Neural Representations and Mechanisms for the Performance of Simple Speech Sequences

Neural Representations and Mechanisms for the Performance of Simple Speech Sequences

Jason W. Bohland[1], Daniel Bullock[2], and Frank H. Guenther[2,3]

Abstract

■ Speakers plan the phonological content of their utterances before their release as speech motor acts. Using a finite alphabet of learned phonemes and a relatively small number of syllable structures, speakers are able to rapidly plan and produce arbitrary syllable sequences that fall within the rules of their language. The class of computational models of sequence planning and performance termed *competitive queuing* models have followed K. S. Lashley [The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley, 1951] in assuming that inherently parallel neural representations underlie serial action, and this idea is increasingly supported by experimental evidence. In this article, we developed a neural model that extends the existing DIVA model of speech production in two complementary ways. The new model includes paired structure and content subsystems [cf. MacNeilage, P. F. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences, 21,* 499–511, 1998] that provide parallel representations of a forthcoming speech plan as well as mechanisms for interfacing these phonological planning representations with learned sensorimotor programs to enable stepping through multisyllabic speech plans. On the basis of previous reports, the model's components are hypothesized to be localized to specific cortical and subcortical structures, including the left inferior frontal sulcus, the medial premotor cortex, the basal ganglia, and the thalamus. The new model, called gradient order DIVA, thus fills a void in current speech research by providing formal mechanistic hypotheses about both phonological and phonetic processes that are grounded by neuroanatomy and physiology. This framework also generates predictions that can be tested in future neuroimaging and clinical case studies. ■

INTRODUCTION

Here we present a neural model that describes how the brain may represent and produce sequences of simple, learned speech sounds. This model addresses the question of how, using a finite inventory of learned speech motor actions, a speaker can produce arbitrarily many utterances that fall within the phonotactic and linguistic rules of her language. At the phonological level of representation, the model implements two complementary subsystems, corresponding to the structure and content of planned speech utterances within a neurobiologically realistic framework that simulates interacting cortical and subcortical structures. This phonological representation is hypothesized to interface between the higher level conceptual and morphosyntactic language centers and the lower level speech motor control system, which itself implements only a limited set of learned motor programs. In the current formulation, syllable-sized representations are ultimately selected through phonological encoding, and these activate the most appropriate sensorimotor programs, commanding the execution of the planned sound. Construction of the model was guided by previous theoretical work as well as clinical and experimental results, most notably a companion fMRI study (Bohland & Guenther, 2006).

Much theoretical research has focused on the processes involved in language production. One approach has been to delineate abstract stages through which a communicative concept is subjected to linguistic rules and ultimately transformed into a series of muscle activations used for speech production (Garrett, 1975). This approach has led to the development of the influential *Nijmegen model* (Levelt, Roelofs, & Meyer, 1999), which casts the speech system as a set of hierarchical processing stages, each of which transforms an input representation of a certain form (at a certain linguistic "level") to an output representation in a different "lower level" form. The current work addresses the proposed phonological encoding and phonetic encoding stages and interfaces with an existing model, the Directions Into Velocities of Articulators (DIVA) model of speech production (Guenther, Ghosh, & Tourville, 2006; Guenther, Hampson, & Johnson, 1998; Guenther, 1995), which describes the stage of articulation. The present model, called gradient order DIVA (GODIVA), describes the ongoing parallel representation of a speech plan as it cascades through the stages of production. Although we do not address higher level linguistic processes, the proposed architecture is designed to be extensible to address these in future work.

[1]Cold Spring Harbor Laboratory, [2]Boston University, [3]Harvard University-Massachusetts Institute of Technology

A limitation of the DIVA model, which accounts for how sensorimotor programs for speech sounds[1] can be learned and executed, is that it contains no explicit planning representations outside the activation of a single speech sound's stored representation, nor does it address the related issue of appropriately releasing planned speech sounds to the motor apparatus (referred to here as *initiation*). GODIVA adds abstract phonological representations for planned speech sounds and their serial order and simulates various aspects of serial speech planning and production. Furthermore, the model follows recent instantiations of the DIVA model (Guenther, 2006; Guenther et al., 2006) by proposing specific neuroanatomical substrates for its components.

Models of Serial Behavior

At the heart of any system for planning speech must be a mechanism for representing items to be spoken in the correct order. A number of concrete theoretical proposals have emerged to model serial behaviors (see Bullock, 2004; Rhodes, Bullock, Verwey, Averbeck, & Page, 2004; Houghton & Hartley, 1995). *Associative chaining* theories postulate that serial order is stored through learned connections between cells representing successive sequence elements and that each node's activation in turn causes activation of the subsequent node, enabling sequential readout. In their simplest form, however, these models cannot learn to unambiguously readout different sequences defined over the same component items. Wickelgren's (1969) speech production model addressed this problem by introducing many context-sensitive allophones (e.g., /$_{k}$æ$_{t}$/ for phoneme /æ/ when preceded by /k/ and followed by /t/) as nodes through which a serial chain could proceed. However, such a model does not capture the relationship between the same phonemes in different contexts and suffers a combinatorial explosion in the requisite number of nodes when allowing sequences that can overlap by several consecutive phonemes. More recent neural network models (Beiser & Houk, 1998; Elman, 1990; Lewandowsky & Murdock, 1989; Jordan, 1986) proposed revisions that rely on a series of sequence-specific internal states that must be learned to allow sequence recall. Although such networks overcome the central problem above, they provide no basis for novel sequence performance and have difficulty simulating cognitive error data due to the fact that if a "wrong link" is followed in error, there is no means to recover and correctly produce the remaining items (Henson, Norris, Page, & Baddeley, 1996).

Alternatively, strict *positional models* represent serial order by the use of memory "slots" that signify a specific ordinal position in a sequence. Sequence performance then simply involves stepping through the series of slots (always in the same order) and executing each associated component item. Unfortunately, there is no obvious neural mechanism to allow the insertion of an arbitrary memory (or memory pointer) into a particular "slot." Such models either require the ability to "label" a positional node with a particular representation or component item or require a set of all possible representations to be available at all serial positions, which is often infeasible. Recent models within this lineage hypothesize order to be accounted for by some contextual signal such as the state of an oscillatory circuit or other time-varying function (Brown, Preece, & Hulme, 2000; Vousden, Brown, & Harley, 2000; Burgess & Hitch, 1999; Henson, 1998). Recall then involves "replaying" this contextual signal that, in turn, preferentially activates the items associated with the current state of the signal. Such models require the ability to form associations between context signal and component item through one-shot learning to allow for novel sequence performance.

Lashley (1951) deduced that associative chaining models could not sufficiently describe the performance of sequences including those comprising speech and language and that serial behavior might instead be performed based on an underlying parallel planning representation. Grossberg (1978a, 1978b) developed a computational theory of short-term sequence memory in which items and their serial order are stored via a primacy gradient using the simultaneous parallel activation of a set of nodes, where relative activation levels of the content-addressable nodes code their relative order in the sequence. This parallel working memory plan, which can be characterized as a *spatial pattern* in a neuronal map, can be converted to serial performance through an iterative competitive choice process in which (i) the item with the highest activation is chosen for performance, (ii) the chosen item's activation is then suppressed, and (iii) the process is repeated until the sequence reaches completion. These types of constructions have been collectively termed *competitive queuing* (CQ) models (Bullock & Rhodes, 2003; Houghton, 1990). Figure 1 illustrates the basic CQ architecture. Recently, this class of CQ models has received substantial support from direct neurophysiological recordings in monkeys (Averbeck, Chafee, Crowe, & Georgopoulos, 2003) and from chronometric analyses of seriation errors (Farrell & Lewandowsky, 2004). A CQ-compatible architecture forms the basis of various modules used in GODIVA.

Linguistic Representations

Although the majority of serial order theories have addressed STM without explicit treatment of linguistic units, some models have been introduced to account for the processing of such representations for word production. Such models generally follow the theoretical work of Levelt (1989), Garrett (1975), and others. Existing linguistic models typically seek to address either patterns observed in speech error data or chronometric data concerning speaker RTs under certain experimental manipulations. Error data have highlighted the importance of

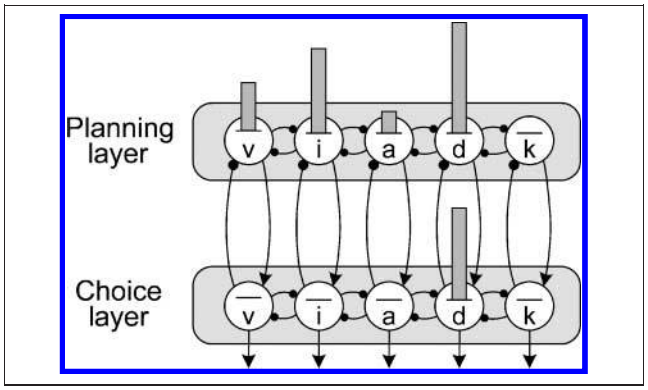

Figure 1. CQ model architecture for the representation and performance of the letter sequence "diva." The serial position of each letter is encoded by its strength of representation (height of bar) in the planning layer (top). The choice layer (bottom) realizes a competitive (winner take all) process that allows only the strongest input to remain active, in this case "d." Upon selection of "d," its representation in the planning layer would be suppressed, leaving "i" as the most active node. This entire process iterates through time, enabling performance of the entire letter sequence.

considering speech planning and production at multiple levels of organization, specifically including *word*, *syllable*, *phoneme*, and *feature* as possible representational units. These are conceptually hierarchical, with higher units comprising one or more lower.

MacNeilage (1998) proposed a close link between the syllable unit and the motor act of open-closed jaw alternation. In this proposal, a behaviorally relevant motor frame roughly demarcates syllable boundaries. Such a motor frame might be useful in delineating and learning individual "chunks" that contribute to a finite library of "performance units." However, the phonological syllable does not only circumscribe a set of phonemes but also appears useful to provide a schema describing the abstract "rules" governing which phonemes can occur in each serial position within the syllable (e.g., Fudge, 1969). To this end, the syllable can be broken into, at least, an *onset* and a *rime*, the latter of which contains subelements *nucleus* and *coda*. Syllables contain phonemes, which are categorical and exhibit a many-to-one relationship between acoustic signals and perceptual labels, with all realizations of a particular phoneme being cognitively equivalent. The GODIVA model assumes the reality of phonemic categories, but its framework is amenable to alternative discrete categorizations. Although our proposal is for a segmental model that lacks explicit representation of either articulatory or acoustic features, we allowed room for the implicit representation of featural similarity, which has been shown to have a significant influence on speech error patterns (MacKay, 1970) and production latencies (Rogers & Storkel, 1998).

Nearly all previous theories of phonological encoding or serial speech planning have proposed some form of factorization of the structure and the phonological content of an utterance, often in the form of syllable- or word-sized structural frames and phoneme-sized content.[2] Such a division is motivated, in part, by the pattern of errors observed in spontaneously occurring slips of the tongue. MacKay (1970), in his study of spoonerisms, or phoneme exchange errors (e.g., saying "*h*eft *l*emisphere" instead of the intended "*l*eft *h*emisphere"), noted the prominence of the syllable position constraint, in which exchanges are greatly biased to occur between phonemes occupying the same positional "slot" in different planned syllables. This constraint appears to be the strongest pattern observed in speech errors. Shattuck-Hufnagel (1979), for example, found that 207 of 211 exchange errors involved transpositions to and from similar syllabic positions. More recently, Vousden et al. (2000) found that approximately 90% of consonant movement errors followed this constraint. Treiman and Danis (1988) also noted that during nonword repetition, most errors are phonemic substitutions that preserve syllable structure. Such exchanges also follow a transposition distance constraint (MacKay, 1970), in that phonemes are more likely to exchange between neighboring rather than distant syllables. Beyond speech error data, priming studies have demonstrated effects in speech production based purely on CV structure (while controlling for content) at the syllable and word level (Meijer, 1996; Sevald, Dell, & Cole, 1995). Together, such data argue for factorizing abstract frames from phonemic content, an organizational principle that is exploited in the GODIVA model at the level of syllables.

Neurobiological Foundations

A shortcoming of previous theoretical psycholinguistic proposals has been their general failure to account for how linguistic behavior can emerge from specific neural structures (Nadeau, 2001). GODIVA makes use of information-processing constructs similar to those proposed elsewhere but embeds these in a biologically realistic architecture with specific hypotheses about cortical and subcortical substrates. These hypotheses are based on integrating the sparse available data from clinical and functional imaging studies and from inferences drawn from nonspeech sequencing tasks in other species, under the assumption that similar mechanisms should underlie linguistic processes and other complex serial behaviors. We emphasize that the use of artificial neural network architectures alone does not establish biological plausibility; rather, it is essential to explicitly consider what is known about the functional architecture of specific systems. In so doing, the GODIVA framework offers the ability to treat additional data sets that cannot be directly addressed by previous models. In particular, region-level effects in functional imaging and lesion studies can be related to specific model components. Although a review of the possible roles of various brain areas in speech planning and production is beyond the scope of this article, here we have elaborated on certain anatomical and physiological considerations involved in the new model development.

Prefrontal Cortex

The left prefrontal cortex, specifically in and surrounding the ventral inferior frontal sulcus (IFS), showed increased activity in a memory-guided speaking task when the serial complexity of the utterance was increased (Bohland & Guenther, 2006). This is consistent with a hypothesis that this region contains a representation of a forthcoming speech plan. A similar region in left dorsal inferior frontal gyrus has been suggested to be involved in sequencing discrete units, including phonemes (Gelfand & Bookheimer, 2003), and in phonological encoding tasks (Papoutsi et al., 2009; Chein & Fiez, 2001; Burton, Small, & Blumstein, 2000; Poldrack et al., 1999).

In a nonspeech sequencing study in macaque monkeys, Averbeck et al. (2003) and Averbeck, Chafee, Crowe, and Georgopoulos (2002) recorded single-cell activity from the right hemisphere prefrontal cortex during a sequential shape copying task. These recording sites were within approximately 5 mm of the ventral portion of the arcuate sulcus, which has been proposed to be homologous to the IFS in humans (Rizzolatti & Arbib, 1998). Activities of cell ensembles coding for specific segments in the shape were recorded during a delay period before the first stroke and throughout the performance of the stroke sequence. In the delay period, a cotemporal representation of all of the forthcoming segments was found, and the relative activity in each neuron ensemble predicted the relative priority (i.e., order) in which the segments were performed. After execution of each segment, the activation of its ensemble representation was strongly reduced, and the other ensembles' activations increased. Such item-specific primacy gradients were observed even after sequences were highly practiced, to the point of demonstrating "coarticulation." Further analyses have shown a partial normalization of total activation distributed among the representation for planned items (Averbeck et al., 2002; Cisek & Kalaska, 2002). This agrees with predictions based on planning layer dynamics in CQ models (Grossberg, 1978a, 1978b). Because total activity growth is a decelerating and saturating function of the number of planned items in a sequence, relative activation levels become more difficult to distinguish, and more readily corrupted by noise (Page & Norris, 1998), in longer sequences. These properties help explain why there is a low limit (see Cowan, 2000) to the number of items that can be planned in working memory and recalled in the correct order. Taken together, these electrophysiological findings provide compelling evidence for CQ-like dynamics in the prefrontal cortex, in a location near the possible homologue for human IFS. The GODIVA model posits that forthcoming *phonemes* are planned in parallel in or around the left hemisphere IFS, consistent with evidence from fMRI.

Medial Frontal Cortex

The medial frontal cortex, consisting of the more posterior SMA and the more anterior pre-SMA (Picard & Strick, 1996; Matsuzaka, Aizawa, & Tanji, 1992), has been implicated in sequencing and speech production tasks. Lesions to the medial wall cause speech problems (Pai, 1999; Ziegler, Kilian, & Deger, 1997; Jonas, 1981, 1987), usually resulting in reduced propositional speech with nonpropositional speech (e.g., counting, repeating words) largely intact. Other problems include involuntary vocalizations, echolalia, lack of prosody, stuttering-like output, and variable speech rate. Both Ziegler et al. (1997) and Jonas (1987) have suggested that the SMA plays a role in sequencing and self-initiation of speech sounds but that it is unlikely that these areas code for specific speech sounds.

