HHS Public Access

Author manuscript

Nat Rev Neurosci. Author manuscript; available in PMC 2017 March 27.

Published in final edited form as:

Nat Rev Neurosci. ; 13(2): 135–145. doi:10.1038/nrn3158.

Computational neuroanatomy of speech production

Gregory Hickok

Department of Cognitive Sciences, University of California, Irvine, California, 92697, USA

Abstract

Speech production has been studied predominantly from within two traditions, psycholinguistics and motor control. These traditions have rarely interacted and the resulting chasm between these approaches seems to reflect a level of analysis difference: while motor control is concerned with lower-level articulatory control, psycholinguistics focuses on higher-level linguistic processing. However, closer examination of both approaches reveals a substantial convergence of ideas. The goal of this article is to integrate psycholinguistic and motor control approaches to speech production. The result of this synthesis is a neuroanatomically grounded hierarchical state feedback control model of speech production.

Most research on speech production has been conducted from within two different traditions: a psycholinguistic tradition that seeks generalizations at the level of phonemes, morphemes and phrasal units,[1–4] and a motor control tradition that is more concerned with kinematic forces, movement trajectories and feedback control[5–7]. Despite their common goal, to understand how speech is produced, little interaction has occurred between these traditions. The reason for this disconnect seems fairly clear: the two approaches are focused on different levels of the speech production problem, with the psycholinguists working at a more abstract, perhaps even amodal level of analysis and the motor control scientists largely examining lower-level articulatory control processes. The question posed here is whether the level-driven chasm between these two traditions reflects a real distinction in the systems underlying speech production, such that the vocabularies, architectures and computations that are associated with the respective traditions are necessarily different, or whether the chasm is a vestige of the history of the two approaches. I will suggest that the disconnect is more apparent than it is real and, more importantly, that both approaches have much to gain by paying attention to each other.

The article begins with an introduction to the motor control perspective of speech production through highlighting a fundamental engineering problem in motor control and how internal models solve this problem. The next section briefly summarizes psycholinguistic approaches to speech production, and points out some similarities and differences between these approaches and those from the motor control perspective. The core of the article outlines a hierarchical state feedback control model of speech production that incorporates components from both traditions and data from recent neuroscience research on sensorimotor integration. This model is based on the assumption that sensory representations in both auditory and somatosensory cortex define a hierarchy of targets for speech gestures. In this model, auditory targets are predominantly syllabic and comprise higher-level sensory goals, whereas somatosensory targets represent lower-level goals that correspond loosely to

phonemic-level targets. Movement plans that are coded in a corresponding cortical motor hierarchy are selected to hit the sensory targets. This selection process involves an internal feedback control loop (involving forward prediction and correction) that is integral to the motor selection process rather than serving to evaluate and correct motor execution errors. Sensorimotor integration (that is, coordinate transform) is achieved in the temporal–parietal cortex (in area Spt) for the higher-level system and via the cerebellum for the lower-level circuit. A simple simulation of one aspect of the model is presented to demonstrate the feasibility of the proposed architecture and computational assumptions.

Motor control and internal models

Sensory feedback is a critical component of motor control, yet the delay in this feedback presents an engineering problem that can be illustrated by considering the following hypothetical task. Imagine driving a car on a racetrack while only looking in the rear view mirror. From this perspective, it is possible to determine whether the car is on the track and pointed roughly in the right direction. It is also possible to drive the track reasonably successfully under one of two conditions: the track is perfectly straight or you drive extremely slowly, inching forward, checking the car's position, making a correction, and inching forward again. It might be possible to learn to drive the track more quickly after considerable practice; that is, by learning to predict when to turn on the basis of landmarks that you can see in the mirror. However, you will never win a race against someone who can look out of the front window, and an unexpected event such as an obstacle in the road ahead could prove catastrophic. The reason for these outcomes is obvious: the rear view mirror can only provide direct information about where you have been, not where you are or what is coming in the future.

Motor control presents the nervous system with precisely the same problem[8, 9]. As we reach for a cup, we receive visual and somatosensory feedback. However, as a result of neural transmission and processing delays, which can be significant, by the time the brain can determine the position of the arm based on sensory feedback, it is no longer at that position. This discrepancy between the actual and directly perceived state of the arm is not much of an issue if the movement is highly practiced and is on target. If a correction to a movement is needed, however, the nervous system has a problem because the required correctional forces are dependent on the position of the limb at the time of the arrival of the correction signal — that is, in the future. Sensory feedback alone cannot support such a correction efficiently. As with the car analogy, one way to get around this problem is to execute only very slow, piecemeal movements. The CNS, however, clearly does not adopt this strategy. Rather, it favours a solution that involves looking out of the 'front window' or, in motor control terms, comprises generating an internal forward model that can make accurate predictions regarding the current and future states of motor effectors.

