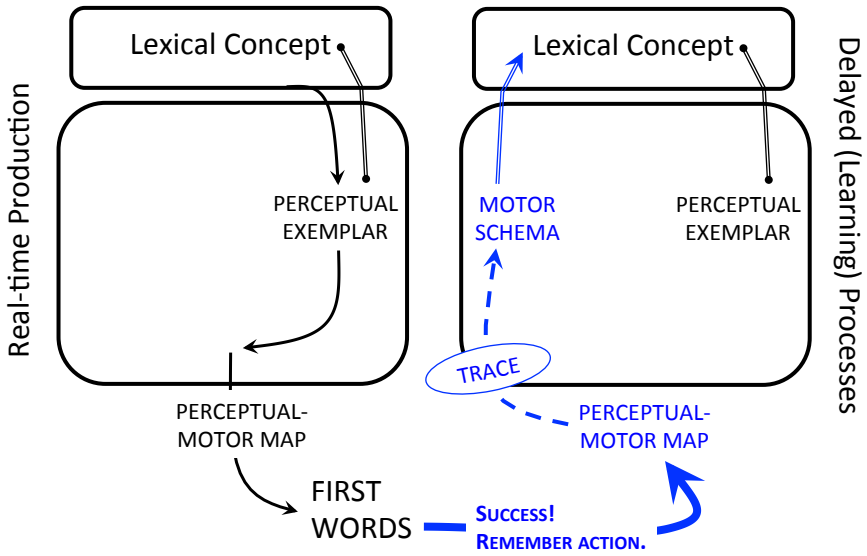

Figure 2. The onset of speech coincides with attempts to produce specific meanings (concept) associated with perceptual word forms learned from the ambient language (left). Specifically, infants engage their perceptual-motor map to derive a best motoric approximation of the exogenous perceptual form, or "perceptual exemplars". The shape of the approximation will depend on how the map has been warped through vocal exploration, which itself is constrained by motor development. The motor routines used to convey specific concepts are abstracted and stored during production (right). These abstractions, or "motor schemas", are associated with the concept attempted and so serve as one half of the phonological representation of a word. Solid lines with arrows represent feedforward processes; dotted lines with arrows represent feedback processes.

Schemas are continually updated with production. This means that they become more abstract over time as the a one-to-one relationship with a single motor routine gives way to timing generalizations that are common to all attempts of a particular word production. Note that the protracted development of articulatory timing control, which results in highly variable speech output, ensures that the schema-encoded generalizations become abstract quite quickly. Ultimately, schemas may encode little else than number of syllables as iterations of the open-close cycle of the

vocal tract, the relative durations of these cycles, plus the initial posture and direction of major articulators for each cycle. This hypothesis is consistent (or at least reconcilable) with evidence for serial timing control and frame-based plans generated in the supplementary motor area (SMA) and pre-SMA, respectively, during adult speech production (see, e.g., MacNeilage, 1998; Bohland & Guenther, 2006).

Step 3 : Onset of Perceptually-based Control

Once schemas are abstracted, they are activated with the perceptual form when a concept is selected for production. The motor and perceptual forms are integrated in the perceptual-motor map. Hickok, Houde, and Rong (2011:413) adopt a similar hypothesis, albeit with an emphasis on sensorimotor integration at the level of phoneme production. They note that the hypothesis "is consistent with Wernicke's early model in which he argued that the representation of speech, e.g., a word, has two components, one sensory (what the word sounds like) and one motor (what sequence of movements will generate that sequence of sounds)." Wernicke's exact hypothesis of dual word form representations is adopted here to explain both why child forms deviate from adult forms and how the forms change over time.

With respect to children's deviant forms, schemas are assumed to initially weight production in such a way that it appears motorically constrained. The weighting is the result of a very small productive vocabulary, which serves to entrench particular trajectories through motor space. For awhile, this entrenchment may even limit the child's ability to form new motor trajectories. At this stage, children's productions of novel words may appear more template-like than in first word production. In Vihman and Croft's (2007:696) words: "the child (implicitly) impos(es) one or more preexisting templates, or familiar phonological patterns, on an adult form that is… similar to those patterns."

Around 18 months of age, significant vocabulary expansion results in a developmental shift away from forms that suggest production constraints and towards those that suggest perceptual

ones due to increasing homophony among expressive word forms (Redford & Miikkulainen, 2007). This shift heralds the next critical step in the evolution of speech production: a newfound focus on how self-productions *should* sound. The onset of predictive encoding (state feedback control) emerges from this focus.