In monkey studies of complex nonspeech tasks, sequence-selective cells have been identified in both SMA and pre-SMA (Shima & Tanji, 2000) that fire during a delay period before the performance of a specific sequence of movements in a particular order. This study also identified interval-selective cells, mostly in the SMA, that fired in the time between two particular component movements. Rank-order-selective cells have also been found, primarily in the pre-SMA, whose activity increased before the *n*th movement in the sequence, regardless of the particular movement (see also Clower & Alexander, 1998). Finally, Shima and Tanji (2000) found that only 6% of pre-SMA cells were selective to particular movements as opposed to 61% of cells in the SMA, indicating a cognitive-motor functional division.

Bohland and Guenther (2006) described differences between anterior and posterior medial areas, with the pre-SMA increasing activity for more complex syllable frames (e.g., CCCV[3] vs. CV) and with the SMA increasing activity during overt speaking trials relative to preparation-only trials. These findings suggest differential roles for the pre-SMA and SMA in speech, which has also been demonstrated elsewhere (Alario, Chainay, Lehericy, & Cohen, 2006). In the present proposal, we hypothesized that the pre-SMA encodes structural "frames" at an abstract level (cf. MacNeilage, 1998), whereas the SMA serves to initiate or release planned speech acts. We view these specific hypotheses as only tentative, although consistent with available evidence, and suggest that further experiments are needed to determine both the localization and the level of representation of any such abstract representations.

Basal Ganglia Loops

Interactions between the cortex and the BG are organized into multiple loop circuits (Middleton & Strick, 2000; Alexander & Crutcher, 1990; Alexander, DeLong, & Strick, 1986). The BG are involved in sequencing motor acts (e.g., Harrington & Haaland, 1998), and abnormalities in these regions variously impact speech production (Murdoch, 2001; Kent, 2000), with some patients having particular difficulty fluently progressing through a sequence of articulatory targets (Ho, Bradshaw, Cunnington, Phillips, & Iansek, 1998; Pickett, Kuniholm, Protopapas, Friedman, & Lieberman, 1998). Damage to the caudate is also associated with perseverations and paraphasias (Kreisler et al., 2000), and both structural and functional abnormalities have been observed in the caudate in patients with inherited verbal dyspraxia characterized by particular difficulties with complex speech sequences (Watkins et al., 2002; Vargha-Khadem et al., 1998). The architecture within BG loops is intricate (e.g., Parent & Hazrati, 1995), but here we adopt a simplified view to limit the model's overall complexity. The striatum, comprising the caudate nucleus and the putamen, receives inputs from different cortical regions. The striatum is dominated by GABA-ergic medium spiny projection neurons (Kemp & Powell, 1971), which are hyperpolarized and normally quiescent, requiring convergent cortical input to become active (Wilson, 1995). Also found in the striatum, but less prevalent (only 2–3% of striatal cells in rats, but perhaps as many as 23% in humans; Graveland, 1985), are heterogeneous interneurons, many of which exhibit resting firing rates and receive cortical, thalamic, and dopaminergic input (Tepper, Koos, & Wilson, 2004; Kawaguchi, 1993). Some of these cells are GABA-ergic and suitable to provide feedforward surround inhibition (Plenz & Kitai, 1998; Jaeger, Kita, & Wilson, 1994). Medium spiny neurons send inhibitory projections to cells in the pallidum including the globus pallidus internal (GPi) segment, which are tonically active and inhibitory to cells in the thalamus, which in turn project back to cortex (e.g., Deniau & Chevalier, 1985). Hikosaka and Wurtz (1989) found that BG output neurons enable voluntary saccades by means of a pause in the normally tonic inhibition delivered to spatially specific targets in the superior colliculus and motor thalamus.

Mink (1996) and Mink and Thach (1993) outlined a conceptual model of BG function, suggesting that BG loops are used to selectively enable a motor program for output among competing alternatives. Such action selection models (also Gurney, Prescott, & Redgrave, 2001a, 2001b; Kropotov & Etlinger, 1999; Redgrave, Prescott, & Gurney, 1999), which suggest that the BG do not generate movements but rather select and enable them, are relevant for sequence performance. Mink, for instance, suggested that each component movement must be selected whereas other potential movements (e.g., those to be performed later in the sequence) must be inhibited. In response to such early models' omission of distinct planning and executive layers of cortex, Brown, Bullock, and Grossberg (2004) described a computational neural model (TELOS) for the control of saccades, in which cortico-BG loops make voluntary behavior highly context sensitive by acting as a "large set of programmable gates" that control planning layers' abilities to fully activate executive (output) layers, which drive action. In TELOS, the striatal stage of the gating system receives cortical inputs from cue-sensitive planning cells in various gateable cortical zones. By opening gates when it senses coherence among multiple cortical representations, the striatum promotes the most apt among competing cortical plans while also deferring execution of a plan until conditions favor its achieving success. The TELOS theory also proposed that BG loops divide the labor with higher resolution, laminar target structures in the frontal cortex (and superior colliculus). The BG output stages lack sufficient cell numbers to provide high-resolution action specifications; however, they can enable one frontal zone much more (than another), and within the favored zone, competition between representations ensures the required specificity of output. Here we simplified the BG circuit model but make essential use of the ideas of action selection by gating circuits that enable plan execution only when the striatum detects that multiple criteria have been satisfied.

Interface with Speech Motor Control

DIVA is a neural network model of speech motor control and acquisition that offers unified explanations for a large number of speech phenomena including motor equivalence, contextual variability, speaking rate effects, and coarticulation (Guenther et al., 1998; Guenther, 1995). DIVA posits the existence of a speech sound map (SSM) module in the left ventral premotor cortex and/or posterior inferior frontal gyrus pars opercularis that contains cell groups coding for well-learned speech sounds. SSM representations are functionally similar to a mental syllabary (Levelt & Wheeldon, 1994; Crompton, 1982), suggested by Levelt et al. (1999) to consist of a "repository of gestural scores for the frequently used syllables of the language" (p. 5). Using alternative terminology, SSM representations can be thought of as sensorimotor chunks or programs, learned higher order representations of frequently specified spatio-temporal motor patterns. As noted above, the GODIVA model follows these proposals as well as MacNeilage (1998) in placing the syllable as a key unit for speech motor output, but our general approach is amenable to output units of other sizes that exhibit repeating structural patterns (e.g., words or morphemes).

A syllable frequency effect, in which frequently encountered syllables are produced with a shorter latency than uncommon (but well-formed) syllables, has been reported by several researchers (Cholin, Levelt, & Schiller, 2006; Laganaro & Alario, 2006; Alario et al., 2004; Carreiras & Perea, 2004; Levelt & Wheeldon, 1994). Although Levelt and Wheeldon (1994) argued that the syllable frequency effect implied the use of stored syllable motor programs, it proved difficult to rule out that the effect could be due to phonological processing. Laganaro and Alario (2006) used a delayed naming task with and without an interfering articulatory suppression task to provide strong evidence that this effect is localized to the phonetic rather than phonological encoding stage, consistent with the role of the SSM module in DIVA.

In the extended GODIVA model, the SSM forms the interface between the phonological encoding system and the phonetic/articulatory system. Sensorimotor programs for frequently encountered syllables can be selected from the SSM in full, whereas infrequent syllables must be performed from smaller (e.g., phoneme-sized) targets. We note that, although this points to two routes for syllable articulation, in neither case does GODIVA posit a bypass of the segmental specification of forthcoming sounds in a phonological queue. Some dual-path models (Varley & Whiteside, 2001) predict that whereas novel or rare sequences are obliged to use an "indirect" process that requires working memory for sequence assembly, high-frequency sequence production instead uses a "direct" route that bypasses assembly in working memory and instead produces sequences as "articulatory gestalts." Contrary to such dual-path models, Rogers and Spencer (2001) argued that sequence assembly in a working memory buffer is obligatory even for production of high-frequency words. Besides exchange error data, a key basis for their argument was the finding that "speech onset latencies are consistently … longer when the onsets of successive words are phonologically similar" (p. 71), even when the successive words are high frequency. This transient inhibitory effect is an expected "aftereffect" of a fundamental CQ property: active suppression of a chosen item's representation. Consistently, although both the Nijmegen model and the GODIVA model have provisions for modeling differences in the production of high- versus low-frequency sequences, neither make the problematic assumption that automatization eventually entails bypassing assembly. That assumption is incompatible with the exquisite control of vocal performance that speakers/singers retain for even the highest frequency syllables.

A detailed description of the DIVA model is available elsewhere (Guenther, 2006; Guenther et al., 2006). Below, we specified a computational neural model that extends DIVA to address the planning and initiation of sequences of connected speech and the brain regions likely to be involved in those processes.

MODEL DESCRIPTION

Here we provided a high-level overview of the GODIVA model, followed by the formal specification. A comprehensive illustration of the theoretical model is shown in Figure 2. In this figure, boxes with rounded edges refer to components that have been addressed in the DIVA model (Guenther, 2006; Guenther et al., 2006). Here we further conceptualized some brain areas attributed to the phonetic/articulatory levels of processing but focus on higher level processing, providing implementation details only for boxes drawn with dotted borders. The model contains dual CQ-like representations of the forthcoming utterance at the phonological level, hypothesized to exist in the pre-SMA and IFS regions, with selection mediated by a BG "planning loop." Selected phonological codes interface with an elaborated SSM to select best matching sensorimotor programs for execution. Because available data are sparse, the GODIVA model should only be considered as a starting point: one possible interpretation of existing data. The key advancements that we hope to achieve in this proposal are (i) to tie information processing accounts of phonological processes in speech production to hypothetical neural substrates and (ii) to bring together models of articulation and theories of speech planning and preparation. Explicit calls for bridging this latter gap have been made in the study of communication disorders (McNeil, Pratt, & Fossett, 2004; Ziegler, 2002).

The "input" to the GODIVA model during ordinary speech production is assumed to arrive from higher level lexical/semantic and/or syntactic processing areas, possibly including the inferior or ventrolateral prefrontal regions of the cortex, or from posterior regions in repetition tasks. In most cases, these inputs are thought to code lexical items (words) or short phrases and arrive sequentially as incremental processing is completed by the higher level modules. These inputs initiate the activation of two parallel and complementary representations for a forthcoming utterance: a *phonological content representation* hypothesized to exist in the left hemisphere IFS and a *structural frame representation* hypothesized to exist in the pre-SMA. Both representations constitute planning spaces or forms of working memory, where representative neurons or populations of neurons maintain a cortical code for the potential phonemes (in the IFS) or abstract syllable frames (in the pre-SMA) that define the utterance. In GODIVA, these representations simultaneously, cotemporally code for multiple forthcoming phonemes and syllable frames by use of primacy gradients, in which relative activation levels code for the serial order in which the items are to be produced. These gradients over plan cells are maintained for a short duration through recurrent dynamics and can robustly handle new inputs as they arrive without disruption of ongoing performance, up to a certain item capacity determined by the signal-to-noise ratio of the representation. Both the IFS and the pre-SMA plan layers thus take the form of "item and order memories" (Grossberg, 1978a, 1978b) or, equivalently, planning layers in CQ circuits (Bullock & Rhodes, 2003).

The model's production process begins when the most active frame in the pre-SMA planning layer is selected within a second set of pre-SMA cells, the *choice layer*. The division of cortical representations into plan and choice layers within a columnar architecture is repeated throughout the model (see Figure 2). Activation of a pre-SMA choice cell initiates the firing of a chain of additional pre-SMA cells, each corresponding to an abstract position (but not a specific phoneme) in the forthcoming syllable. These pre-SMA cells give input to a BG-mediated planning loop, which serves as an input gate to the choice layer in the IFS region, effectively controlling access to

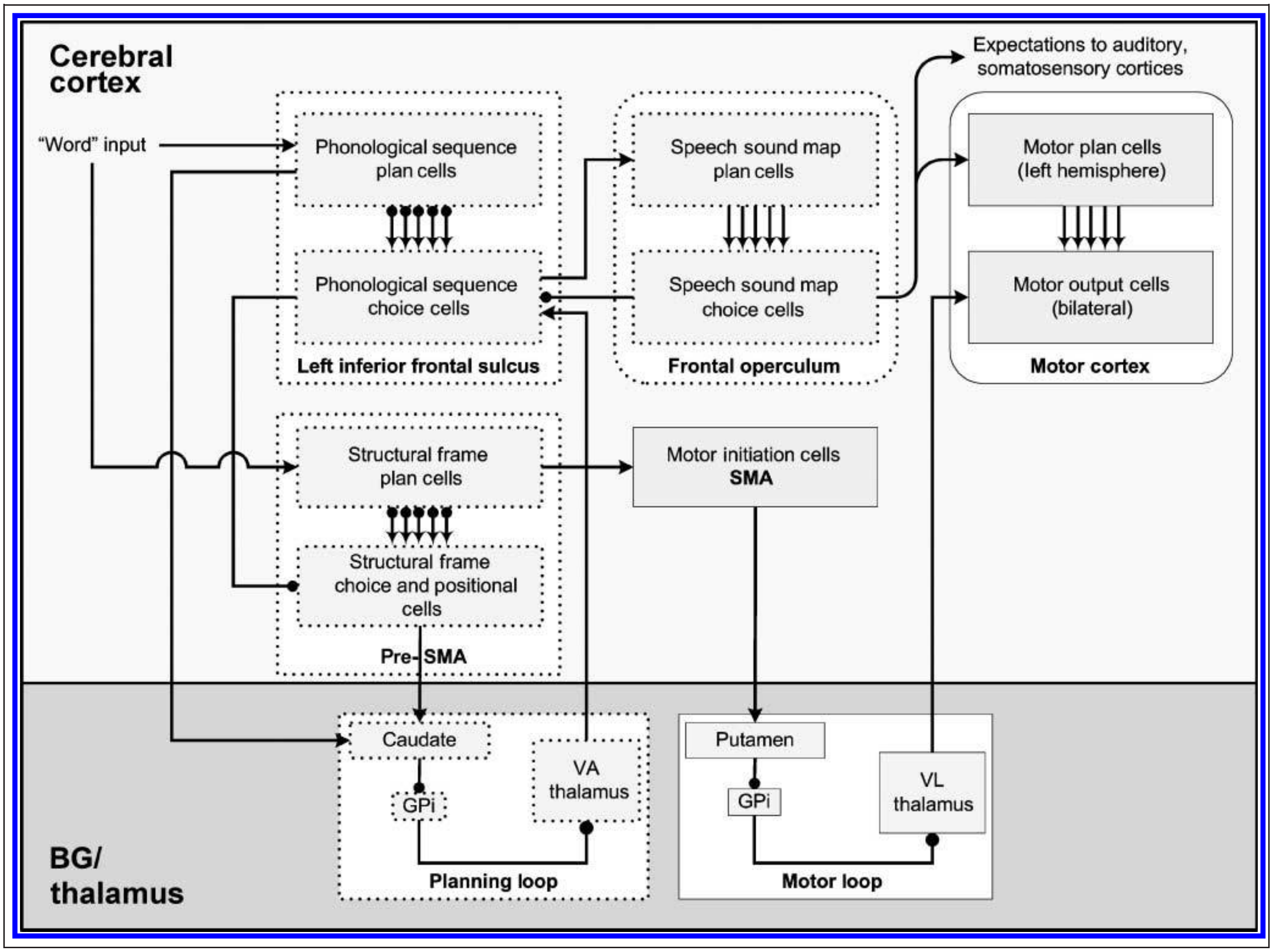
Cerebral cortex
"Word" input
Phonological sequence plan cells
Phonological sequence choice cells
Left inferior frontal sulcus
Speech sound map plan cells
Speech sound map choice cells
Frontal operculum
Expectations to auditory, somatosensory cortices
Motor plan cells (left hemisphere)
Motor output cells (bilateral)
Motor cortex
Structural frame plan cells
Structural frame choice and positional cells
Pre-SMA
Motor initiation cells SMA
Caudate
GPi
VA thalamus
Planning loop
Putamen
GPi
VL thalamus
Motor loop
BG/ thalamus

Figure 2. Schematic of the overall proposed architecture of the GODIVA model, indicating their hypothesized cortical and subcortical correlates. Boxes with rounded edges have received treatment previously in the DIVA model but may be further elucidated here. Boxes with dotted borders are given explicit computational treatment here, whereas others are outlined conceptually. Lines with arrows represent excitatory pathways, and lines with filled circles represent inhibitory pathways. Lines with both arrowheads indicate that connectivity between these modules features top–down excitatory connections and bottom–up inhibitory connections. The inhibitory pathways shown in the cortical portion of the model are feedback pathways that suppress planning representations after their corresponding action has been taken.

the output phonological representation, which drives activation in the planning layer of the SSM, a component of the DIVA model storing phonetic representations, which is further elaborated here. This planning loop specifically enables topographic zones in the IFS choice layer that correspond to appropriate syllable positions for the immediately forthcoming syllable. Strong competition among IFS choice cells in each positional zone results in a single "winning" representation within each active positional zone. As in standard CQ-based models, any IFS and pre-SMA choice cells that become active ("win") selectively suppress the planning representations to which they correspond.