Recent models of motor control circuits incorporate such a forward-looking component (Figure 1)[9]. These circuits include a motor controller that sends signals to an effector (often called the 'plant') and a sensory system that can detect changes in the state of the effector and other sensory consequences of the action. A key additional component of these circuits is the so-called internal forward model, which receives a corollary discharge or efference

copy of the motor command that is issued to the motor effector. The internal forward model allows the circuit to make predictions regarding the current state of the effector (that is, its position and trajectory) and the sensory consequences of a movement. Thus, these recently proposed motor control circuits have both a mechanism that allows the brain to 'look out the rear view mirror' and measure the actual sensory consequences of an action and an internal mechanism to look forward and make predictions regarding the probable consequences of a programmed movement. Both mechanisms are critical for effective motor control. The internal forward-looking mechanism is particularly useful for online movement control because the effects of a movement command can be evaluated for accuracy and potentially corrected before overt sensory feedback. By contrast, external feedback is critical for three purposes: to learn the relationship between motor commands and their sensory consequences in the first place (that is, to learn the internal model); to update the internal model in case of persistent mismatches (errors) between the predicted and measured states owing to system drift or shifting sensory–motor conditions (such as during motor fatigue, switching from a light to a heavy tool, or donning prism goggles); and to detect and correct for sudden perturbations (for example, getting bumped in the middle of a movement). In many cases, the two sources of feedback work together such as when a perturbation is detected via sensory feedback and a correction signal is generated using internal forward predictions of the state of the effector. Motor control models with these feedback properties are often referred to as state feedback control models because feedback from the predicted (internal) state as well as the measured state of the plant is used as input to the controller.

The inclusion of internal forward prediction in motor control circuits as a source of state feedback control provides a solution to the engineering problem outlined above and the existence of such internal models in state feedback control has been supported experimentally[10–12]. For these reasons, the state feedback control approach has been highly influential and widely accepted within the visuomotor domain[8, 13, 14]. Feedback control generally, as opposed to internal feedback control specifically, has also been empirically demonstrated in the speech domain using overt sensory feedback alteration paradigms and other approaches[15]. This work has shown that when speaking, people adjust their speech output to compensate for sensory feedback 'errors' (experimentally induced shifts) in both the auditory[6, 16–18] and somatosensory systems[19]. Evidence for internal state feedback is less prevalent in motor speech control than in the visuomotor domain. However, if one looks outside of the motor control tradition, strong evidence can be found for the existence of internal state feedback control in speech production (see below).

Of particular interest to the present discussion is the suggestion in the visuomotor literature that state feedback models for motor control are hierarchically organized[20–22]. The concept of a sensorimotor hierarchy has a long history[23] and is well accepted. Application of this notion to state feedback models of motor control, including those of speech[24], is therefore a natural extension of existing motor control models. Indeed, if we introduce the notion of a hierarchy of internal feedback control, then hierarchical motor control models of speech production begin to overlap with hierarchical linguistic models of speech production; that is, the traditions begin to merge.

The psycholinguistic perspective

Psycholinguists have traditionally been concerned with higher-level aspects of the speech production process, specifically, the nature of the speech planning units and the processing steps involved in transforming a thought into a speech act[4]. As such, psycholinguistic speech production models typically start with a conceptual or message level representation and end with a phonological or phonetic representation (that is, the output) that feeds into the motor control system. Thus, phonological representations are considered abstract representations that are distinct from motor control structures in most, but not all[25, 26], psycholinguistic or linguistic models of speech production.

For the present purposes, it is worth highlighting two important points of consensus that have emerged from the psycholinguistic research tradition. One is that speech production is planned across multiple hierarchically organized levels of analysis that span phonetic–phonological, morphological and phrase-level units[2, 4, 27–29]. Such planning is consistent with (and perhaps predictable from) not only the observation that language structure is strongly hierarchical but also the notion that motor control circuits are hierarchically organized. The second point of consensus is that word production involves at least two stages of processing: a lexical (or 'lemma') level and a phonological level[1, 3, 30].