In particular, the proposed process by which the 18-month-old infant begins to forge new paths through motor space takes as its inspiration the hierarchical state feedback control model of production (Hickok et al., 2011; Hickok, 2012; 2014), where state feedback control is described as having two functions. The first is to adjust motor commands so that the articulators reach desired perceptual targets; the second is to use external feedback to update the representations that guide speech. In the present proposal, both functions are thought to emerge with a communication-driven shift in production towards better matching endogenous to exogenously-derived perceptual forms. Further, function 2 is proposed to drive function 1 in that function 1 may begin as a delayed comparison between the perceptual trace of a production and the intended target, absent any motor adjustments (Figure 3).

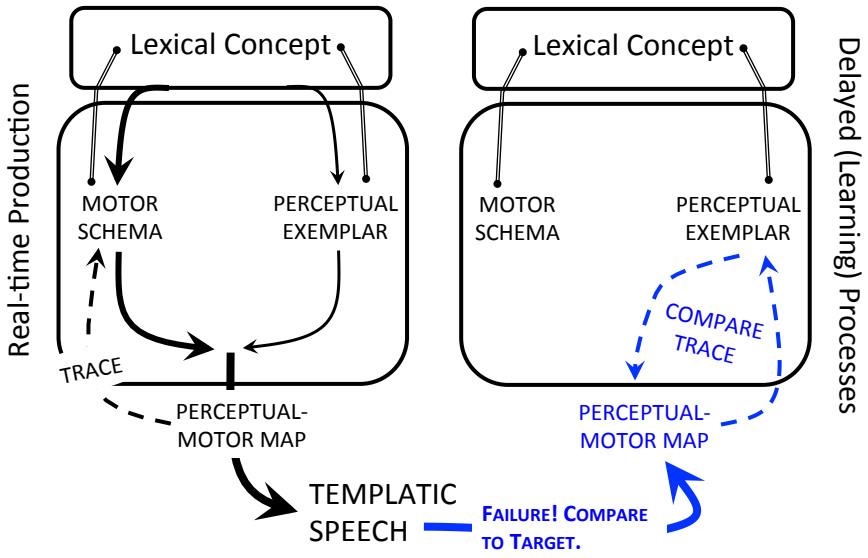

Figure 3. Following early word production, the next major developmental change is hypothesized to occur when motorically-driven homophony begins to threaten the young child's ability to effectively communicate. At this stage, the child begins to focus on how words should sound. As a result, production shifts from an entirely feedforward process to one where feedforward routines

are adjusted to match perceptual representations. The adjustment process, carried out through interactions between the endogenous perceptual-motor map and the repository of exogenous word form representations, or "perceptual exemplars", sets the stage for state feedback control (SFC), which nonetheless begins with a delayed comparison between the perceptual trace and target absent adjustment (left). Solid lines with arrows represent feedforward processes; dotted lines with arrows represent feedback processes.

How might a delayed matching evolve into real time state feedback control? One possibility is that the matching process creates a bidirectional connection between the exogenously-derived exemplar targets and the perceptual-motor map, where the connections between motor routines and perceptual patterns are already robust and bidirectional. Now, the perceptual outcomes of schema-associated routines can be matched in real time against perceptual exemplars. Any discrepancies between the expected self-outcomes and other-based representations could force new paths through motor space by stretching entrenched motor routines in the direction of the exogenously-derived perceptual form.

Step 4 : Self-Monitoring

Speech production does not become adult-like until children begin to externally monitor their own speech and consciously recognize its divergence from (chosen) adult norms. The evidence suggests that this may not occur until around age 4 years. In particular, feedback perturbation experiments with young children suggest that perceptual input plays little role in speech production before age 4; for example, toddlers neither immediately compensate nor adapt over time with articulatory changes to their vowel productions when hearing spectrally perturbed alterations of their own speech during a word production task (MacDonald, Johnson, Forsythe, Plante, & Munhall, 2012). At age 4 years, children begin to compensate, but do not adapt over the long term to perturbed feedback (Ménard, Perrier, Aubin, Savariaux, & Thibeault, 2008; MacDonald et al., 2012); for example, Ménard and colleagues showed that 4-year-old children return immediately to preferred productions after compensating on-line to an articulatory perturbation. Failures to adapt suggest that although 4-year-old children may use auditory information to help guide speech production, they do not yet use external feedback to update existing production

representations and processes. Still, the ability to adapt appears to emerge soon after 4 years of age in typically developing children (Terband, Van Brenk, van Doornik-van der Zee, 2014).