IFS choice cells form cortico-cortical synapses with cell populations in the SSM that, following the hypotheses of the DIVA model, enable the "readout" of motor programs as well as auditory and somatosensory expectations for simple learned speech sounds. The SSM is hypothesized to occupy the left posterior inferior frontal gyrus (opercular region) and adjoining left ventral premotor cortex (Guenther, 2006); we will use the term *frontal operculum* as shorthand for this region. Learning of the IFS → SSM synapses is suggested to occur somewhat late in development, after a child has developed well-defined phonetic/phonological perceptual categories for his or her language. These tuned synapses (which are defined algorithmically in the model) allow the set of winning choice cells in the IFS choice layer to activate a set of potential "matching" sensorimotor programs represented by SSM plan cells, with better matching programs receiving higher activations. Because one IFS choice cell is active for each position in the forthcoming syllable, this projection transforms a phonological syllable into a speech motor program.

SSM plan cells give input to SSM choice cells, which provide output to hypothesized lower level motor units. Competition via recurrent inhibition among SSM choice cells allows a single sensorimotor program to be chosen for output to the motor apparatus. We postulate an additional BG loop (motor loop in Figure 2) that handles the appropriate release of planned speech sounds to the execution system. The chosen SSM output cell is hypothesized to activate motor plan cells primarily in the left-hemisphere motor cortex that, together with inputs from the SMA, bid for motor initiation. A new motor program will be initiated only upon completion of the previous program. The uncoupling of the selection of motor programs from the timing of initiation allows the system to proceed with selection before the completion of the previous chosen program. This provides a simple mechanism to explain differences between preparation and production and between covert and overt speech.

Model Specification

The GODIVA model is defined as a system of differential equations describing the activity changes of simulated neuron populations through time. Equations were numerically integrated in MATLAB using a Runge–Kutta method for noise-free simulations and the Euler method when noise was added. Table 1 provides a legend for the symbols used in the following equations to define cell groups. To reduce complexity, cortico-cortical inhibitory projections, which likely involve a set of intervening interneurons between two sets of excitatory neurons, are modeled as a single inhibitory synapse from a cell that can also give excitatory projections. Note that the present model is "hand wired." That is, synaptic weights that are assumed to be tuned through learning are set algorithmically within the range of values that learning must achieve for proper operation. Possible methods by which these weights can be learned are suggested in the Discussion section. In this version of the model, we have not fully explored the parameter space to provide particularly optimal settings but reported simulations use the same parameter set (excepting noise).

Table 1. Legend of Symbols Used to Refer to Cell Populations in the GODIVA Model Specification

Cell Group	*Symbol*
External input to IFS	u^p
External input to pre-SMA	u^f
IFS phonological content plan cells	p
IFS phonological content choice cells	q
Pre-SMA frame plan cells	f
Pre-SMA frame choice cells	g
Pre-SMA positional chain cells	h
Planning loop striatal projection cells	b
Planning loop striatal interneurons	$\underline{b}$
Planning loop GPi cells	c
Planning loop anterior thalamic cells	d

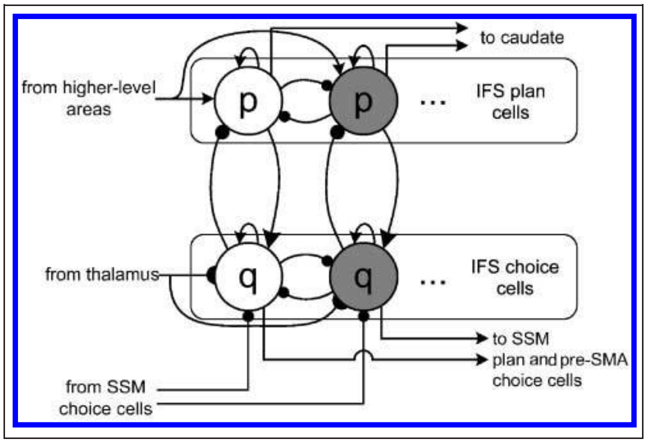

Figure 3. Schematic illustration of the structure of the GODIVA model's IFS phonological content representation. The region is hypothesized to consist of a layer of plan cells (p; top) and a layer of choice cells (q; bottom), arranged into columns, each of which codes for a planned phoneme in a given syllable position. The plan cells are loaded in parallel from other cortical or cerebellar regions. Choice cells, whose input from plan cells is gated by a syllable position-specific signal from the anterior thalamus, undergo a winner-take-all process within each gated zone. The activation of a choice cell suppresses its corresponding plan cell. This process results in the activation of a phonological syllable in the IFS choice field that can activate potentially matching syllable motor programs in the SSM. Choice cell activations can be suppressed upon the selection of a specific SSM motor program.

Phonological Content Representation in IFS

The IFS representation consists of two layers, one containing plan cells and one containing a corresponding set of choice cells. A plan cell and a corresponding choice cell represent a simplified cortical column. Figure 3 illustrates two such IFS columns from a single positional zone as well as their theoretical inputs and outputs. The idealized IFS columns are hypothesized to be tuned to a particular phoneme and to a particular abstract syllable position. The IFS map can thus be thought of as a two-dimensional grid, where each row corresponds to a particular phoneme and each column to a particular syllable position (see Figure 4). Seven syllable positions are included in the model. These correspond to a generic syllable template, such as that introduced by Fudge (1969) and used in the model of verbal STM introduced by Hartley and Houghton (1996). The vast majority of English syllables can be represented in this template by assigning particular phonemes to particular template slots.[4] In GODIVA, the middle (fourth) position is always used to represent the syllable nucleus (vowel), preceding consonants are represented in preceding positional zones, and succeeding consonants in succeeding positional zones.[5] Within a particular positional zone (corresponding to the long axis in Figure 4), an activity gradient across plan cells defines the serial order of the phonemic elements. For example, Figure 4 schematizes the representation of the planned utterance "go.di.və" ("go diva") in the IFS phonological planning layer. Competitive interactions in the IFS map model are restricted to within-position interactions; in essence, therefore, this representation can be thought of as having multiple competitive queues, one for each syllable position. The model includes representations for 53 phonemes (30 consonants and 23 vowels) derived from the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995).

The cells in the IFS form an efficient categorical basis set for representing arbitrary speech sequences from a speaker's language. This is an important principle because it enables the representation and ultimately production of both common (hence well-learned) utterances and novel phonological "words" for which there are no (complete) stored motor associations. These novel phonological words can be effectively "spelled out" during production using motor programs for the individual phonemes rather than a single motor program for the entire phonological word. This representation also allows the speaker to simultaneously plan multiple forthcoming syllables in this categorical space, a faculty that is crucial to rapid, fluent performance. It is important to note that, as depicted, the representation fails to handle repeated elements in a speech plan (e.g., "ta.ka"). If only one cell were available to code /a/ in the nucleus position, it would be impossible to simultaneously represent the order of two occurrences of that phoneme. Although not shown in Figure 4, we therefore assumed the existence of multiple "copies" of each cell in the 53 × 7 representation. This expansion requires some additional machinery to handle loading and response suppression that is discussed further below. For simplicity, the equations below make reference to only one copy of each representational cell. The activity of cell p_{ij}, representing phoneme i at syllable position j in the two-dimensional IFS phonological planning layer matrix $\mathbf{p}$, is governed by

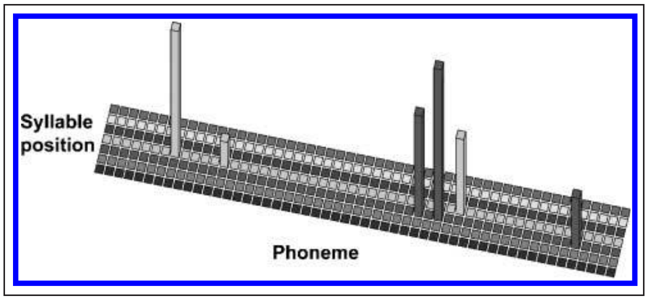

Figure 4. Illustration of the layout of cells in the IFS phonological content representation. Both plan and choice layers in the region use the same representation; shown here is the plan layer, which has dynamics that allow multiple parallel items to be cotemporally active. The long axis in the IFS map corresponds to specific phonemes, and the short axis corresponds to abstract serial positions in a generic syllable template. Cells compete with one another through lateral inhibition along the long axis. This map illustrates an idealized plan that corresponds to the syllable sequence "go.di.və." The height of the vertical bar at a particular entry in the map corresponds to a cell's activation level. Note that entries in the schematic of the same color indicate these cells code for the same syllable position; in this representation, there are three active cells each in Syllable Positions 3 and 4 in the template, corresponding to three [CV] syllables.

$$\dot{\mathrm{p}}_{ij} = -A_{\mathrm{p}}\mathrm{p}_{ij} + (B_{\mathrm{p}} - \mathrm{p}_{ij})\left(\alpha u_{ij}^{\mathrm{p}} + [\mathrm{p}_{ij} - \theta_{\mathrm{p}}]^{+}\right) - \mathrm{p}_{ij}\left(\sum_{k \neq i}\mathbf{W}_{ik}\mathrm{p}_{kj} + 10y\left([q_{ij} - \theta_{q}]^{+}\right)\right) + N(0, \sigma_{\mathrm{p}}) \tag{1}$$

Here the first term yields a passive decay such that, in the absence of external inputs, activity will return to resting potential (identically zero for all cells) at a rate controlled by A_{p}. The second term models excitatory input to the cell, which drives activity in the positive direction. The multiplicative term $(B_{\mathrm{p}} - \mathrm{p}_{ij})$ enforces an upper bound B_{p} to cell activity. Such multiplicative or shunting terms (Grossberg, 1973) are motivated by empirically observed cell membrane properties (e.g., Hodgkin & Huxley, 1952). The third term models the inhibitory inputs to the cell, which drive activity in the negative direction, with lower bound of zero. Many of the equations that follow are written in similar form. The final term, unique to Equation 1, models an additive zero-mean Gaussian noise source, with values independent across time. The model is typically run without noise (e.g., $\sigma_{\mathrm{p}} = 0$), but its inclusion within the IFS can be used to produce stochastic phonological errors.

In Equation 1, there are two primary sources of excitatory input. First, u_{ij}^{p} corresponds to a "word"[6] representation that gives parallel excitation to the IFS plan layer. This input is assumed to arrive from one or more of three brain areas not explicitly treated in the model:

1. a higher level linguistic processing area involved in the morphosyntactic processing of an internally generated communicative concept, likely also in left prefrontal cortex;
2. a phonological processing region in the parietal cortex that can load the modeled phonological output system when the task is, for instance, reading or repetition; or
3. the inferior right-hemisphere cerebellum, which is hypothesized to assist in "fast-loading" of phonological content into this buffer (Rhodes & Bullock, 2002).

This transient input instantiates a gradient across IFS plan units that represents the ordered set of phonemes in the input "word." The input is multiplicatively gated by a term α that can be used to ensure that the activity of cells receiving new inputs corresponding to words to be spoken later does not exceed the activity level of cells representing sounds to be spoken sooner, thus maintaining the correct order of planned speech sounds (e.g., Bradski, Carpenter, & Grossberg, 1994). The second excitatory input to cell p_{ij} is from itself. The constant θ_p is a noise threshold set to some small value and $[\ ]^+$ indicates half-wave rectification, a function that returns the value of its argument (e.g., $p_{ij} - \theta_p$) if positive, otherwise zero. Recurrent self-excitation allows this layer to maintain a loaded plan over a short duration even in the absence of external inputs.

Cell p_{ij} is inhibited by other plan cells coding different phonemes at the same syllable position. The inhibitory inputs are weighted by entries in the adjacency matrix $\mathbf{W}$. In the simplest case (used in simulations here), entry W_{ik} is 1 for $i \neq k$ and 0 for $i = k$. This matrix can be modified to change the strength of phoneme–phoneme interactions, allowing for a partial explanation of phonemic similarity effects (see Discussion). Cell p_{ij} also receives strong inhibition from cell q_{ij}, its corresponding cell (in the same column) in the IFS choice layer. This input is thresholded by θ_q and amplified via a faster-than-linear activation function, $y(x) = x^2$ (Grossberg, 1973). This function can be thought of as a nonlinear neural response (e.g., spike rate varies nonlinearly with membrane potential) inherent to choice cells. The same activation function also guides self-excitatory activity among the choice cells in Equation 2.

The activity of a cell q_{ij} in the IFS choice layer q is given by

$$\dot{q}_{ij} = -A_q q_{ij} + (B_q - q_{ij})\Big(d_j[p_{ij} - \theta_p]^+ + y(q_{ij})\Big) - q_{ij}\left(\sum_{kj,k\neq i} \mathbf{W}_{ik} y(q_{kj}) + \Gamma_{ij}\right) \tag{2}$$

Excitatory inputs include a self-excitation term $y(q_{ij})$ and a selective input from the IFS plan cells in the same cortical column. The latter input is modulated by d_j, which represents a signal hypothesized to arise from the ventral anterior thalamus as the output of the BG-mediated planning loop (see Figure 2). The dynamics of this loop are specified below. The signal d_j serves as a "gate" that, when opened, allows plan cells to activate corresponding cells in the IFS choice layer and thereby initiate a competition among cells in that zone. In the model, such gateable zones (cf. Brown et al., 2004) constitute positional representations (the strips counted along the minor axis in Figure 4) within the IFS map.

The IFS choice cell q_{ij} is inhibited by all other cells within the same gateable zone. The action of the inhibitory cells is again faster-than-linear via signal function y. The resulting dynamics are such that choice cells are typically quiescent, but when a thalamic input gates on a positional zone, IFS plan cells are able to activate their corresponding choice cells, which in turn compete in a winner-take-all process (cf. the competitive layer in the CQ framework; Figure 1) within that positional zone. Once a choice cell "wins," it will maintain its activation for a short time through recurrent interactions. That cell's activity may be quenched via the potentially strong inhibitory input Γ_{ij}. This response suppression signal arrives from the SSM choice layer, described below. The value of Γ_{ij} is given by:

$$\Gamma_{ij}(t) = 10 Z_k^{ij} s_k(t) \tag{3}$$

where Z_k^{ij} is 1 if phoneme i occurs at syllable position j in the sensorimotor program k and 0 otherwise, and $s_k(t)$ is the activation of SSM choice cell k at time t (see Equation 14). Γ_{ij} therefore models the suppression of active phonological choice cells by chosen speech motor program cells in the SSM. It should be noted that only the phonemes that comprise the currently chosen motor program in the SSM are suppressed. This allows the model to produce unfamiliar syllables from targets representing its constituent segments (see below).

Structural "Frame" Representations in Pre-SMA

Cells in the pre-SMA are hypothesized to serve as representations of structural frames that code abstract structure at a level above the phonemic content represented in the IFS. Although alternative representations are also plausible, in the current proposal, pre-SMA cells code for common syllable types and for their abstract "slots" or positions. For example, the model pre-SMA contains cells that code for the syllable type CVC as well as for C in onset position, V in nucleus position, and C in coda position. Acquisition of this set of representations is feasible because of linguistic regularities; most languages use a relatively small number of syllable types. An analysis of frequency of usage tables in the CELEX lexical database (Baayen et al., 1995) revealed that just eight syllable frames account for over 96% of syllable productions in English.