In typical word-level psycholinguistic models of speech production[1, 31] (FIG. 2), input to the system comes from the conceptual system; that is, the particular concept or message that the speaker wishes to express. The concept is mapped onto a corresponding lexical item, often referred to as a lemma representation, which codes abstract word properties such as a word's grammatical features but does not code a word's phonological form. Phonological information is coded at the next level of processing. Evidence for such a two-stage model comes from a variety of sources including the distribution of speech error types[2, 4, 28], chronometric studies of interference in picture naming[29], tip-of-the-tongue phenomena[32] and speech disruption patterns in patients with aphasia[30].

Interestingly, feedback correction mechanisms — including both internal and external (overt) feedback monitoring loops — have been proposed to form part of psycholinguistic models of speech production[33]. That external feedback is monitored and used for error correction is evident in everyday experience when the occasional misspoken word or phrase is noticed by a speaker and is corrected. The timing of such error detection in some cases reveals that internal error detection is also operating. For example, Nozari *et al.* point out that documented error corrections such as "v-horizontal" (incorrectly starting to utter "vertical" with subsequent correction) occur too rapidly to be carried out by an external feedback mechanism[34]. Within the psycholinguistic tradition, the nature of the internal and external feedback correction mechanisms in speech production has received increasing empirical and theoretical attention over the past two decades[34–38], including the suggestion that error detection and correction in speech may not rely on sensory systems[39], a notion that is not consistent with assumptions in the motor control literature.

Integrating the traditions

In this section, I start with the assumptions that speech production is fundamentally a motor control problem and that motor control is hierarchically organized. Thus, the engineering problems that exist at one level also hold for other levels in the hierarchy. In other words, there is no fundamental distinction between the problems and solutions at different levels of analysis in speech production. What has been learned about motor control at lower levels (for example, internal forward models) can, and should, be applied to the problems at higher levels and *vice versa*. Thus, when thinking about how, say, phonological forms are accessed we need to consider forward prediction as part of the process. Likewise, when thinking about control architectures for speech motor control, we need to consider the hierarchical structure of the system as a whole, as revealed by linguistic approaches. At this point, I would also like to note another source of constraint on the development of a model of speech production, namely neuroscience. A fair amount of information is now available regarding the neural circuits that are involved in motor control generally[9, 13] and speech motor control more specifically[40, 41]. This information also needs to be integrated in any new model.

My colleagues and I have already sketched a first attempt at an integration of the psycholinguistic and motor control literatures[41]. Here, I will briefly review that model as a starting point.

An integrated state feedback control model

The integrated state feedback control (SFC) model of speech production[41] (FIG. 3) builds on models proposed by Guenther and colleagues[5], Tian and Poeppel[12] and Houde and Nagarajan[42]. Consistent with state feedback control models generally, the SFC model includes a corollary discharge to an internal model of the state of the motor effector (the vocal tract), which in turn generates forward predictions of the sensory consequences of the motor effector states. It also incorporates the two-stage model of speech production from psycholinguistics, a lexical–conceptual level and a phonological level. It further includes a translation component, labeled auditory–motor translation, that is assumed to compute a coordinate transform between auditory and motor space, which is a concept that comes out of the neuroscience literature[41, 43, 44].

The SFC model diverges from typical ‘within tradition’ assumptions in several respects as a result of its integrated design. In contrast to the typical low-level focus of motor control models, this model includes a higher-level circuit involving phonological representations, a key level of processing in psycholinguistic models. Unlike some psycholinguistic models[1], however, the phonological level is split into two components, a motor and a sensory phonological system. There has been some discussion within the psycholinguistic tradition of distinctions within the phonological system[3], including a distinction (in neuropsychological theories) between a sensory or input component and a motor or output component[45–47]. The latter distinction fits well with feedback control architectures for speech, which include an internal model of the motor effector and a separate system that codes the targets in auditory space[5, 48, 49]. The idea that the lexical–conceptual system sends parallel inputs to the sensory and motor components of the system is not characteristic of either the motor control or psycholinguistic traditions, although the idea does have roots in

classical 19th century models of the neural organization of language[50, 51] and provides an explanation for certain forms of language disruption following brain injury (BOX 1).