Psycholinguistic evidence is consistent with the hypothesis that self-monitoring emerges late in the preschool years during spoken language development. For example, preschool children understand unfamiliar adult speech better than their own unadult-like speech (Dodd, 1975). In addition, self-initiated speech repairs increase over developmental time, with many fewer repairs observed in the speech of 5-year-old children than in the speech of older school-aged children (Rogers, 1978; Evans, 1985). Moreover, if we imagine the self-monitoring process as one where the speaker must identify particular discrepancies between what they intended to produce and what they actually produced, then its slow development is consistent with the slow development of selective attention (see, e.g., Plude, Enns, & Brodeur, 1994; Wellman, Cross, & Watson, 2001). The speculation here is that selective attention to ones own speech is motivated also by a developing self-concept. When the child begins to appreciate those aspects of her own speech that signal an undesired social distance between herself and others, she shifts her attentional focus to identifying discrepancy between how she sounds and who she wants to sound like. This motivates a final marked disruption of entrenched motor routines in service of better approximating the exogenously-derived exemplars.

Self-concept emerges with theory of mind during the preschool years (see Symons, 2004). Self-identity, which is part of the self-concept (Gecas, 1982; Baumeister, 1999), manifests in speech with socio-indexical marking. For example, voice onset time for stops varies differently as a function of gender across languages (Whiteside & Irving, 1998; Oh, 2011; Li, 2013), suggesting social as opposed to physiological reasons for this speech production difference. How does the child acquire female versus male gendered speech? The suggestion here is that a burgeoning sense of identity leads the child to selectively attend to those adult productions she is most interested in approximating. In identifying a discrepancy between how they sound and who they want to sound

like, children may highlight exemplars associated with those individuals thereby highlighting aspects of the perceptual form that need special attention in production. At the same time, self-monitoring focuses more attention on the perceptual consequences of one's own speech, which further increases the weight of exemplars in the production process, thus pushing motor routines and resulting schema ever more in the adult direction (Figure 4).

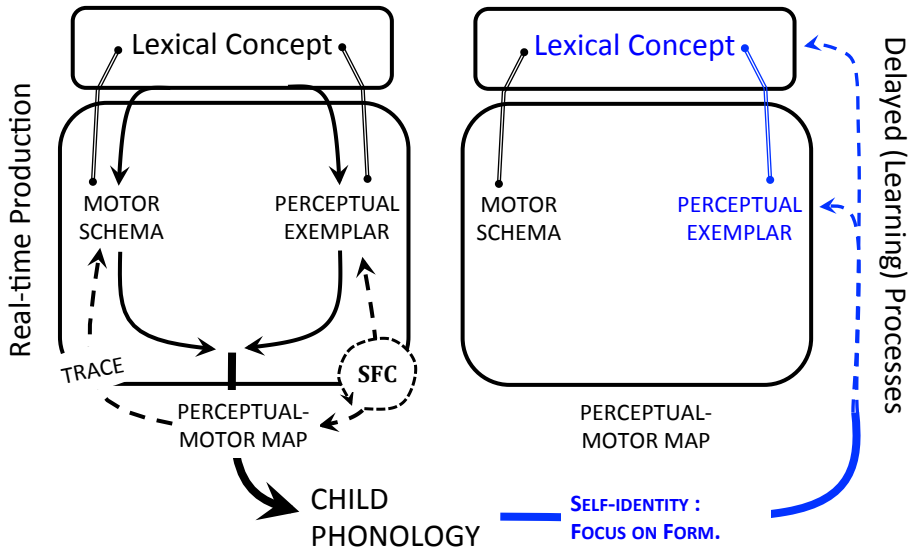

Figure 4. During the preschool years, children begin to self-monitor based on external perceptual feedback to identify deviations between how they sound and who they want to sound like. The perceived deviations highlight aspects of the stored perceptual representations, driving the perceptual-motor mapping and resulting endogenous motoric representations (i.e., schemas) ever more towards matching exogenous perceptual goals (i.e., exemplars). Solid lines with arrows represent feedforward processes; dotted lines with arrows represent feedback processes.