The pre-SMA frame representations are activated in parallel with the IFS phonological content representation. Like the IFS planning layer, multiple pre-SMA frame cells can be active simultaneously in the plan layer. The relative activation levels of pre-SMA plan cells encode the serial order of the forthcoming syllable frames, with more activity indicating that a frame will be used earlier. The model thus represents a speech plan in two parallel and complementary queues, one in the IFS and one in the pre-SMA. This division of labor helps to solve a combinatorial problem that would result if all possible combinations of frame and content required their own singular representation. Such a scheme would require tremendous neural resources in comparison to the method proposed, which separates the representational bases into two relatively small discrete sets. The syllable frames [CV], [CVC], [VC], [CVCC], [CCV], [CCVC], and [VCC], the most common in English according to the CELEX database, were implemented. To allow for repeating frame types in a forthcoming speech plan, the model included multiple "copies" of each syllable frame cell.

The model pre-SMA contains not only cells that code for the entire abstract frame of a forthcoming syllable but also chains of cells that fire in rapid succession because they code for the individual abstract serial (phoneme-level) positions within the syllable frame. These two types of cells, one type that codes for an entire sequence (in this case a sequence of the constituent syllable positions within a syllable frame) and another type that codes for a specific serial position within that sequence, are similar to cells that have been identified in the pre-SMA in monkey studies (Shima & Tanji, 2000; Clower & Alexander, 1998). In the GODIVA model, the selection of a syllable frame cell (e.g., activation of a pre-SMA choice cell) also initiates the firing of the chain of cells coding its constituent structural positions (but not specific phonemes). The structure and the operation of the pre-SMA in the model are schematized in Figure 5.

For a single syllable, the temporal activity pattern in the pre-SMA proceeds as follows. First, a single choice cell is activated, corresponding to the most active syllable frame among a set of pre-SMA plan cells; upon the instantiation of this choice, the corresponding pre-SMA plan cell is suppressed. Next, the choice cell activates the first position cell in the positional chain corresponding to this syllable type. This cell and the subsequent cells in the positional chain give their inputs to zones in the caudate that have a one-to-one correspondence with positions in the syllable template and, equivalently, gateable zones in the IFS. These cortico-striatal projections form the inputs to the BG planning loop, which eventually enables the selection of the forthcoming syllable's constituent phonemes in the IFS choice field. When the positional chain has reached its completion, the last cell activates a corresponding cell in the SMA proper, which effectively signals to the motor portion of the circuit that the planning loop has prepared a new phonological syllable.

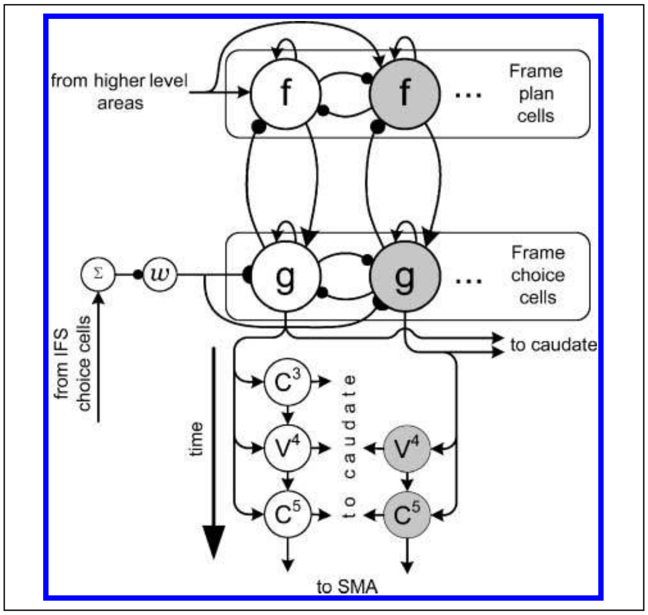

Figure 5. Schematic illustration of the structure and function of model cells hypothesized to exist in the pre-SMA. This region consists of a layer of plan cells (top) and a layer of choice cells arranged into columns, each of which corresponds to the same abstract syllable frame. When a pre-SMA choice cell is activated (i.e., the forthcoming frame is chosen), the cell gives inputs to a chain of cells, each of which corresponds to a position within the abstract syllable frame. These cells fire rapidly and in order, according to the vertical arrow labeled "time". In this schematic, the first pre-SMA cortical column codes for the syllable frame type [CVC], and the second column codes for the frame type [VC]. Note that the inputs to caudate are aligned such that the [V] position in both cases gives input to the same caudate channel (corresponding to Positional Zone 4). Cell *w* gates the pre-SMA frame choice process.

The pre-SMA frame cells are modeled by equations very similar to those describing IFS cell activity. These layers, again, embody a CQ architecture as described above. The activity of the *i*th frame cell in the pre-SMA plan layer, *f*, is given by

$$\begin{aligned}\dot{f}_i &= -A_f f_i + (B_f - f_i)\left(\alpha u_i^f + [f_i - \theta_f]^+\right)\\ &\quad - f_i\left(\sum_{k\neq i} f_k + 10y\left([g_i - \theta_g]^+\right)\right)\end{aligned} \tag{4}$$

where u^f is the external input to the pre-SMA, assumed to arrive from the same source that provides input u^p to the IFS. The activity of pre-SMA choice cell g_i is given by:

$$\begin{aligned}\dot{g}_i &= -A_g g_i + (B_g - g_i)\left(\omega[f_i - \theta_f]^+ + y(g_i)\right)\\ &\quad - g_i\left(\sum_{k\neq i} y(g_k)\right)\end{aligned} \tag{5}$$

Here the signal ω is modeled as a binary input, with value 1 when the IFS choice field is completely inactive and 0 when one or more cells are significantly active in that field, thereby serving as a "gate" to the pre-SMA frame choice process. Without such a gate, the pre-SMA choice process could proceed without pause through selection of each of the syllable frames represented in the graded pattern *f*. Instead, this gate requires the pre-SMA module to wait until the currently active syllable has been chosen for production on the motor side of the circuit. At such times, the choice of the *next* frame may proceed. This gating is implemented algorithmically but can be achieved through a cortico-cortical projection between IFS and pre-SMA via an inhibitory interneuron. This is schematized in Figure 5, where it is assumed that tonically active cell ω is quenched when any IFS choice cells are active above some low noise threshold.

As noted above, activation of a pre-SMA choice cell initiates a serial chain of cells that code for individual abstract positions in the syllable. The activity of the *j*th cell in the positional chain corresponding to syllable frame k is specified by

$$b_j^k(t) = \begin{cases} 1 & \text{if} \qquad (t_0 + (j-1)\tau) \le t \le (t_0 + j\tau) \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where t_0 is the time at which the pre-SMA choice cell g_k exceeds a threshold θ_g (the time at which it is "chosen"), and τ is a short duration for which each cell in the chain is uniquely active. Each of these cells gives input to a cell in the striatum corresponding to the same positional zone (see below). The deactivation of the final cell in the chain activates a model SMA cell that codes for the appropriate syllable type k.

Cortico-striato-pallido-thalamo-cortical Planning Loop

The GODIVA model posits that two parallel BG loop circuits form competitive gating mechanisms for cortical modules during the production of syllable sequences. The first loop, the planning loop, serves to enable activation of cortical zones in the choice layer of the model's left IFS. This loop receives inputs from the IFS plan cells (p) as well as from the more abstract pre-SMA positional cells (h). Following Brown et al. (2004), GODIVA hypothesizes that the one-to-many projection from thalamic output cells to a cortical zone serves to gate that zone's activation—ultimately allowing the flow of output signals via cortico-cortical projections. The model's subcortical circuitry is much simplified in comparison to other detailed treatments but remains compatible with Brown et al. In GODIVA, the critical role of the subcortical circuits is to coordinate signal flows throughout multiple levels of cortical representation. Although there is significant convergence within cortico-striato-pallidal pathways, the model treats these projections as a set of competitive channels, each represented by one striatal (caudate) projection neuron (b), one striatal interneuron ($\underline{b}$), and one pallidal (GPi) cell (c). This highly idealized circuitry is depicted in Figure 6. These channels correspond one-to-one with the gateable cortical zones in the IFS choice layer that, as described above, correspond to a set of abstract syllable positions. The activity of the striatal projection neuron in BG channel j is given by

$$\dot{b}_j = -A_b b_j + (B_b - b_j)\left(h_j \wedge \left[\sum_k p_{kj} - \delta\right]^+\right) - b_j\left(\sum_{k\neq j} y(\underline{b}_k)\right) \tag{7}$$

where $\wedge$ is the Boolean AND operator, assumed here to output 1 when both of its operands have value greater than zero, and 0 otherwise. A coincidence of suprathreshold activity in one or more IFS phonological plan cells tuned to position j, and significant input from pre-SMA cells coding for position j is required to drive activation of this striatal projection cell. The cell b_j also receives strong (modeled as faster than linear) feedforward inhibition from striatal interneurons $\underline{b}_k$ in the other BG channels ($k \neq j$). The activity of a striatal inhibitory interneuron in channel j is governed similarly by

$$\dot{\underline{b}}_j = -A_{\underline{b}}\underline{b}_j + (B_{\underline{b}} - \underline{b}_j)\left(h_j \wedge \left[\sum_k p_{kj} - \delta\right]^+\right) - \underline{b}_j\left(\sum_{k\neq j} y(\underline{b}_k)\right) \tag{8}$$

Thus, the model's cortico-striatal cells in both the IFS and the pre-SMA give inputs to the projection neurons and inhibitory interneurons in the model's caudate. Striatal projection cells connect to GPi cells within the same BG channel via an inhibitory synapse. The activity of the GPi cell c_j, which is itself inhibitory to a corresponding thalamic cell d_j, is given by

$$\dot{c}_j = -A_c c_j + \beta_c(B_c - c_j) - c_j(b_j) \tag{9}$$

where β_c and B_c control the level of spontaneous tonic activation of the GPi cell. Such tonic activation is required

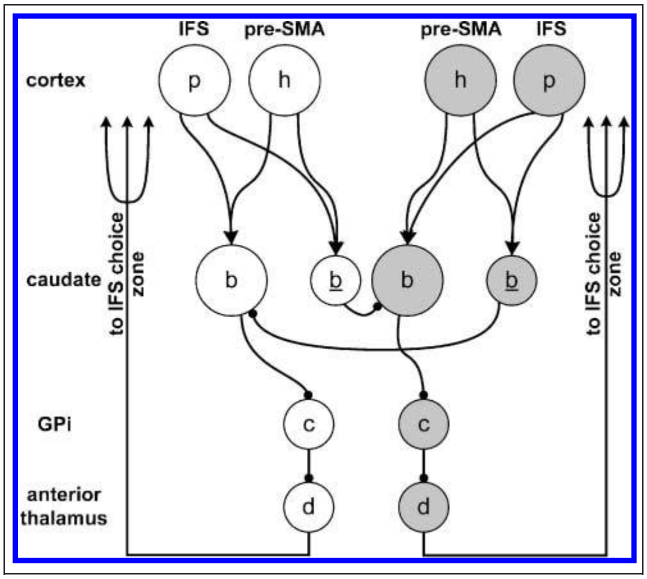

Figure 6. Schematic illustration of "channel" architecture through the BG planning loop. Each channel corresponds to an abstract serial position in the generic syllable template. The modeled caudate consists of one projection neuron (b) and one inhibitory interneuron ($\underline{b}$) in each channel. The channels compete via feedforward inhibition in the caudate. Caudate projection neurons give inhibitory projections to a modeled GPi cell (c). The GPi cell, in turn, inhibits the anterior thalamic cell d. The successful activation of a channel disinhibits its specific thalamic cell, which in turn "opens the gate" to a zone in the IFS phonological choice layer through a multiplicative interaction.

for the BG model to achieve the correct net effect within a channel. Specifically, the corresponding thalamic cell d_j must be normally silent but should become transiently activated upon the selective competitive activation of BG channel j. To achieve this result, GPi cells are tonically active but demonstrate a pause response when they are inhibited by the striatal projection neuron in the same channel. Because the projection from the GPi cell c_j to the anterior thalamus cell d_j is inhibitory, a pause response in c_j will disinhibit d_j and thereby enable the cortical selection process in zone j of the IFS choice layer (see Equation 2). Activity in thalamic cell d_j, which diffusely projects to zone j in the IFS choice layer, is given by

$$\dot{d}_j = -A_d d_j + \beta_d(B_d - d_j) - d_j(c_j) \tag{10}$$

Here β_d and B_d control the amplitude of the rebound excitation of the thalamic cell. A transient decrease in inhibitory input c_j leads to transient activation of d_j, enabling the cortical selection process for syllable position j in the IFS choice field.

Speech Sound Map

The SSM is a component of the DIVA model (Guenther et al., 1998, 2006; Guenther, 1995) that is hypothesized to contain cells that, when activated, "readout" motor programs and sensory expectations for well-learned speech sounds. In DIVA, tonic activation of an SSM cell (or ensemble of cells) is required to readout the stored sensory and motor programs throughout the production of the sound. To properly couple the system described herein with the DIVA model, GODIVA must provide this selective, sustained excitation to the appropriate SSM cells.

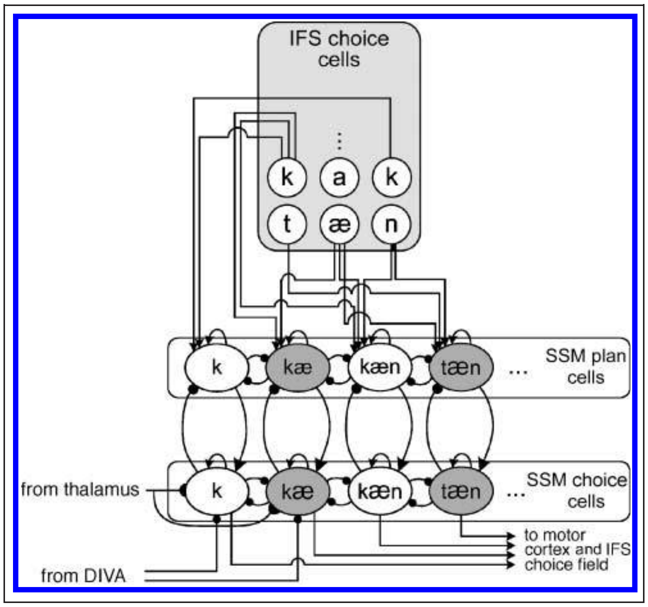

Figure 7. Illustration of the functional architecture of the model's SSM module. Columns consisting of a plan cell and a choice cell code for specific phonetic targets (for phonemes and syllables). IFS phonological choice cells give input to SSM plan cells that contain the phoneme for which they code. System dynamics allow only one SSM choice cell to remain active at a time. SSM choice cells give strong inhibitory input (not shown for simplicity) back to IFS choice cells to quench their constituent phonemes after their activation.