Box 1

Conduction aphasia: a sensorimotor deficit

One major empirical benefit of the integrated state feedback control (SFC) model is its ability to explain the central features of conduction aphasia. People with such aphasia have fluent speech yet produce relatively frequent and predominantly phonemic speech errors (paraphasias) that they often detect and attempt to correct, mostly unsuccessfully. Although speech perception and auditory comprehension at the word and conversational level are well preserved in such individuals, verbatim repetition is impaired, particularly for complex phonological forms and non-words[86]. Reconciliation of the co-occurrence of these features — that is, generally fluent output, impaired phonemic planning, and preserved speech perception — has proved difficult. A central phonological deficit could yield phonemic output problems but would also be expected to affect perception. Alternatively, assuming that separate phonological input and output systems exist, impairment to a phonological output system could explain the paraphasias but should also cause dysfluency. Furthermore, the lesions in conduction aphasia are in auditory-related temporal–parietal cortex [83–85], not in frontal cortex where one would expect to find motor-related systems. Damage to a phonological input system is more consistent with the lesion location, explains the preserved fluency because the motor phonological system is still intact, and could explain paraphasias if one assumes a role for the input system in speech production. However, again there is no explanation for why the system can easily recognize errors perceptually that it fails to prevent in production.

Wernicke's original hypothesis that conduction aphasia is a disconnection between sensory and motor speech systems is a viable solution[117]: fluency is preserved because the motor system is intact, perception is preserved because the sensory system is intact, and paraphasias occur because the sensory system can no longer play its role in speech production once the systems are disconnected (see also [46] for similar arguments). What was lacking from Wernicke's account, though, was a principled explanation for why the sensory system plays a role in production. Internal feedback control (as included in the SFC model) provides such a principled explanation: the sensory speech system is involved in production because the sensory system defines the targets of speech actions, and without access to information about the targets, actions will sometimes miss their mark; this is especially true for actions that are not highly automated (complex phonological forms) or are novel (non-words). The only other modern adjustment that is needed to Wernicke's account is the anatomy. He proposed a white matter tract as the source of the disconnection, for which there is little evidence[118–120]. Modern findings instead implicate a cortical system that computes a sensorimotor coordinate transformation[43, 44, 84, 90]. In short, the integrated SFC model improves our understanding of conduction aphasia[41].

Extending the model

Here, I will outline an extension of the integrated SFC model, which will be referred to simply as the hierarchical state feedback control (HSFC) model (FIG. 4). In the HSFC model there are two hierarchically organized levels of state feedback control, which are similar to those proposed by Gracco and Lofqvist[24, 52]. The higher level codes speech information predominantly at the syllable level (that is, vocal tract opening and closing cycles) and involves a sensory–motor loop that includes sensory targets in auditory cortex, motor programs coded in the Brodmann area (BA) 44 portion of Broca's area and/or lower BA6, and area Spt, which computes a coordinate transform between the sensory and motor areas. This is the loop described in the earlier SFC model[41] that was discussed in the previous section. The lower level of feedback control codes speech information at the level of articulatory feature clusters; that is, the collection of feature values that are associated with the targets of a vocal tract opening or closing. These feature clusters roughly correspond to phonemes[24, 52] and involve a sensory–motor loop that includes sensory targets coded primarily in somatosensory cortex (as suggested by V. Gracco, personal communication), motor programs coded in lower primary motor cortex (M1), and a cerebellar circuit mediating the relation between the two. The inclusion of auditory and somatosensory targets and a cerebellar loop is not unique to this proposed model: Guenther and colleagues' directions into velocities of articulators (DIVA) model also includes these components[5, 40]. The DIVA model, however, does not make use of an internal feedback control system (control is achieved using overt feedback) and does not distinguish hierarchically organized levels.

Sensory targets

Convincing arguments regarding auditory targets for speech gestures have been made previously[5, 15, 53]. Here, I will only add the point that activation of an auditory speech form, whether internally or externally, seems to automatically define a potential target for action and consequently excites a corresponding motor program, regardless of whether there is an intention to speak. This assertion is based on the observation that the perception of others' speech activates motor speech systems[54, 55] and that the speech of others, even if it is ambient, can be unintentionally imitated by a listener or speaker[56–58]. The existence of echolalia, the tendency of individuals with certain acquired or developmental speech disorders to repeat heard speech[59, 60], provides additional evidence for this assertion in that it suggests that an underlying, almost reflexive sensory-to-motor activation loop exists. In the normal brain, listening to speech and the consequent activation of the sensory-to-motor circuit does not normally result in motor execution (and hence repetition of heard speech) presumably because motor selection mechanisms inhibit this behaviour at some level. Echolalia seems to be induced by an abnormal release from inhibition of this motor selection system.