Thus, the full proposal is that during the preschool years socially-directed listening induces changes in speech production through a self-monitoring led shift towards perceptually-weighted production. Prior to this point, self-productions are (unconsciously) heard as being the same as other productions. Consider, for example, the toddler who points to a picture of a fish in a picture book and utters "fifth," to which the parent responds "fifth?" and the child answers, "No, fifth!" (see Menn, 1983). Updates to both the perceptual-motor map and schema representations follow from this shift, soon resulting in adult-like representations. This proposed final stage in the development of speech production is consistent with the evidence that socio-indexical information, such as

gender-specific use of phonetic features, begin to emerge in children's speech around age 4 years (see Foulkes & Docherty, 2006:422-424). This observation bring us back to an earlier one that closes the gap between work in speech motor control and real-world speaker behavior; namely, the observation that participants' behavior in auditory feedback perturbation experiments resembles phonetic convergence, normally understood as a socially-driven behavior meant to lubricate interactions between interlocutors.

Discussion

Current approaches to speech production aim to explain adult behavior and in so doing frequently make at least some assumptions that, when taken to their logical conclusion, fail to adequately account for how the system develops. This failure is problematic from a developmental perspective. According to this perspective, the representations and processes of adult speech and language should emerge from the developmental process (for a similar view see Vihman & Croft, 2007; Menn et al., 2013).

Development is particularly relevant for theories of speech production because of the paradox of early speech onset despite slowly developing speech motor control. Here, this paradox was taken to suggest the working hypothesis that feedforward processes mature earlier than central feedback control processes in speech production. This hypothesis structured a developmentally sensitive theory of speech production that was elaborated in stages, with each stage building on the previous one. The stages proposed were designed to accommodate developmental patterns. At the same time, developmental patterns were given new meaning and grouped in novel ways by the working hypothesis. The accommodation of speech production theory to developmental findings and vice versa results in many new testable hypotheses that could motivate future empirical work and usher in new knowledge and even new clinical practice. For example, the hypothesis that perceptual-motor integration relies on the development of a non-linguistic perceptual-motor map suggests that therapeutic uses of speech sound practice should cover as broad a range of sound combinations as

possible. By hypothesis, these sound combinations need not be tied to lexical content and so the therapy could involve a fun and silly random sound sequence generating game using, say, magnetic letters that could be arranged and then rearranged on a board. Such a game would allow the set of possible sound combinations in a language to be more fully explored than is possible when that set is constrained by picturable words in the language. The benefits of this therapy for generalization to novel or known word production could be tested against current therapies where speech sound practice typically involves the use of visual props to elicit specific lexical items. Intriguingly, this idea echoes to some extent Gierut's (2007) differently motivated contention that words with complex speech sound sequences allow for better generalization of treatment in children with phonological disorder than words that have simple phonological structure.

The hypothesized disassociation of the perceptual-motor map and perceptual exemplar representation of word forms also has implications for the clinical assessment of speech sound disorder. For example, when this hypothesis is taken together with the idea articulatory change is motivated by weighting perceptual exemplar representations more heavily during production it suggests that the aforementioned fun and silly random sound sequence generating game could be used to supplement a comprehensive evaluation of speech sound disorder. Performance in the game could help diagnose whether the articulation problem is due to a poorly developed perceptual-motor map or to poorly specified perceptual exemplars. The diagnosis would then lead to therapy that focuses either on speech sound practice or on developing perceptual exemplars. Finally, the theory-dependent hypothesis that perceptual weighting of production is driven in part by the emergence of a self-concept and the ensuing selective attention to self-productions suggests not only a testable hypothesis regarding the development of convergence behaviors in spoken language interactions, but also a novel way to understand the absence of convergence behaviors and other mild segmental speech sound disorders in individuals on the autism spectrum.

Another major implication of the developmentally sensitive theory elaborated in this paper is a new adult model of speech production. This model, illustrated in Figure 5, incorporates insights from many existing theories. Some of these insights were explicitly acknowledged in the preceding text; others were merely implied. For example, the reference to "self-monitoring" indicates an acceptance of the evidence in favor of this well-established hypothesis (see Postma, 2000, for a review). Otherwise, the model diverges from most adult-focused theories in assuming distinct action-based and perception-based representations (though see Hickok, 2012; 2014). This aspect of the model provides a framework for understanding phenomena that have been traditionally ignored in adult-focused theories of speech production. For example, the model very obviously allows for the different possible speaking modes that are thought to correspond with speaking style differences: one in which the motor pathway is emphasized over the perceptual pathway—this is Lindblom's (1990) hypo or system-oriented mode; one in which the reverse occurs—this is Lindblom's hyper or output-oriented mode (shown); and a mode in which the two pathways are in equilibrium—this is likely the default mode.