Like its other cortical representations, the GODIVA model's SSM is divided into two layers, again labeled plan and choice (Figure 7). Here, each idealized cortical column represents a well-learned syllable or phoneme. Unlike plan layers in the IFS and pre-SMA, however, the activation pattern across SSM plan cells does not code for serial order but rather indicates the degree of match between the set of active phonological cells in the IFS choice layer (the forthcoming phonological syllable) and the stored sensorimotor programs associated with the SSM columns. This match is computed via an inner product of the IFS choice layer inputs with synaptic weights that are assumed to be learned between these cells and the SSM plan cells. In the current implementation, these weights are "hand wired" such that the synapse Z_k^{ij} from IFS choice cell q_{ij} (which codes phoneme i at syllable position j) to SSM plan cell r_k is given by

$$Z_k^{ij} = \begin{cases} \frac{1}{N_k} & \text{if } r_k \text{ includes phoneme } i \text{ at position } j \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where N_k is the number of phonemes in the syllable coded by r_k. This specification indicates that an SSM plan cell receives equally weighted input from each IFS choice cell that codes its constituent phonemes in their proper syllabic positions and receives no input from other IFS choice cells. Furthermore, the sum of synaptic weights projecting to any syllable program in the SSM plan layer is equal to 1. Learning rules that conserve total synaptic strength have been proposed elsewhere (Grossberg, 1976; von der Malsburg, 1973), and similar conservational principles have been observed empirically (Markram & Tsodyks, 1996). An exception to the synaptic weight rule is made for SSM cells that code single phoneme targets (as opposed to entire syllables). These cells have synaptic inputs set equal to

$$Z_k^{ij} = \begin{cases} 0.85 - 0.05j & \text{if } r_k \text{ codes phoneme } i \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

In other words, the input to SSM plan cells coding for single phoneme targets is weighted by the position of IFS choice cells, such that inputs from earlier positions in the syllable have greater efficacy. This allows SSM plan cell inputs to maintain the serial order of the constituent phonemes in the IFS choice field in the case that the syllable must be produced from subsyllabic motor programs (e.g., when there is no matching syllable-sized SSM representation for the forthcoming phonological syllable). The activity level of cell k in the SSM plan layer representation r is governed by

$$\dot{r}_k = -A_r r_k + (B_r - r_k)\left(\sum_i \sum_j \mathbf{Z}_k^{ij} y\left([q_{ij} - \theta_q]^+\right) + [r_k - \theta_r]^+\right) - r_k\left(\sum_{n \neq k} r_n\right) \tag{13}$$

The double sum in the excitatory term above computes the net excitatory input from cells in the IFS choice field (q), which is weighted by the synaptic strengths specified in the input weight matrix $\mathbf{Z}_k$. Cell r_k also receives self-excitatory feedback (subject to a low noise threshold θ_r) and lateral inhibitory input from all other cells in the SSM plan layer. As in the other plan layers in the model, the interactions described by this equation allow multiple cells to sustain their activation cotemporally.

The SSM plan cell r_k gives specific excitatory input to the SSM choice cell s_k within the same idealized cortical column. The activation of s_k is given by

$$\dot{s}_k = -A_s s_k + (B_s - s_k)\left(r_k + 10y\left([s_k - \theta_s]^+\right)\right) - s_k\left(\sum_{j \neq k} [s_j - \theta_s]^+ + \Omega\right) \tag{14}$$

where y is again a faster-than-linear signal activation function, resulting in winner-take-all dynamics within the layer s. Ω models a nonspecific response suppression signal that arrives from the articulatory portion of the model, indicating the impending completion of production of the current syllable motor program. When Ω is large, activity is quenched in s, and a new winner is then instantiated, corresponding to the most active SSM program in the plan layer r. The DIVA model can provide such a suppression signal before actual completion of articulation but still related to the expected duration of the planned sound because of the inherent delay between sending a motor command and the effect that that motor command has on the articulators. Such delays in the production model have been considered by Guenther et al. (2006). Alternatively, in covert or internal speech, this completion signal might arrive from elsewhere, allowing the model to proceed through SSM programs without overtly articulating them.

Response Release via the "Motor Loop"

The initiation or release of chosen speech motor programs for overt articulation is hypothesized to be controlled by a second loop through the BG, the motor loop. The proposal that two loops through the BG, one mediated by the head of the caudate nucleus and one mediated by the putamen, are important in cognitive and motor aspects of speech production, respectively, was supported by intraoperative stimulation results demonstrating dysarthria and articulatory deficits when stimulating the anterior putamen and higher level deficits including perseveration when stimulating the head of the caudate (Robles, Gatignol, Capelle, Mitchell, & Duffau, 2005). In GODIVA, the motor loop receives convergent input from the SMA and motor cortex and gates choice (or execution) cells in the motor cortex (Figure 2). In keeping with established views of distinct BG–thalamo-cortical loops, the motor loop receives inputs at the putamen, whereas the planning loop receives inputs from "higher level" prefrontal regions at the caudate nucleus (Alexander & Crutcher, 1990; Alexander et al., 1986). The motor loop gives output to the ventrolateral thalamus, as opposed to the ventral anterior thalamic targets of the model's planning loop.

Currently, this motor loop in the GODIVA model is not specified with the same level of detail as the previously discussed planning and selection mechanisms in the model. To achieve the same level of detail, it will be necessary to fully integrate the circuits described above with the existing DIVA model. For the sake of clarity and tractability, we leave such integration for future work while focusing on the new circuitry embodied by the GODIVA model. Nevertheless, a conceptual description of these mechanisms is possible and follows from the general architecture of the higher level portions of the model. Specifically, the activation of an SSM choice cell representing the forthcoming speech motor program is hypothesized to activate plan cells in the left motor cortex. These plan cells do not directly drive movement of the articulators, just as plan cell activity in other modules in GODIVA does not directly drive activity beyond that cortical region. Instead, overt articulation in the model requires the enabling of motor cortex choice cells via the BG-mediated motor loop. To "open the gate" and initiate articulation, the motor loop requires convergent inputs from motor cortex plan cells and from the SMA proper. This notion is based on three major findings from Bohland and Guenther (2006), which are consistent with other reports in the literature. These results are (i) that overt articulation involves specific additional engagement of the SMA-proper, (ii) that the putamen is particularly involved when speech production is overt, and (iii) that whereas the left hemisphere motor cortex may become active for covert speech or speech preparation, overt speech engages the motor cortex in both hemispheres.

Figure 8 provides a summary of the process by which the model produces a sequence of syllables.

RESULTS

Computer simulations were performed to verify the model's operation for a variety of speech plans. Figures 9 and 10 show the time courses of activity in several key model components during the planning and production of the

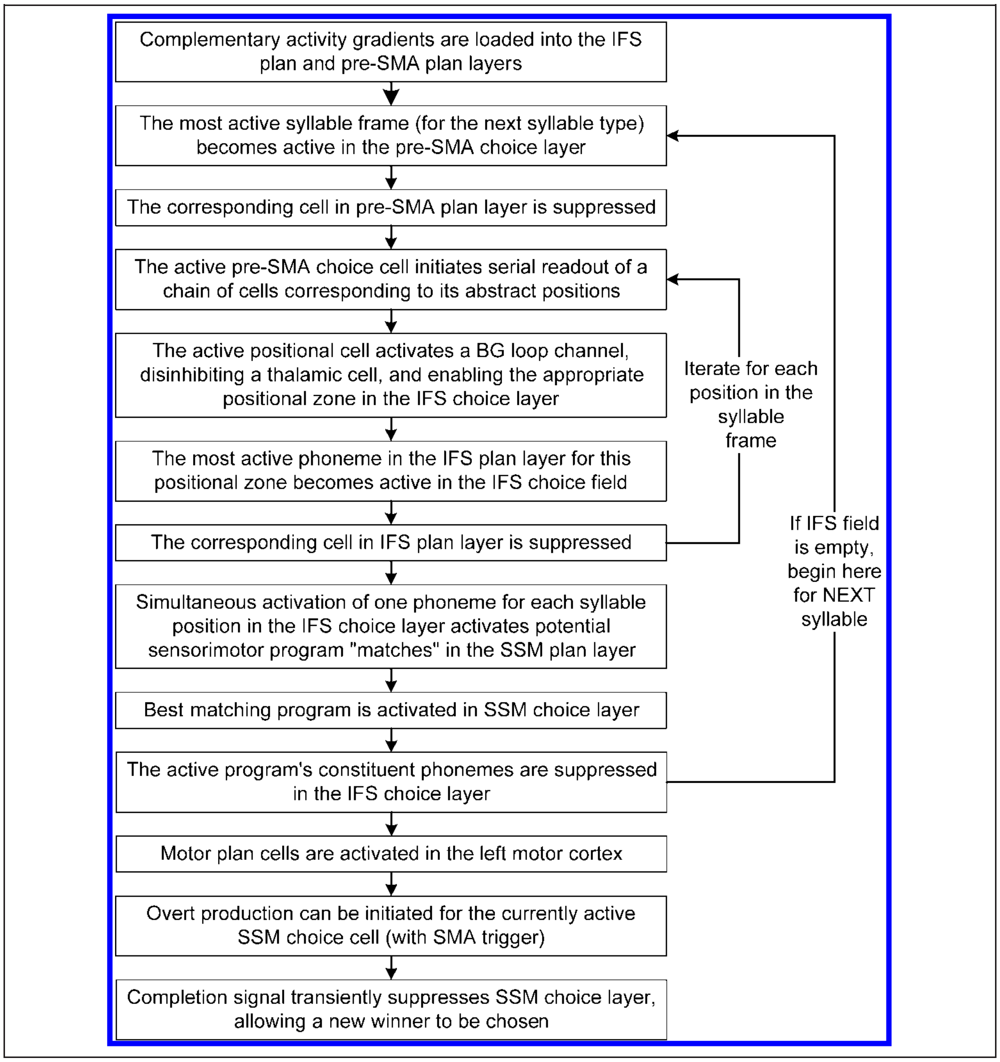
Complementary activity gradients are loaded into the IFS plan and pre-SMA plan layers
The most active syllable frame (for the next syllable type) becomes active in the pre-SMA choice layer
The corresponding cell in pre-SMA plan layer is suppressed
The active pre-SMA choice cell initiates serial readout of a chain of cells corresponding to its abstract positions
The active positional cell activates a BG loop channel, disinhibiting a thalamic cell, and enabling the appropriate positional zone in the IFS choice layer
The most active phoneme in the IFS plan layer for this positional zone becomes active in the IFS choice field
The corresponding cell in IFS plan layer is suppressed
Simultaneous activation of one phoneme for each syllable position in the IFS choice layer activates potential sensorimotor program "matches" in the SSM plan layer
Best matching program is activated in SSM choice layer
The active program's constituent phonemes are suppressed in the IFS choice layer
Motor plan cells are activated in the left motor cortex
Overt production can be initiated for the currently active SSM choice cell (with SMA trigger)
Completion signal transiently suppresses SSM choice layer, allowing a new winner to be chosen
Iterate for each position in the syllable frame
If IFS field is empty, begin here for NEXT syllable

Figure 8. An algorithmic summary of the steps that the GODIVA model takes to perform a syllable sequence.

syllable sequence "go.di.və" given two different assumptions about the model's initial performance repertoire. Figure 11 illustrates a typical phonological error made when noise is added to the model.

Performance of a Sequence of Well-learned Syllables

In the first simulation, the model's task is to produce this sequence assuming that each individual syllable ("go," "di," and "və) has been learned by the speaker and thus a corresponding representation is stored in the model's SSM. Sensorimotor programs for these syllables must be acquired by the DIVA portion of the circuit; this learning process is described elsewhere (Guenther et al., 2006; Guenther, 1995). In this simulation, the 1,000 most frequent syllables from the CELEX database (which include the three syllables to be performed here) are represented in the SSM. The "input" to this simulation is a graded set of parallel pulses, applied at the time indicated by the first arrow in each panel of Figure 9. This input activates the two complementary gradients in the IFS plan layer (Figure 9A and B) and in the pre-SMA plan layer (not shown). This mimics the input signals that are hypothesized to arise from higher order linguistic areas. These inputs create an activity gradient across the /g/, /d/, and /v/ phoneme cells in Syllable Position 3 (onset consonant) and a gradient across the /o/, /i/, and /ə/ phoneme cells in Syllable Position 4 (vowel nucleus) in the IFS plan layer as well as a gradient across three "copies" of the [CV] frame cell in the pre-SMA. Figure 9A and B shows that the activation levels of the phonemes in these positional zones rise from the initial state of 0 and begin to equilibrate with each cell taking on a distinct activation level, thus creating the activity gradients that drive sequence performance.

After the first frame representation is activated in the pre-SMA choice layer, Positional Zones 3 and 4 are enabled in rapid succession in the IFS choice layer. This allows the most active phoneme in each IFS positional zone to become active. Figure 9C and D reveals this choice process, which results in the strong, sustained activation of the phonemes /g/ and /o/ in IFS Choice Zones 3 and 4, respectively, with the activation of Zone 4 occurring slightly later than activation in Zone 3. Immediately after the choice of /g/ and /o/ (Figure 9C and D), the representations for these phonemes in the IFS plan layer (Figure 9A and B) are rapidly suppressed. IFS plan layer activity then re-equilibrates, leaving only two active phonemes in each zone, now with a larger difference in their relative activation levels.

The cotemporal activation of cells coding /g/ and /o/ in the IFS choice layer (Figure 9C and D) drives activity in the model's SSM plan layer (Figure 9E). Multiple cells representing sensorimotor programs for syllables and phonemes become active, each partially matching the phonological sequence represented in the IFS choice layer. The most active of these SSM plan cells codes for the best matching syllable (in this case "go"). This most active syllable becomes active also in the SSM choice layer (Figure 9F). As soon as "go" becomes active in Figure 9F, its constituent phonemes in the IFS choice layer (Figure 9C and D) are suppressed. The resulting lack of activity in the IFS choice layer then enables the choice of the next CV syllable frame in the pre-SMA, allowing the model to begin preparing the syllable "di" (up to the stage of activating potential SSM matches in the SSM plan cells) while it is still producing the syllable "go" (compare Figure 9C–E with Figure 9F). The syllable "di" however, can only be chosen in the SSM choice layer (Figure 9F)

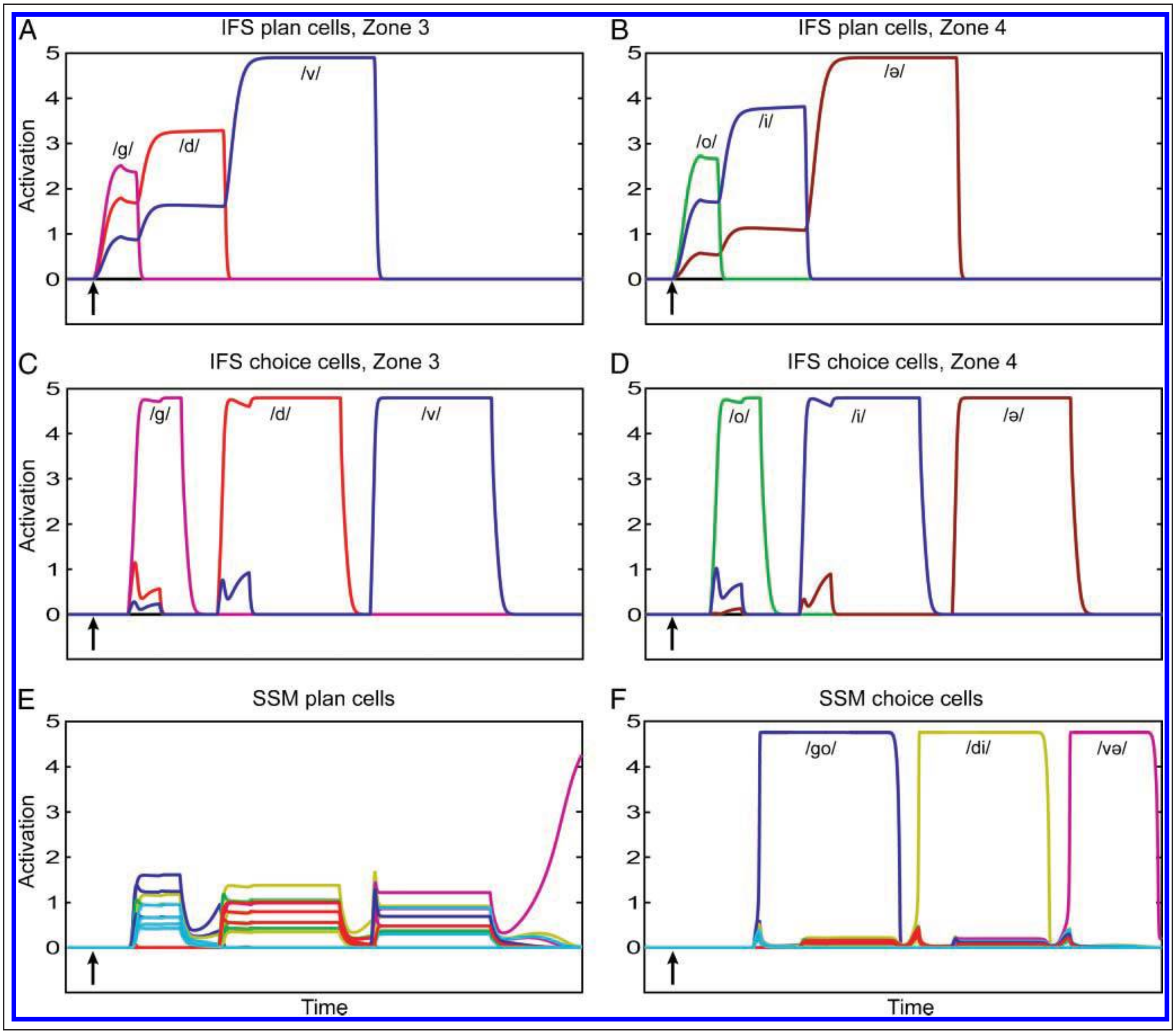

Figure 9. Simulation result showing the production of the three syllable sequence "go.di.və." In this simulation, each of the three syllables has a corresponding stored SSM representation. Each plot shows time courses of cell activity in different model components. The *x*-axis in each plot is time, and the *y*-axis is activation level (both in arbitrary model units). The arrows in each plot indicate the onset of the external input at the start of the simulation. See text for details.