Regarding somatosensory targets, clear evidence exists that the somatosensory system has an important role in speech production. Just as speakers adapt to altered auditory feedback, they also correct for unexpected mechanical alterations of the jaw (somatosensory feedback) even when there are no acoustic consequences associated with the alteration[19]. Furthermore, transient or permanent disruption of lingual nerve feedback has been found to affect speech

articulation even for phonemes with clear acoustic consequences (vowels and sibilants)[61–63]. For these reasons, motor–somatosensory loops are prominent components of motor control models for speech[5, 40, 64].

The logic behind the idea that articulatory feature clusters (roughly equating to phonemes) are defined predominantly in terms of somatosensory rather than auditory targets requires justification. If we think of speech production as a cycle of opening and closing of the vocal tract, we can then view phonemes as the articulatory configurations that are defined by the endpoints (the open or closed positions) of each half cycle of movement. In other words, the feature clusters that define phonemes represent the articulatory features at closed (consonants) or open (vowels) positions. Owing to coarticulation, however, the acoustic consequences of the articulatory configurations that define phonemes are not restricted to and indeed often not apparent at the precise time point when these articulatory configurations are achieved. Put differently, the vocal tract configurations that define individual phoneme segments do not, in isolation, have reliably identifiable acoustic consequences (particularly for stop consonants). This inconsistency forms the basis of the lack of invariance problem in speech perception[65] (BOX 2). However, the vocal tract configurations that define phonemes, the endpoints of an articulatory half cycle, do have somatosensory consequences. Lip closure or tongue raising, for example, have detectable somatosensory consequences at the endpoint vocal tract configurations — the point in time that defines the phoneme — even in the absence of a clear auditory signature during that same time window. Thus, I hypothesize that the higher-level goal of a speech act is to hit an auditory target (roughly equating to syllable units), which can be defined as an articulatory cycle or half-cycle. This goal can be decomposed into subgoals, namely to hit somatosensory targets (roughly equating to articulatory feature clusters or phoneme units) at the endpoints of each half-cycle.

Box 2

The lack of invariance problem

The lack of invariance problem refers to the fact that there is not a one-to-one mapping between acoustic features and perceptual categorization of speech sounds. For example, the same phoneme, say /d/, can have different acoustic patterns in different syllable contexts, such as in /di/ and /da/[65]. This lack of invariance between acoustics and perception is arguably the fundamental problem in speech perception[65, 121, 122]. An early solution to this problem was the motor theory, which held that the target of speech perception is not acoustic but motor gestures[65, 122]. However, the idea that low-level articulatory plans form the basis of perception has been rejected on empirical grounds[123–125]. In response, variants of the model have been proposed in which the objects of speech perception are more abstract gestural goals[121, 126], but this idea is functionally indistinguishable from an auditory theory that assumes that the goals of speech gestures are sensory states.

Several other approaches have been taken to resolve the lack of invariance problem. These include the search for possibly overlooked acoustic features that do hold an invariant relation to phonemic categories[127], and a range of approaches that accept that a

variable acoustic–phonemic relationship exists but use various active processes such as motor prediction[128], normalization[129], or top-down lexical constraint[130] to circumvent the problem.

Another class of solutions, broadly consistent with the hierarchical state feedback control (HSFC) model, rejects the idea that the basic acoustic unit of speech perception is the phoneme and argues instead for a larger unit such as the syllable[125, 131–134]; that is, units that have more consistent acoustic consequences. Exemplar- or episodic-based approaches, which code acoustic patterns more broadly, are another class of models that resolve the lack of invariance problem by using the broader acoustic context to code speech representations[135–137]. However, the idea that the basic unit of speech perception is larger than the phoneme has met with resistance, as is evident from the fact that the dominant models of speech recognition include a phoneme-level component[130, 138, 139]. I suggest that some of the resistance comes from the assumption that by doing away with the phoneme in speech recognition, one must do away with the phoneme (or feature clusters) altogether, which flies in the face of decades of research on phonology. The present conceptualization (that is, the HSFC model) accommodates the idea of syllable-based auditory speech recognition, yet retains the phoneme, albeit predominantly at a lower (somatosensory) level in the speech sensory–motor hierarchy, which is less involved in speech recognition.

I have roughly equated lower-level somatosensory targets with phoneme units and higher-level auditory targets with syllable units. It is important to note that this is only an approximate alignment. Isolated phonemes on their own can have acoustic consequences (such as, fricatives, liquids and sibilants) and vowels are both phonemes and syllabic nuclei, so some segments can have both auditory and somatosensory targets, with different weightings depending on the particular segment involved[53, 64]. Given these considerations, phonemes and syllables may be distributed, in partially overlapping fashion, across the two hierarchical levels of motor control that are proposed above. The relevant generalization here, however, is not over linguistic units. Rather, the generalization is over control units, with the somatosensory system driving lower-level online control of vocal tract trajectories that target the endpoint of a vocal tract opening or closing, and the auditory system driving higher-level control of the cycles and half-cycles themselves.