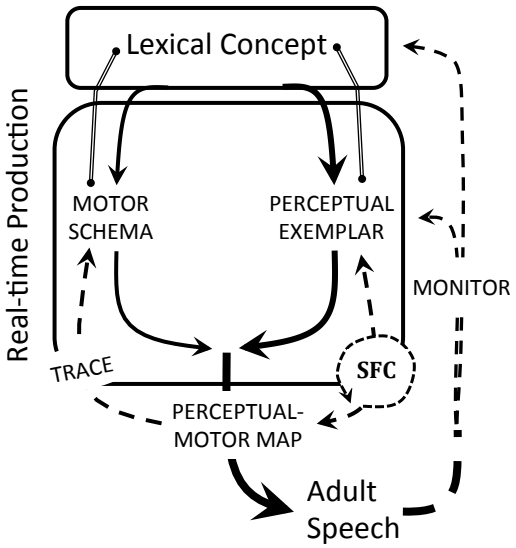

Figure 5. The adult model of speech production implied by the developmental model outlined in this paper. Solid lines with arrows represent feedforward processes; dotted lines with arrows represent feedback processes. The linkages between the repository of lexical concepts and motor

schemas and between lexical concepts and perceptual exemplars represent the conceptual and phonological aspects of the lexicon.

The implied adult model shown in Figure 5 also diverges from information processing theories in assuming that holistic phonological representations serve as speech plan representations. This developmentally sensitive aspect of the model is not immediately compatible with the evidence for sublexical units in productions, including the speech error data that have long been used to argue for the psychological reality of a phonological encoding process. The developmentally sensitive adult model automatically fails if it cannot account for these data. Accordingly, we are currently pursuing the hypothesis that discreteness emerges at the level of the perceptual-motor map (Davis & Redford, in prep). More specifically, we have formally defined the perceptual-motor map as a linked set of experienced perceptual and motor trajectories that are time-based excursions through speaker-defined perceptual and motor spaces. By hypothesis, nodes appear where motor trajectories intersect in motor space, creating perceptually-linked node-delimited paths that can be recombined. Though weighted in the direction of already experienced paths, exemplar-driven novel word production picks new trajectories through motor space by deforming existing node-delimited paths in systematic ways. These new trajectories may intersect existing trajectories or go on to be intersected themselves. In this way, motor space is reticulated with vocabulary acquisition and discrete speech motor goals emerge absent discrete phonological representations. In future work, we will investigate how this view of discreteness might account for the speech error data. Our initial hypothesis is that these arise from the competing motoric and perceptual pressures of schema and exemplar integration during speech production.

Conclusion

Theories of spoken language production provide frameworks for understanding developmental speech sound disorders. Even the distinction between motor speech, articulation, and phonological disorders reflects this fact. In so far as the types of interventions chosen to address a disorder

follow from how the disorder is understood, theory informs practice. This is as it should be. But the relationship between theory and practice should also motivate a reconsideration of theory when it fails to address a problem that is relevant to practice. The problem of development clearly falls into this category. A major aim of this paper was to show that current adult-focused approaches to speech production fail to address the paradox of slow developing speech motor control in spite of early speech onset because they depart from perspectives that are not developmental. A developmental perspective assumes change over time and those who adopt it focus on explaining how this change occurs. A second major aim of this paper was to show how a commitment to this perspective leads to a theory of speech production that is different in many respects from existing theories. Thus, even if the various ideas presented herein are dismissed after testing, the conclusion should be that a developmental approach to understanding speech production should be pursued if theory is to be useful for practice.

Acknowledgments

The author is grateful to several anonymous reviewers, Maya Davis, and Jill Potratz for their comments on a previous version of this paper. Manuscript preparation was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) under grant R01HD087452. The content is solely the author's responsibility and does not necessarily reflect the views of NICHD.

REFERENCES

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology*, 51(4), 705-723.

Arbib, M. A. (1992). Schema theory. In S. Shapiro (ed.), *The Encyclopedia of Artificial Intelligence*, *Vol. 2* (pp. 1427-1443). Wiley.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 177-189.

Barbier, G. (2016). *Contrôle de la production de la parole chez l'enfant de 4 ans : l'anticipation comme indice de maturité motrice*. PhD thesis. Université Grenoble Alpes. Français. NNT : 2016GREAS013

Baumeister, R. F. (1999). Self-concept, self-esteem, and identity. In V. J. Derlega, B. A. Winstead, & W. H. Jones (eds.), *Nelson-Hall Series in Psychology. Personality: Contemporary Theory and Research* (pp. 339-375). Nelson-Hall Publishers.

Best, C. T. (1995). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). MIT Press.