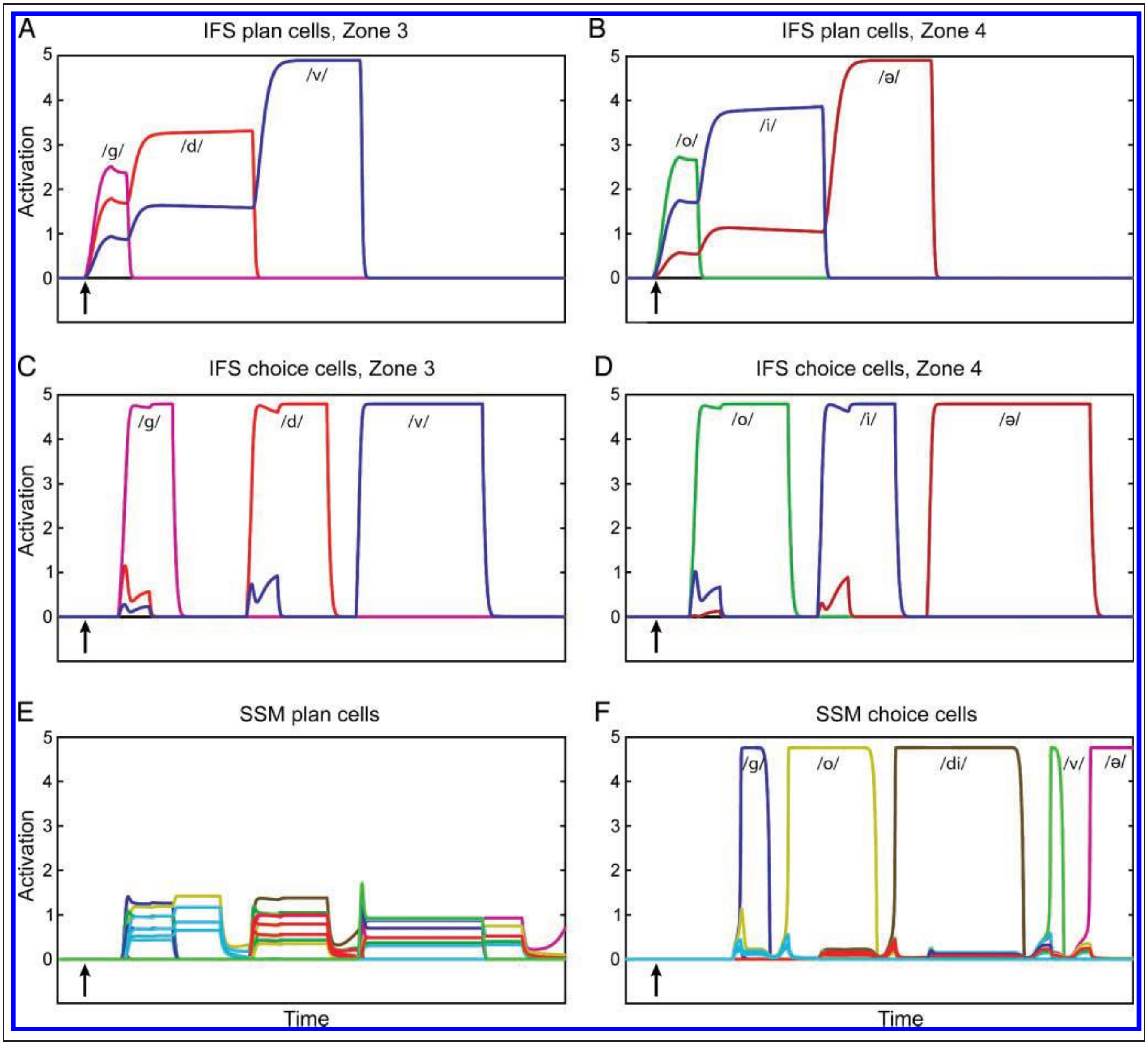

Figure 10. Simulation result showing the production of the syllable sequence "go.di.və." using piece-wise sensorimotor programs. In this simulation, only the second syllable ("di") has a corresponding representation in the SSM. The model must perform the first and the third syllables, therefore, by sequentially activating targets for the constituent phonemes in those syllables. Each plot shows time courses of cell activity in different model components. The *x*-axis in each plot is time, and the *y*-axis is activation level (both in arbitrary model units). The arrows in each plot indicate the onset of external input at the start of the simulation. See text for details.

upon the receipt of a nonspecific suppression signal from the articulatory control circuit. This inhibitory signal transiently quenches all activity in the SSM choice layer, which can be seen by the rapid decrease in activation of the cell coding for "go" in Figure 9F. Upon removal of this suppression signal, "di," the most active SSM plan representation, is chosen in the SSM choice layer. This entire process iterates until there are no remaining active cells in the pre-SMA or IFS plan layers. It can be seen from Figure 9F that the syllable motor programs corresponding to the desired syllables receive sustained activation, one at a time, in the proper order. This is precisely what is required to interface GODIVA with the DIVA model, which can then be used to control a computer-simulated vocal tract to realize the desired acoustic output for each syllable.

Performance from Subsyllabic Targets

We have emphasized a speaker's ability to represent and to produce arbitrary syllable sequences that fall within the rules of her language. By planning in the phonological space encompassed by the IFS and pre-SMA categorical representations, the GODIVA model does not rely on having acquired phonetic or motor knowledge for every syllable it is capable of planning and/or producing. Instead, the model is capable of producing unfamiliar syllables by activating

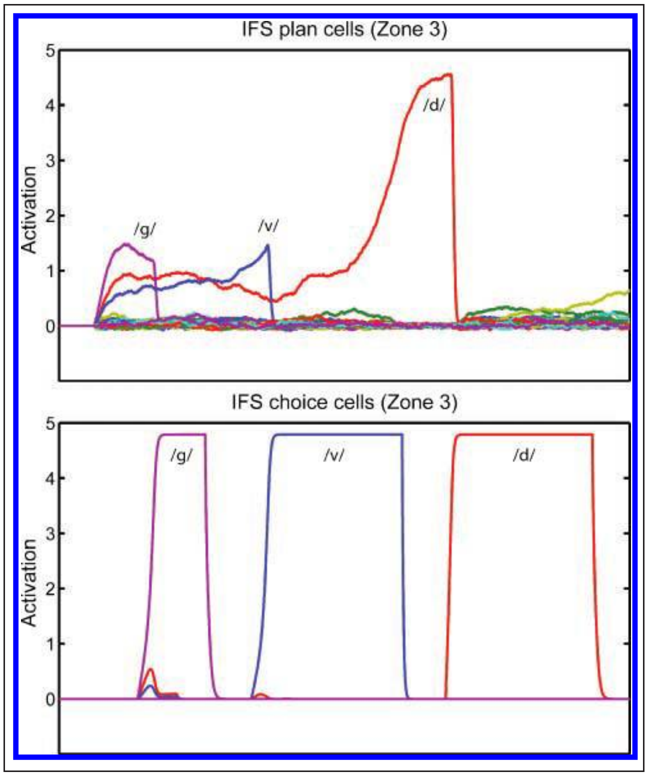

Figure 11. Simulated results in the IFS Zone 3 plan and choice cell layers for a simulation of the intended syllable sequence "go.di.və" with Gaussian noise added to IFS plan cells. The simulation was chosen from multiple stochastic versions to illustrate how the model can produce phoneme exchange errors that obey syllable position constraints (cf. MacKay, 1970). Because of noise, the plan representation for /v/ (blue) becomes greater than that for /d/ (red) and is thus selected as part of the second syllable in the sequence. The plan for /d/ remains active and is chosen as the onset of the third syllable. Thus, the model produces the sequence "go.vi.də" in error.

an appropriate sequence of subsyllabic phonetic programs. This point is addressed in a simulation that parallels the one described above but makes different assumptions about the initial state of the model's SSM.

Figure 10 shows the model again producing the syllable sequence "go.di.və," but in this simulation the syllables "go" and "və" have each been removed from the model's SSM. The system thus no longer has the requisite motor programs for these syllables, and it must effect production using a sequence of smaller stored programs corresponding to the syllables' individual phonemes. Figure 10F shows that the model activates SSM choice cells coding the constituent phonemes, in the correct order, for the first and third syllables of the planned utterance. The SSM program associated with the second syllable, "di," remains as a possible match in the SSM and is correctly chosen for production.

Figure 10C–E demonstrates the model's operation when individual syllables must be created from phonemic motor programs. A comparison of Figure 10C and D reveals that the IFS choice cell for the first phoneme (/g/ of the syllable "go") is suppressed before suppression of the phoneme /o/. This is because the inhibition of IFS choice cells is dictated by which sensorimotor program is chosen in the SSM choice layer. Because no SSM cell matches "go" exactly, the best matching cell (as determined by the dot product of IFS choice layer activity with each SSM plan cell's stored synaptic weights; see Equation 13) codes for the phonetic representation of the phoneme /g/. This cell is activated in the SSM choice field (see Figure 10F) and inhibits only the phonological representation of /g/ in the IFS choice layer (Figure 10C). Because the phoneme /o/ remains active in IFS choice field zone 4 (Figure 10D), the preparation of the *next* syllable cannot yet begin. Instead, SSM plan cell activity (Figure 10E) is automatically regulated to code for degree of match to the remaining phonological representation in the IFS choice field (in this case the single phoneme /o/). Once the nonspecific quenching signal arrives at the SSM choice field to indicate impending completion of the motor program for /g/, the program for /o/ can be chosen. Finally, the entire IFS choice field (in both Zones 3 and 4; Figure 10C and D) is inactive, allowing the pre-SMA to choose the next syllable frame and continue the sequencing process.

Production of Phonological Errors

Figure 11 shows the time course of activity within Zone 4 of the IFS during another simulation of the intended sequence "go.di.və," which results in a phonological error due to the addition of noise. In this example $\sigma_f = 1.0$ (see Equation 1), corresponding to a large Gaussian noise source giving independent input to each cell at each time step. Here, after the correct choice of the onset phoneme in the first syllable, noise is able to drive the IFS plan for /v/ (blue) to a higher activation level than the plan for /d/ (red) before its selection in the IFS choice layer. Despite the improper choice of the onset phoneme for the second syllable, the system continues to behave as usual, ultimately resulting in production of the syllable "vi" in place of the intended "di" (SSM cell activations not shown for brevity). Activation of /v/ in the IFS choice field causes the corresponding /v/ plan cell to be quenched in the IFS plan layer, leaving /d/ as the remaining significantly active onset plan. The /d/ plan is subsequently chosen and paired with the vowel /ə/, resulting in the completion of the syllable sequence "go.vi.də," an example of a simple *exchange* error that obeys syllable position constraints as noted in previous studies (e.g., Shattuck-Hufnagel, 1979; MacKay, 1970). It should be noted that whether this error follows a word position constraint is determined merely by the placement of word boundaries (e.g., *go diva* vs. *godi va*). In the present treatment, we do not explicitly deal with word-level representations but suggest that substitutions across word boundaries will be produced, provided multiple word representations can be activated simultaneously in the IFS, which we strongly suggest is the case.

DISCUSSION

We have presented a neurobiologically based model that begins to describe how syllable sequences can be planned and produced by adult speakers. The above simulations demonstrate the model's performance for sequences of both well-learned and uncommon syllables. Although the model attempts to index syllable-sized performance units, its underlying planning representation consists of categorical phonemes and abstracted syllable frames. Although we have not focused on modeling the rich patterns in speech error data, these representations naturally account for the most typical slips of the tongue. GODIVA builds on much previous theoretical work, beginning with the seminal contributions of Lashley (1951). Lashley's ideas can be viewed as a precursor to CQ proposals (Bullock, 2004; Bullock & Rhodes, 2003; Houghton & Hartley, 1995; Houghton, 1990; Grossberg, 1978a, 1978b), which are used in multiple places within GODIVA. The encoding of serial order by a primacy gradient is a fundamental prediction of CQ-style models that has received experimental support (Averbeck et al., 2002, 2003). Such order-encoding activity gradients underlie the choice of the model's name (Gradient Order DIVA; GODIVA). The GODIVA modules specified here operate largely "above" the DIVA model in the speech production hierarchy; these modules act to select and activate the proper sensorimotor programs and to initiate the production of chosen speech sounds. Online motor control for the individual speech motor programs as well as their acquisition is the function of the DIVA model itself, which has been described in detail elsewhere (Guenther, 1994, 1995, 2006; Guenther et al., 1998, 2006). DIVA also is responsible for coarticulation, which can be absorbed into learned programs, and can also cross the boundaries of individual "chunks."

That GODIVA was not developed in isolation, but rather as a continuation of an existing model of the neural circuits for speech and language production, is an important characteristic. Although future effort will be required to more fully integrate GODIVA with DIVA, here we have laid the groundwork for a comprehensive computational and biologically grounded treatment of speech sound planning and production. Each component of the GODIVA model, after previous efforts with DIVA (Guenther, 2006; Guenther et al., 2006), has hypothesized cortical and/or subcortical correlates. GODIVA thus appears to be the first treatment of the sequential organization and production of speech sounds that is described both formally and with detailed reference to known neuroanatomy and neurophysiology.

Representations of Serial Order

Although the CQ architecture plays a fundamental role in GODIVA, it is not the only ordinal representation used. The IFS representation combines elements of both CQ and positional models. Specifically, the minor axis of this two-dimensional map (Figure 4) is proposed to code for abstract serial position within a syllable. The inclusion of cells that code for both a particular phoneme and a particular syllable position may seem unappealing; the use of multiple "copies" of nodes coding for a single phoneme has often been criticized for failing to encapsulate any relationship between the nodes (e.g., Dell, 1986). In the proposed IFS representation that relationship is captured, to an extent, because "copies" of the same phoneme will always appear in close topographic proximity so long as the two-dimensional grid is mapped contiguously onto the cortical sheet. Additionally, and perhaps more importantly, this position-specific representation, which was motivated here and elsewhere by the strong syllable position constraint in speech errors, is computationally useful. Because IFS cells interact only within a positional zone, the IFS field can be thought to contain multiple queues. The capacity of a single queue (i.e., a planning layer) in CQ models is limited by noise; as additional elements are added, the difference between activation levels of any two elements to be performed successively is reduced. With the addition of zero-mean Gaussian noise, the probability of serial order errors at readout also becomes larger with additional elements. By dividing a phonological plan among multiple queues, the effect of noise is less destructive than it would be in a single queue, and the overall planning capacity is effectively increased. Although we have implemented a system with effectively seven queues, based on a generic syllable template, we remain agnostic to the precise details of the actual phonological map but suggest that the general principles outlined here present a plausible representational scheme in view of the sparse existing evidence. The idea of serial position-specific representations, while useful and supported by behavioral data, is less appealing for modeling simple list memory, general movement planning, and many other sequential behaviors because the number of "slots" is often ambiguous and the number of possible items that must be available at any serial position can be quite large. The phonotactic constraints of a language, however, reduce the set of possible phonemes at any given position.