Best, C. T., Goldstein, L. M., Nam, H., & Tyler, M. D. (2016). Articulating what infants attune to in native speech. *Ecological Psychology*, 28(4), 216-261.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9), 3253-3258.

Bladon, R. A. W., & Al-Bamerni, A. (1976). Coarticulation Resistance in English/l/. *Journal of Phonetics*, 4(2), 137-150.

Bock, K., & Levelt, W. (2002). Language production. In G.T. Altman (ed.), *Psycholinguistics: Critical Concepts in Psychology*, (pp. 405-450). Routledge.

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.

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, 45(2-4), 140-155.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6(2), 201-251.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4), 155-180.

Daniloff, R., & Hammarberg, R. (1973). On defining coarticulation. *Journal of Phonetics*, 1(3), 239-248.

Cooper, R. P., & Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychological Review*, 113(4), 887-916.

Davis, B. L., & MacNeilage, P. F. (2000). An embodiment perspective on the acquisition of speech perception. *Phonetica*, 57(2-4), 229-241.

Davis, B. L., MacNeilage, P. F., & Matyear, C. L. (2002). Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production. *Phonetica*, 59(2-3), 75-107.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149-179.

Dodd, B. (1975). Children's understanding of their own phonological forms. *The Quarterly Journal of Experimental Psychology*, 27(2), 165-172

Evans, M. A. (1985). Self-initiated speech repairs: A reflection of communicative monitoring in young children. *Developmental Psychology*, 21(2), 365-371.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language* 51, 491-439.

Flege, J. E. (1988). Anticipatory and carry-over nasal coarticulation in the speech of children and adults. *Journal of Speech, Language, and Hearing Research*, 31(4), 525-536.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, 34(4), 409-438.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8(1), 113-133.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective, *Journal of Phonetics*, 14, 3-28.

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36(2-3), 171-195.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361-377.

Garrett, M. F. (1988). Processes in language production. *Linguistics: the Cambridge Survey*, 3, 69-96.

Gecas, V. (1982). The self-concept. *Annual Review of Sociology*, 8(1), 1-33.

Gierut, J. A. (2007). Phonological complexity and language learnability. *American Journal of Speech-Language Pathology*, 16(1), 6-17.

Giles, H., & Powesland, P. (1997). Accommodation theory. In N. Coupland & A. Jaworski (eds.), *Sociolinguistics* (pp. 232-239). Palgrave, London.

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M.A. Arbib (ed.), *Action to Language via the Mirror Neuron System*, (pp. 215-249). CUP.

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N.O. Schiller & A. Meyer (eds), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 159-207). de Gruyter.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251-279.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*, 21(7-8), 817-855.

Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research*, 43, 239-255.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102(3), 594-621.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280-301.

Guenther, F. H., Perkell, J. S., (2004). A neural model of speech production and its application to studies of the role of auditory feedback in speech. In Maassen, B., Kent, R. D., Peters, H. F. M., van Lieshout, P. H. H. M., & Hulstijn, W. (eds), *Speech Motor Control in Normal and Disordered Speech*, (pp. 29-49). OUP.

Haken, H., Kelso, J. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, 51(5), 347-356.

Harris, M., Yeeles, C., Chasin, J., & Oakley, Y. (1995). Symmetries and asymmetries in early lexical comprehension and production. *Journal of Child Language*, 22(1), 1-18.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135-145.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, 29(1), 2-20.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, 15(5), 673-682.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, 69(3), 407-422.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4(4), 131-138.

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5, 82. https://doi.org/10.3389/fnhum.2011.00082

Johnson, K. (2007). Decisions and mechanisms in exemplar-based phonology. In. M.-J. Sole, P. Beddor, & M. Ohala. (eds.), *Experimental Approaches to Phonology* (pp. 25-40). OUP.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, 505-528.

Kager, R., Pater, J., & Zonneveld, W. (Eds.). (2004). *Constraints in phonological acquisition*. CUP.

Katseff, S., Houde, J., & Johnson, K. (2012). Partial compensation for altered auditory feedback: a tradeoff with somatosensory feedback? *Language and Speech*, 55(2), 295-308.

Keating, P. A. (1990). The window model of coarticulation: articulatory evidence. In J. Kingston & M.E. Beckman (eds.), *Papers in Laboratory Phonology I*, 451-470. CUP.

Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, 112-156.

Kelso, J. A., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, 14(1), 29-59.

Kent, R. D. (1983). The segmental organization of speech. In P. F. MacNeilage (ed.), *The Production of Speech* (pp. 57-89). New York: Springer-Verlag.

Kent, R. D., & Forner, L. L. (1980). Speech segment duration in sentence recitations by children and adults. *Journal of Phonetics*, 8, 157-168.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, 97(22), 11850-11857.

Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J. F. L., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *Proceedings of the National Academy of Sciences*, 111(31), 11238-11245.

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*, 32(27), 9351-9358.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, 105, 1455-1468.

Levelt, W. J. (1989). *Speaking: From intention to articulation*. MIT press.

Li, F. (2013). The effect of speakers' sex on voice onset time in Mandarin stops. *Journal of the Acoustical Society of America*, 133(2), EL142-EL147.

Lindblom, B. E. (1964). Articulatory activity in vowels. *Journal of the Acoustical Society of America*, 36(5), 1038-1038.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W.J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403-439). Springer, Dordrecht.

Lindblom, B., Lubker, J., & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of motor programming by predictive simulation. *Journal of Phonetics*, 7, 147-161.

Löfqvist, A. (1990). Speech as audible gestures. In H.W. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modelling* (pp. 289-322). Springer, Dordrecht.

Locke, J. L. (1983). *Phonological acquisition and change*. Academic Press.

MacDonald, E. N., Goldberg, R., & Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *Journal of the Acoustical Society of America*, 127(2), 1059-1068.

MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., & Munhall, K. G. (2012). Children's development of self-regulation in speech production. *Current Biology*, 22(2), 113-117.

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, 8(3), 323-350.

Macken, M. A., & Ferguson, C. A. (1981). Phonological universals in language acquisition. *Annals of the New York Academy of Sciences*, 379(1), 110-129.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review*, 77(3), 182-196.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21(4), 499-511.

Mandler, G. (2007). *A history of modern experimental psychology*. MIT Press.

McCune, L., & Vihman, M.M. (1987). Vocal motor schemes. *Papers and Reports on Child Language Development*, 26, 72-79.

McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language, and Hearing Research*, 44(3), 670-684.

Ménard, L., Perrier, P., Aubin, J., Savariaux, C., & Thibeault, M. (2008). Compensation strategies for a lip-tube perturbation of French [u]: An acoustic and perceptual study of 4-year-old children. *Journal of the Acoustical Society of America*, 124(2), 1192-1206.

Menn, L. (1983). Development of articulatory, phonetic, and phonological capabilities. In B. Butterworth (ed.), *Language Production, Vol. 2* (pp. 3-50). Academic Press.

Menn, L., Schmidt, E., & Nicholas, B. (2013). Challenges to theories, charges to a model: The Linked-Attractor model of phonological development. In M.M. Vihman & T. Keren-Portnoy (eds.), *The Emergence of Phonology: Whole-Word Approaches and Cross-Linguistic Evidence*, (pp. 460-502). CUP.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In F. Pellegrino, I. Chitoran, E. Marsico, & C. Coupé (eds.), *Approaches to Phonological Complexity* (pp. 299-328). Mouton.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132-142.

Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech, Language, and Hearing Research*, 36(5), 959-972.

Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America*, 97(1), 520-530.

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), 120-132.

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 and Hearing Research*, 39, 379–389.

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience*, 33(41), 16110-16116.

Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., & Tiede, M. (2018). How do children organize their speech in the first years of life? Insight from ultrasound imaging. *Journal of Speech, Language, and Hearing Research*, 1-14.

Noiray, A., Ménard, L., & Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. *Journal of the Acoustical Society of America*, 133(1), 444-452.

Norman, D. A., & Shallice, T. (1986). Attention to action. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (eds.), *Consciousness and Self-Regulation* (pp. 1-18). Plenum Press.

Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics*, 39(1), 59-67.

Oller, D. K. (2000). *The emergence of the speech capacity*. Psychology Press.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1993). Trading relations between tongue-body raising and lip rounding in production of the vowel/u/: A pilot ''motor equivalence'' study. *Journal of the Acoustical Society of America*, 93(5), 2948-2961.

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica*, 86(2-3), 227-272.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, 77(2), 97-132.

Recasens, D. (1989). Long range coarticulation effects for tongue dorsum contact in VCVCV sequences. *Speech Communication*, 8(4), 293-307.

Redford, M.A. (2015). Unifying speech and language in a developmentally sensitive model of production. *Journal of Phonetics*, 53, 141-152.

Redford, M.A. (2018). Grammatical word production across metrical contexts in school-aged children's and adults' speech. *Journal of Speech, Language, and Hearing Research*, 61, 1339-1354

Redford, M. A., & Miikkulainen, R. (2007). Effects of acquisition rate on emergent structure in phonological development. *Language*, 83(4), 737-769.