GODIVA also includes "serial chain" representations within the pre-SMA module. The inclusion of these specific chains does not, however, invite all of the same criticisms that pertain to associative chaining as an exhaustive theory of serial order. This is because the total number of sequences that are encoded in this manner is small, corresponding to the number of abstract structural syllable frames available to the speaker (as discussed above, just eight syllable types account for almost all productions), and the order of the elements within a particular frame type is fixed. This leads to some general guiding principles that appear useful in modeling hierarchical sequential behavior. When sequence production must be generative,[7] associative chaining quickly becomes problematic, and the use of CQ-type activation gradients to encode order is preferred. When a small set of sequences becomes highly stereotyped, however, readout by serial or "synfire" chains (e.g., Pulvermüller, 1999, 2002; Abeles, 1991) can offer greater efficiency. The GODIVA model thus leverages these different representations when sequencing demands differ. Similarly, Dell (1986) speculated that a principle explanation for the appearance of speech errors in normal speech is the speaker's need for productivity or generativity. To produce novel linguistic sequences, it is necessary to "fill" slots in a sequence, and this allows the possibility of error due to difficulties with the "filling-in" mechanism(s). Dell argues that the set of possible phonemes is closed (after acquisition), whereas the set of possible phoneme combinations is open. CQ provides an efficient and physiologically plausible mechanism for representing this open set of combinations, which is problematic for other proposals. In this framework, it is intuitive that the units that slip during sequence production should be the units that form the bases in the CQ planning layer, in this case phonemes.

Development of the Speech and Language System

Our model is built around an assumed division of syllabic frames from phonemic content, but this system must be learned by speakers. MacNeilage (1998) has proposed that speech evolved the capability to program syllabic frames with phonological content elements and that every speaker learns to make use of this capacity during his or her own period of speech acquisition. Developing speakers follow a trajectory that may give rise to this factorization of frame and content. At approximately 7 months, children enter a canonical babbling stage in which they rhythmically alternate an open and a closed vocal tract configuration while phonating, resulting in repeated utterances such as "ba.ba.ba.ba." MacNeilage and Davis (1990) have suggested that these productions represent "pure frames." These reduplicated babbles dominate the early canonical babbling stage but are largely replaced by variegated babbling at around 10–13 months. This stage involves modifications of the consonant and vowel sounds in babbles, resulting in syllable strings such as "ba.gi.da.bu." MacNeilage and Davis suggest that this stage may represent the earliest period of content development.

Locke (1997) presented a theory of neurolinguistic development involving four stages: (1) vocal learning, (2) utterance acquisition, (3) structure analysis and computation, and (4) integration and elaboration. Locke (p. 273) suggests that in Stage 2, "every utterance [children] know is an idiom, an irreducible and unalterable 'figure of speech.'" This irreducibility was supported by the finding that very young children make far fewer slips of the tongue than adult speakers (Warren, 1986). It is only at the onset of Stage 3, around 18 to 20 months, that children gain the ability to "analyze" the structure of their utterances, recognizing, for example, recurring elements. This stage may provide the child with the representations needed for phonology, enabling generativity and the efficient storage of linguistic material. Importantly, at around 18 months of age, the rate of word acquisition in children may quadruple (Goldfield & Reznick, 1990). The timing of this explosion in a child's available vocabulary also roughly coincides with development in the perceptual system at approximately 19 months, at which time children can effectively discriminate the phonetic categories in their language (Werker & Pegg, 1992).

We take the position that the stages of speech acquisition up to and including variegated babbling are particularly important for tuning speech-motor mappings such as those described by the DIVA model of speech production (Guenther et al., 1998; Guenther, 1995). These stages also provide a "protosyllabary" of motor programs that are "purely motoric," having little to no linguistic significance (Levelt et al., 1999). A later stage, perhaps Locke's (1997) Stage 3, leads to development of phonological representations that can become associated with the phonetic programs that realize those speech sounds. This allows the learning speaker to insert content items into common learned syllable frames, thus offering an explanation for the rapid increase in the vocabulary at this time. Furthermore, this representation of the common sound elements in a speaker's language should remain largely unchanged after learning and can be used by the adult speaker to interface both words and nonwords with a more plastic speech motor system. In a sense, this representation provides a basis for representing any utterance in the language. The GODIVA model describes the speech system after the development of this stage and leverages this basis to allow generative production of novel sound sequences.

Comparison with Other Computational Models

The WEAVER (and later WEAVER++) model (Levelt et al., 1999; Roelofs, 1997) is broadly a computer implementation of the Nijmegen model. In WEAVER, a selected morpheme activates nodes representing its constituent phonemes and a metrical structure, which specifies the number of syllables and stress pattern. The order of the activated phonemes is assumed to be encoded by links between the morpheme and the phoneme nodes; likewise, links between phoneme nodes and nodes that represent phonetic syllables (e.g., motor programs) are also "labeled" with positional information (indicating onset, nucleus, or coda). Although WEAVER(++) is an important formalization of an influential language production model and shares certain similarities with GODIVA, its focus is somewhat different. Although GODIVA makes specific proposals about representations for order and their instantiations in neural circuits, the WEAVER model's use of rule-based labeling of nodes and links is difficult to reconcile in terms of potential brain mechanisms. The flow of information in the model is also not explicitly linked to regions and pathways in the cortex; thus, the ability to make inferences about neural function based on this model is limited. The GODIVA model is intended to bridge this gap between theoretical information processing and the neural substrates that implement such processes.

Dell's (1986) spreading activation model offers a formal explanation for various speech error data and represents the archetypal "frame-based" model. The proposal uses representations at several hierarchically organized linguistic levels such that nodes at one level are activated by nodes one level higher. Representations of the forthcoming utterance are built through a process of tagging the most active nodes at each level, largely in parallel. Nodes are then labeled with linguistic categories; in phonological encoding, for example, phonemes are labeled as onset, nucleus, or coda position in a syllable. A syllable frame, or ordered set of categories, is used to tag the most active nodes within the appropriate categories. The frame thus dictates not which elements are tagged but which are eligible to be tagged, much like in GODIVA. Dell's model formalized several theoretical proposals that had been proposed to explain speech errors within a network architecture, rendering the theories somewhat more biological but leaving possible anatomical substrates unspecified.

Hartley and Houghton (1996) described a CQ-based model that also exploits the frame-content division to explain learning and recall of unfamiliar nonwords in verbal STM. Individual syllables are represented in terms of their constituent phonemes and the "slots" that they use in a generic syllable template adapted from Fudge (1969). A pair of nodes is allocated for each syllable presented for recall, representing the syllable onset and rime, and temporary associative links are formed between these node pairs and both the appropriate syllable template slots and the phoneme content nodes for each syllable presented. During recall, an endogenous control signal imparts a gradient across syllable nodes, with the immediately forthcoming syllable receiving highest activation (see also Burgess & Hitch, 1992). The most active syllable pair is chosen for output and gives its learned input to the syllable template and phoneme nodes. As each syllable slot becomes activated (iteratively), phoneme nodes also become activated, with the most active nodes generally corresponding to phonemes from forthcoming syllables that occupy the same slot. The most active phoneme node is then chosen for "output," its activity suppressed, and so on until the sequence is completed. The model advances earlier proposals and does not require multiple versions of each phoneme for different serial positions. This requirement is obviated by using a single-shot learning rule to make appropriate associations between position and phonemes; it is not clear, however, how such learning would be used in self-generated speech.

Vousden et al. (2000) presented a similarly motivated model that is derived from a previous proposal for serial recall (Brown et al., 2000). The model postulates the existence of a dynamic, semiperiodic control signal (the phonological context signal) that largely drives its operation. A major goal of Vousden et al. was to eliminate the necessity for syllable position-specific codes for phonemes. Although appealing, this "simplification" requires a rather complex control signal derived from a large set of oscillators. The signal is designed to have autocorrelation peaks at specific temporal delays, reflected by the pool of low-frequency oscillators. In the reported simulations, this periodicity occurs every three time steps, which allows each state in a single period to be associated with an onset, a nucleus, or a coda phoneme. Recall of a sequence depends on learning a large set of weight matrices that encode associations between the context signal and a matrix constituting the phoneme representation, which is potentially problematic for novel or self-generated sequences. At recall, the context vector is reset to its initial state and "played back," resulting in a gradient of activations across phonemes for each successive contextual state. The typical CQ mechanisms are then used to allow sequence performance. Several concerns arise from the model's timing, association, and recall processes; see the critiques of this class of models in Lewandowsky, Brown, Wright, and Nimmo (2006) and Agam, Bullock, and Sekuler (2005).

Repeating Elements

One of the weaknesses of CQ theories concerns representing elements that repeat in a sequence. Because cells code for items and those cells' activity levels code for the items' serial order, it is problematic to represent the relative order of the same item occurring twice or more in the planned sequence. GODIVA uses perhaps the simplest strategy to handle repeating elements, by including multiple "copies" of each representative cell in the IFS and pre-SMA plan layers. With this addition, order is maintained simply by using a different copy of the requisite phoneme or frame cell for each occurrence of that phoneme or frame in the sequence. The sequence "pa.ta.ka" would thus require three different copies of the /a/ phoneme cell in Positional Zone 4 of the IFS. The implementation of this strategy requires some additional *ad hoc* machinery. Specifically, the model's external input, when targeting a phoneme cell in the IFS or frame cell in the pre-SMA, must activate a cell of that type that is currently inactive.

When entire syllables (performance units), on the other hand, are to be repeated by the model (e.g., "ta.ta.ta"), a different assumption is made. On the basis of RT data from Schönle, Hong, Benecke, and Conrad (1986) as well as fMRI observations described by Bohland and Guenther (2006), it appears that producing the same syllable *N* times is fundamentally different from producing *N* different syllables. We therefore assumed that planning a sequence such as "ta.ta.ta" only requires the phonological syllable "ta" to be represented in the complementary IFS and pre-SMA planning layers once. An additional mechanism is proposed to iterate the production portion of the circuit *N* times without the need to specify the phonological representation again each time.

A General Framework

Although the current proposal does not model higher level aspects of language production, the general architecture appears to have potential for reuse throughout the language system. The organization of BG circuits into largely parallel loops (Alexander & Crutcher, 1990; Alexander et al., 1986) offers one possible substrate for the cascaded processing stages that enable linguistic selections from competing alternatives; these selections (cf. choice layer activations) can then activate lower level representations through cortico-cortical pathways (as IFS choice cells, for example, activate SSM plan cells). Such loops may be nested to account for various levels of language production (e.g., Ward, 1994; Garrett, 1975). The GODIVA model architecture also offers an account for how learned structural patterns can be combined with an alphabet of "content" items in a biologically realistic circuit. In the same way that an abstract CV structure combines with representative phoneme units, syntactical structure might, for instance, combine with word units from different grammatical categories (cf. different positional zones). There is evidence that BG loops might indeed take part in selection mechanisms for higher level aspects of language. Damage to portions of the caudate can give rise to semantic paraphasia (Kreisler et al., 2000), a condition marked by the wrongful selection of words, in which the selected word has meaning related to the target word. Crinion et al. (2006) have also suggested that the caudate may subserve selection of words from a bilingual lexicon.

Relevance in the Study of Communication Disorders

Many authors have stressed the usefulness of comprehensive models in the study of communication disorders (e.g., McNeil et al., 2004; Ziegler, 2002; Van der Merwe, 1997). Current speech production models have largely failed to shed light on disorders such as apraxia of speech (AOS) because "theories of AOS encounter a dilemma in that they begin where the most powerful models of movement control end and end where most cognitive neurolinguistic models begin" (Ziegler, 2002). The GODIVA model is the first step in an attempt to bring the DIVA model (the "model of movement control") into a broader neurolinguistic setting. In so doing, the hope is that communication disorders such as AOS and stuttering can be better understood in terms of pathological mechanisms within the model that can be localized to brain regions through experimentation. As an example, in GODIVA, the symptoms of AOS, particularly groping and difficulty reaching appropriate articulations, might be explained by at least two mechanistic accounts. The first possibility is that the motor programs for desired sounds are themselves damaged. In the model, this amounts to damage to the SSM (lateral premotor cortex/BA44) or its projections to the motor cortex. An alternative explanation could be that these sensorimotor plans are intact, but the mechanism for selecting the appropriate plan is defective. This would occur in the model with damage to connections between the IFS choice layer and the SSM. A major focus of future research within this modeling framework should be the consideration of speech disorders.

Expected Effects of Model "Lesions"

One of the major reasons for hypothesizing specific neural correlates for model components (cf. existing psycholinguistic models without such specificity) is to make predictions about the effects that focal lesions might have on normal speech function. Although detailed simulations will need to be presented in future work, we can make some preliminary predictions presently. First, specific lesions to the left lateral pFC (in the area of IFS) will likely impact phonological encoding at the phoneme level. This may result in phonemic paraphasias, including substitutions, anticipations, and perseverations, which are observed in some Broca's aphasics. Because choice of syllable frames in the pre-SMA "starts" the production process, damage here could result in reductions in self-initiated speech (Jonas, 1981, 1987) but also may result in "frame deficiencies"—perhaps taking the form of reducing complex frame types to simpler ones. Damage to the BG planning loop may impact selection and notably the timing of selection processes in phonological encoding, which is consistent with some observations (Kreisler et al., 2000; Pickett et al., 1998). Finally, damage to the SSM or the interface between the IFS choice field and the SSM (e.g., damage to cortico-cortical projections) should lead to problems in realizing a phonetic plan or diminished ability to choose a phonetic plan; these deficits would be in line with observations in patients with AOS (Hillis et al., 2004).

Other Experimental Predictions

Any model of a system as complex as that considered here will eventually be found to have significant flaws. One of the most useful aspects of any model that can be simulated under various conditions is to generate predictions that can be tested experimentally. Through the generation of testable predictions, the model may be proven invalid, but new proposals will arise from this knowledge that further our understanding of the system. The GODIVA model makes many such predictions. For example, GODIVA predicts that the set of IFS choice layer to SSM plan layer connections implements a selection process whereby the strength of input to an SSM plan cell depends on how strongly the speech sound corresponding to that cell matches the currently planned syllable in IFS. This leads to the prediction that when many cells in the SSM code for sounds that partially match the syllable planned in IFS, the overall activation of the SSM will be larger than when there are few partial matches. More broadly speaking, planning and producing syllables with dense phonological neighborhoods are predicted to result in greater activation of the SSM than planning and producing syllables with sparse neighborhoods. This type of prediction is readily testable using fMRI or PET. A continued program of model development combined with targeted experimentation will be critical to better understanding the speech system.

Future Directions

The model and the conceptual framework we have presented here should be viewed as a preliminary proposal that will require considerable expansion to fully treat the myriad issues involved in planning and producing fluent speech. These include expansion of the model to address processing at the level of words and higher, which we believe can be incorporated gracefully in the existing framework. In future work, we plan to more fully address the rich patterns of observed speaking errors in normal and aphasic speakers, which may require further examination of the proposed syllable "template" and set of available syllable frames. Further, it is of interest to more closely examine how treatment of speech sequences (at the level of syllables or multisyllabic words) changes as they go from completely novel to familiar, to highly automatized, yet at every stage maintain the ability to be modulated by the speaker in terms of rate, emphasis, or intonation. We plan to explore the probable role of the cerebellum in phonological and phonetic processes (e.g., Ackermann, 2008), including a role in on-line sequencing, for example, by fast parallel loading of phonological memory buffers (cf. Rhodes & Bullock, 2002) and in the coordination and regulation of precise temporal articulation patterns. The cerebellum may also be involved in the generation of prosody (Spencer & Slocomb, 2007), along with other structures, and future instantiations of the GODIVA model should strive to explain how prosody and stress can be encoded at the phonological and phonetic levels.

Acknowledgments

The authors thank Jason Tourville, Satrajit Ghosh, and Oren Civier for valuable comments. Support was provided by NIH R01 DC007683 and NIH R01 DC002852 (F. H. Guenther, PI) and NSF SBE-0354378 (S. Grossberg, PI).