Richardson, M. J., Shockley, K., Fajen, B. R., Riley, M. A., & Turvey, M. T. (2009). Ecological psychology: Six principles for an embodied–embedded approach to behavior. In P. Calvo & A. Gomila (eds.), *Handbook of Cognitive Science: An Embodied Approach* (pp. 159-187), Elsevier.

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. Language and cognitive processes, 14(2), 173-200.

Rogers, S. (1978). Self-initiated corrections in the speech of infant-school children*. Journal of Child Language*, 365-371.

Saltzman, E., & Kelso, J. A. (1987). Skilled actions: a task-dynamic approach. *Psychological Review*, 94(1), 84.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1(4), 333-382.

Savariaux, C., Perrier, P., & Orliaguet, J. P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *Journal of the Acoustical Society of America*, 98(5), 2428-2442.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, 82(4), 225.

Schwartz, J. L., Boë, L. J., Vallée, N., & Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of Phonetics*, 25(3), 255-286.

Sharkey, S. G., & Folkins, J. W. (1985). Variability of lip and jaw movements in children and adults: implications for the development of speech motor control. *Journal of Speech, Language, and Hearing Research*, 28, 8-15.

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.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, 66(3), 422-429.

Smith, A., & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research*, 41, 18-30.

Smith, A., Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*, 45, 22–33.

Smith, B. L. (1992). Relationships between duration and temporal variability in children's speech. *Journal of the Acoustical Society of America*, 91, 2165-2174.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P.D. Eimas & J.L. Miller (eds.), *Perspectives on the Study of Speech* (pp. 1-38). Erlbaum.

Stoel-Gammon, C. (1983). Constraints on consonant–vowel sequences in early words. *Journal of Child Language*, 10(2), 455-457.

Stoel-Gammon, C., & Cooper, J. A. (1984). Patterns of early lexical and phonological development. *Journal of Child Language*, 11(2), 247-271.

Stoel-Gammon, C., & Dunn, C. (1985). *Normal and disordered phonology in children*. Pro Ed.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (eds.), *Language Perception and Production* (pp. 67-84). London: Academic Press.

Symons, D. K. (2004). Mental state discourse, theory of mind, and the internalization of self–other understanding. *Developmental Review*, 24(2), 159-188.

Terband, H., Van Brenk, F., & van Doornik-van der Zee, A. (2014). Auditory feedback perturbation in children with developmental speech sound disorders. *Journal of Communication Disorders*, 51, 64-77.

Thompson, A. E., & Hixon, T. J. (1979). Nasal air flow during normal speech production. *Cleft Palate Journal*, 16, 412-420.

Tilsen, S. (2014). Selection and coordination of articulatory gestures in temporally constrained production. *Journal of Phonetics*, 44, 26-46.

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952-981.

Turk, A., & Shattuck-Hufnagel, S. (2014). Timing in talking: what is it used for, and how is it controlled?. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1658): 20130395.

Turvey, M. T. (1990). Coordination. *American Psychologist*, 45(8), 938-953. http://dx.doi.org/10.1037/0003-066X.45.8.938

Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Blackwell.

Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, 108(1), 1-27.

Vihman, M., & Croft, W. (2007). Phonological development: Toward a "radical" templatic phonology. *Linguistics*, 45, 683-725.

Vihman, M. M., Ferguson, C. A., & Elbert, M. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics*, 7(1), 3-40.

Vihman, M.M., Keren-Portnoy, T. (Eds.) (2013). *The emergence of phonology: Whole-approaches and cross-linguistic evidence*. Cambridge University Press.

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language*, 397-445.

Vihman, M. M., & McCune, L. (1994). When is a word a word?. *Journal of Child Language*, 21(3), 517-542.

Vihman, M. M., Nakai, S., DePaolis, R. A., & Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, 50(3), 336-353.

Waterson, N. (1971). Child phonology: A prosodic view. *Journal of Linguistics*, 7(2), 179-211.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, 72(3), 655-684.

Whiteside, S.P., & Irving, C.J. (1998). Speakers' sex differences in voice onset time: a study of isolated word production. *Perceptual and Motor Skills*, 86(2), 651-654.

Wickelgran, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, 76(1), 1-15.

Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control*, 15(1), 118-140.

Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2012). An ultrasound study of lingual coarticulation in/s V/syllables produced by adults and typically developing children. *Journal of the International Phonetic Association*, 42(2), 193-208.