Reprint requests should be sent to Jason W. Bohland, Boston University, Sargent College of Health & Rehabilitation Sciences, Department of Health and Sciences, 635 Commonwealth Avenue, Boston, MA 02215, or via e-mail: jbohland@gmail.com.

Notes

[1] In DIVA, a speech sound can be a phoneme, a syllable, an entire word, etc. For the purposes of the current model, we made the simplifying assumption that speech sounds comprise syllables and individual phonemes.

[2] An analogous division of a linguistic plan into frames and content can easily be envisaged at higher linguistic levels, but treatment of such issues is beyond the scope of this article.

[3] This notation is used throughout to indicate syllable type. C indicates consonant, V vowel. For example, a CCCV syllable (e.g., "stra") is composed of three consonants followed by a vowel.

[4] The current model treats successive consonants in an onset or coda cluster as independent; however, error data support the notion that consonant clusters may be additionally bound to one another (though not completely). Future work will elaborate the syllable frame specification to account for such data.

[5] Due to the phonotactic rules of English, not all phonemes are eligible at all positions. For simplicity, this notion was not explicitly incorporated in the model, but its implications suggest future work.

[6] The term *word* is used loosely to indicate a portion of a planned utterance that is at least as large as a syllable. This could represent a real word, a morpheme, or a pseudoword, for example.

[7] Here, *generative* is used to mean that, for the behavior in question, the generation of novel and perhaps arbitrary sequences is crucial. In speech, combining words or syllables into novel sequences is commonplace.

REFERENCES

Abeles, M. (1991). *Corticonics—Neural circuits of the cerebral cortex*. Cambridge, UK: Cambridge University Press.

Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception: Psycholinguistic and neurobiological perspectives. *Trends in Neurosciences, 31,* 265–272.

Agam, Y., Bullock, D., & Sekuler, R. (2005). Imitating unfamiliar sequences of connected linear motions. *Journal of Neurophysiology, 94,* 2832–2843.

Alario, F. X., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research, 1076,* 129–143.

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, and Computers, 36,* 140–155.

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends in Neurosciences, 13,* 266–271.

Alexander, G. E., DeLong, M. R., & Strick, K. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience, 9,* 357–381.

Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences, U.S.A., 99,* 13172–13177.

Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2003). Neural activity in prefrontal cortex during copying geometrical shapes: I. Single cells encode shape, sequence, and metric parameters. *Experimental Brain Research, 150,* 127–141.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (release 2) [CD-ROM]*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Beiser, D., & Houk, J. (1998). Model of corical-basal ganglionic processing: Encoding the serial order of sensory events. *Journal of Neurophysiology, 79,* 3168–3188.

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage, 32,* 821–841.

Bradski, G., Carpenter, G. A., & Grossberg, S. (1994). STORE working memory networks for storage and recall of arbitrary temporal sequences. *Biological Cybernetics, 71,* 469–480.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107,* 127–181.

Brown, J. W., Bullock, D., & Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks, 17,* 471–510.

Bullock, D. (2004). Adaptive neural models of queuing and timing in fluent action. *Trends in Cognitive Science, 8,* 426–433.

Bullock, D., & Rhodes, B. (2003). Competitive queuing for serial planning and performance. In M. Arbib (Ed.), *Handbook of brain theory and neural networks* (2nd ed., pp. 241–244). Cambridge, MA: MIT Press.

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31,* 429–460.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106,* 551–581.

Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience, 12,* 679–690.

Carreiras, M., & Perea, M. (2004). Naming pseudowords in Spanish: Effects of syllable frequency. *Brain and Language, 90,* 393–400.

Chein, J. M., & Fiez, J. A. (2001). Dissociation of verbal working memory system components using a delayed serial recall task. *Cerebral Cortex, 11,* 1003–1014.

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition, 99,* 205–235.

Cisek, P., & Kalaska, J. F. (2002). Simultaneous encoding of multiple potential reach direction in dorsal premotor cortex. *Journal of Neurophysiology, 87,* 1149–1154.

Clower, W. T., & Alexander, G. E. (1998). Movement sequence-related activity reflecting numerical order of components in supplementary and presupplementary motor areas. *Journal of Neurophysiology, 80,* 1562–1566.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–185.

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T., et al. (2006). Language control in the bilingual brain. *Science, 312,* 1537–1540.

Crompton, A. (1982). Syllables and segments in speech production. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 109–162). Berlin: Mouton.

Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Deniau, J. M., & Chevalier, G. (1985). Disinhibition as a basic process in the expression of striatal functions: II. The striato-nigral influence on thalamocortical cells of the ventromedial thalamic nucleus. *Brain Research, 334,* 227–233.

Elman, J. L. (1990). Finding structure in time. *Cognition, 14,* 179–211.

Farrell, S., & Lewandowsky, S. (2004). Modelling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language, 51,* 115–135.

Fudge, E. C. (1969). Syllables. *Journal of Linguistics, 5,* 226–320.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 133–177). New York: Academic Press.

Gelfand, J. R., & Bookheimer, S. Y. (2003). Dissociating neural mechanisms of temporal sequencing and processing phonemes. *Neuron, 38,* 831–842.

Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language, 17,* 171–183.

Graveland, G. A. (1985). A Golgi study of the human neostriatum: Neurons and afferent fibers. *Journal of Comparative Neurology, 234,* 317–333.

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics, 52,* 213–257.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23,* 121–134.

Grossberg, S. (1978a). Behavioral contrast in short term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology, 17,* 199–219.

Grossberg, S. (1978b). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (pp. 233–374). New York: Academic Press.

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics, 72,* 43–53.

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

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders, 39,* 350–365.

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,* 280–301.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105,* 611–633.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001a). A computational model of action selection in the basal ganglia I. A new functional anatomy. *Biological Cybernetics, 84,* 401–410.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001b). A computational model of action selection in the basal ganglia: II. Analysis and simulation of behavior. *Biological Cybernetics, 84,* 411–423.

Harrington, D. L., & Haaland, K. Y. (1998). Sequencing and timing operations of the basal ganglia. In D. A. Rosenbaum & C. E. Collyer (Eds.), *Timing of behavior: Neural, psychological, and computational perspectives* (pp. 35–61). Cambridge, MA: MIT Press.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35,* 1–31.

Henson, R. N. (1998). Short-term memory for serial order: The Start-End Model. *Cognitive Psychology, 36,* 73–137.

Henson, R. N., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology, 49A,* 80–115.

Hikosaka, O., & Wurtz, R. H. (1989). The basal ganglia. In R. H. Wurtz & M. E. Goldberg (Eds.), *The neurobiology of saccadic eye movements* (pp. 257–281). Amsterdam: Elsevier.

Hillis, A. E., Work, M., Barker, P. B., Jacobs, M. A., Breese, E. L., & Maurer, K. (2004). Re-examining the brain regions crucial for orchestrating speech articulation. *Brain, 127,* 1479–1487.

Ho, A. K., Bradshaw, J. L., Cunnington, R., Phillips, J. G., & Iansek, R. (1998). Sequence heterogeneity in Parkinsonian speech. *Brain and Language, 64,* 122–145.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology, 117,* 500–544.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In *Current research in natural language generation* (pp. 287–319).

Houghton, G., & Hartley, T. (1995). Parallel models of serial behaviour: Lashley revisited. *Psyche, 2*.

Jaeger, D., Kita, H., & Wilson, C. J. (1994). Surround inhibition among projection neurons is weak or nonresistant in the rat neostriatum. *Journal of Neurophysiology, 72,* 2555–2558.

Jonas, S. (1981). The supplementary motor region and speech emission. *Journal of Communication Disorders, 14,* 349–373.

Jonas, S. (1987). The supplementary motor region and speech. In E. Perceman (Ed.), *The frontal lobes revisited* (pp. 241–250). New York: The IREN Press.

Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach*. La Jolla, CA: University of California, San Diego.

Kawaguchi, Y. (1993). Physiological, morphological, and histochemical characterization of three classes of interneurons in rat neostriatum. *Journal of Neuroscience, 10,* 3421–3438.

Kemp, J. M., & Powell, T. P. S. (1971). The structure of the caudate nucleus of the cat: Light and electron microscopy. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 262,* 383–401.

Kent, R. D. (2000). Research on speech motor control and its disorders: A review and prospective. *Journal of Communication Disorders, 33,* 391–427.

Kreisler, A., Godefrey, O., Delmaire, C., Debachy, B., Leclercq, M., Pruvo, J.-P., et al. (2000). The anatomy of aphasia revisited. *Neurology, 54,* 1117–1123.

Kropotov, J. D., & Etlinger, S. C. (1999). Selection of actions in the basal ganglia-thalamocortical circuits: Review and model. *International Journal of Psychophysiology, 31,* 197–217.

Laganaro, M., & Alario, F.-X. (2006). On the locus of the syllable frequency effect in speech production. *Journal of Memory and Language, 55,* 178–196.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38.

Levelt, W. J., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50,* 239–269.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lewandowsky, S., Brown, G. D. A., Wright, T., & Nimmo, L. M. (2006). Timeless memory: Evidence against temporal distinctiveness models of short term memory for serial order. *Journal of Memory and Language, 54,* 20–38.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review, 96,* 25–57.

Locke, J. L. (1997). A theory of neurolinguistic development. *Brain and Language, 58,* 265–326.

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

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

MacNeilage, P. F., & Davis, B. (1990). Acquisition of speech production: Frames, then content. In M. Jeannerod (Ed.), *Attention and performance: XIII. Motor representation and control* (pp. 453–476). Hillsdale: Erlbaum.

Markram, H., & Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature, 382,* 807–810.

Matsuzaka, Y., Aizawa, H., & Tanji, J. (1992). A motor area rostral to the supplementary motor area (presupplementary motor area) in the monkey: Neuronal activity during a learned motor task. *Journal of Neurophysiology, 68,* 653–662.

McNeil, M. R., Pratt, S. R., & Fossett, T. R. D. (2004). The differential diagnosis of apraxia. In B. Maassen, R. Kent, H. Peters, P. van Lieshout, & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 389–413). New York: Oxford University Press.

Meijer, P. J. A. (1996). Suprasegmental structures in phonological encoding: The CV structure. *Journal of Memory and Language, 35,* 840–853.

Middleton, F. A., & Strick, P. L. (2000). Basal ganglia and cerebellar loops: Motor and cognitive circuits. *Brain Research Reviews, 31,* 236–250.

Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology, 50,* 381–425.

Mink, J. W., & Thach, W. T. (1993). Basal ganglia intrinsic circuits and their role in behavior. *Current Opinion in Neurobiology, 3,* 950–957.

Murdoch, B. E. (2001). Subcortical brain mechanisms in speech and language. *Folia Phoniatrica et Logopaedica, 53,* 233–251.

Nadeau, S. E. (2001). Phonology: A review and proposals from a connectionist perspective. *Brain and Language, 79,* 511–579.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761–781.

Pai, M. C. (1999). Supplementary motor area aphasia: A case report. *Clinical Neurology and Neurosurgery, 101,* 29–32.

Papoutsi, M., de Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., & Horwitz, B. (2009). From phonemes to articulatory codes: An fMRI study of the role of Broca's area in speech production. *Cerebral Cortex*.

Parent, A., & Hazrati, L.-N. (1995). Functional anatomy of the basal ganglia: I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Research Reviews, 20,* 91–127.

Picard, N., & Strick, P. L. (1996). Motor areas of the medial wall: A review of their location and functional activation. *Cerebral Cortex, 6,* 342–353.

Pickett, E. R., Kuniholm, E., Protopapas, A., Friedman, J., & Lieberman, P. (1998). Selective speech motor, syntax and cognitive deficits associated with bilateral damage to the putamen and the head of the caudate nuclues: A case study. *Neuropsychologia, 36,* 173–188.

Plenz, D., & Kitai, S. T. (1998). Up and down states in striatal medium spiny neurons simultaneously recorded with spontaneous activity in fast-spiking interneurons studied in cortex-striatum-substantia nigra organotypic cultures. *Journal of Neuroscience, 18,* 266–283.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage, 10,* 15–35.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences, 22,* 253–336.

Pulvermüller, F. (2002). A brain perspective on language mechanisms: From discrete neuronal ensembles to serial order. *Progress in Neurobiology, 67,* 85–111.

Redgrave, P., Prescott, T., & Gurney, K. (1999). The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience, 89,* 1009–1023.

Rhodes, B. J., & Bullock, D. (2002). *Neural dynamics of learning and performance of fixed sequences: Latency pattern reorganization and the N-STREAMS model*. Boston: Boston University.

Rhodes, B. J., Bullock, D., Verwey, W. B., Averbeck, B. B., & Page, M. P. (2004). Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives. *Human Movement Science, 23,* 699–746.

Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neurosciences, 21,* 188–194.

Robles, S. G., Gatignol, P., Capelle, L., Mitchell, M.-C., & Duffau, H. (2005). The role of dominant striatum in language: A study using intraoperative electrical stimulations. *Journal of Neurology, Neurosurgery, and Psychiatry, 76,* 940–946.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64,* 249–284.

Rogers, M. A., & Spencer, K. A. (2001). Spoken word production without assembly: Is it possible? *Aphasiology, 15,* 68–74.

Rogers, M. A., & Storkel, H. L. (1998). Reprogramming phonologically similar utterances: The role of phonetic features in pre-motor encoding. *Journal of Speech, Language, and Hearing Research, 41,* 258–274.

Schönle, P. W., Hong, G., Benecke, R., & Conrad, B. (1986). Aspects of speech motor control: Programming of repetitive versus non-repetitive speech. *Neuroscience Letters, 63,* 170–174.

Sevald, C. A., Dell, G. S., & Cole, J. S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language, 34,* 807–820.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In W. E. Cooper & E. C. T Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Erlbaum.

Shima, K., & Tanji, J. (2000). Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. *Journal of Neurophysiology, 84,* 2148–2160.

Spencer, K. A., & Slocomb, D. L. (2007). The neural basis of ataxic dysarthria. *Cerebellum, 6,* 58–65.

Tepper, J. M., Koos, T., & Wilson, C. J. (2004). GABAergic microcircuits in the neostriatum. *Trends in Neurosciences, 27,* 662–669.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 145–152.

Van der Merwe, A. (1997). A theoretical framework for the characterization of pathological speech sensorimotor control. In M. R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders* (pp. 1–25). New Yok: Thieme.

Vargha-Khadem, F., Watkins, K. E., Price, C. J., Ashburner, J., Alcock, K. J., Connelly, A., et al. (1998). Neural basis of an inherited speech and language disorder. *Proceedings of the National Academy of Sciences, U.S.A., 95,* 12695–12700.

Varley, R., & Whiteside, S. P. (2001). What is the underlying impairment in acquired apraxia of speech? *Aphasiology, 15,* 39–49.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik, 14,* 85–100.

Vousden, J. I., Brown, D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology, 41,* 101–175.

Ward, N. (1994). *A connectionist language generator*. Norwood, NJ: Ablex Publishing.

Warren, H. (1986). Slips of the tongue in very young children. *Journal of Psycholinguistic Research, 15,* 309–344.

Watkins, K. E., Vargha-Khadem, F., Ashburner, J., Passingham, R. E., Connelly, A., Friston, K. J., et al. (2002). MRI analysis of an inherited speech and language disorder: Structural brain abnormalities. *Brain, 125,* 465–478.

Werker, J. F., & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological acquisition: Models, research, implications* (pp. 285–311). Timonium, MD: York Press.

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

Wilson, C. J. (1995). The contribution of cortical neurons to the firing patterns of striatal spiny neurons. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 29–50). Cambridge, MA: MIT Press.

Ziegler, W. (2002). Psycholinguistic and motor theories of apraxia of speech. *Seminars in Speech and Language, 23,* 231–243.

Ziegler, W., Kilian, B., & Deger, K. (1997). The role of the left mesial frontal cortex in fluent speech: Evidence from a case of left supplementary motor area hemorrhage. *Neuropsychologia, 35,* 1197–1208.