In the DIVA model, auditory goals have primacy during learning; somatosensory correlates are learned later and form another source of control for speech gestures[53]. This order of events seems reasonable given that the ultimate goals of speech production are to reproduce the sounds in one's linguistic environment. Another way to think about auditory and somatosensory control circuits, consistent with the present HSFC model, is that the auditory goals comprise the broad, context-free target space whereas the somatosensory goals are used for fine tuning the movement in particular phonetic contexts. Such thinking is consistent with Levelt and colleagues' notion that phonological code access precedes and is separate from both 'syllabification' and 'phonetic encoding', processes that are context dependent[3]. A large-scale meta-analysis that aimed to localize the neural correlates of these psycholinguistic levels identified posterior temporal lobe regions as being involved in

phonological code retrieval and frontal areas as being involved in syllabification and articulatory processes[66], consistent with the model proposed here. The idea that auditory goals are broadly tuned, with somatosensory goals filling in the fine, context-dependent details, is also consistent with recent suggestions in the manual control literature that actions are selected on the basis of a "motor vocabulary"[67] and then fine-tuned to particular situations, which can vary in terms of muscle fatigue, mechanical loads, obstacles and so on[13].

Ventral premotor cortex and motor vocabularies—Ventral premotor cortex has been implicated in motor vocabularies in both speech and manual gestures[13, 40, 49, 68, 69]. As noted above, Levelt and colleagues' notion of a mental syllabary — a repository of gestural scores for the most highly used syllables in a language[3] — has been linked to ventral premotor cortex in a large-scale meta-analysis of functional imaging studies[66]. A recent prospective functional MRI (fMRI) study that was designed to distinguish phonemic and syllable representations in motor codes provided further evidence for this view by demonstrating adaptation effects in ventral premotor cortex to repeating syllables[70].

Apraxia of speech (AOS) is a motor speech disorder that seems to affect the planning or coordination of speech at the level that has been argued to correspond to syllable-sized units[71, 72]. Although this conclusion should be regarded as being tentative, it is clear that AOS is not a low-level motor disorder such as dysarthria, which manifests as a consistent and predictable error (misarticulation) pattern in speech that is attributable to factors such as muscle weakness or tone. Rather, AOS is a higher-level disorder with a variable error pattern[73]. Ventral premotor cortex has been implicated in the aetiology of AOS[74], as has the nearby anterior insula[75, 76]. It is worth noting that speech errors in AOS and conduction aphasia (BOX 1) are often difficult to distinguish, the difference being most notable in speech fluency, with AOS resulting in more halting effortful speech[73]. The similarity in error type and the distinction in fluency between AOS and conduction aphasia is consistent with the present model if one assumes that the two disorders affect the same level of hierarchical motor control (errors occur at the same level of analysis) but in different components of the circuit (AOS affects access to motor phonological codes and conduction aphasia affects internal state feedback control).

In the visual–manual domain, physiological evidence from monkeys has suggested the existence of grasping-related motor vocabularies in ventral premotor cortex[68, 69]. Grafton has emphasized that such a motor vocabulary codes relatively higher motor programs — for example, correspondences between object geometry and grasp shape — that are then implemented via interactions with primary motor cortex[13]. This conceptualization is very similar to the present hierarchical model for speech actions.

Role of the cerebellum—In addition to the parietal cortex, the cerebellum has long been implicated in internal models of motor control including within the speech domain[18, 40, 77–80] and the cerebellum has been specifically implicated as being part of a forward model[81, 82]. The suggestion here is that parietal and cerebellar circuits are performing a similar sensory–motor coordinate transfer function but at different levels in the sensory–motor hierarchy (see[78] for a review of evidence for coordinate transform in the

cerebellar oculomotor system). Specifically, clinical evidence from the speech domain suggests that cortico-cortical circuits are involved in motor control at a higher (syllable) level whereas cerebellar–cortical circuits are controlling a lower (phonetic) level. For example, although lesions to cortical temporal–parietal structures are associated with phonological-level errors that are characteristic of conduction aphasia[83–86], cerebellar dysfunction results in a characteristic dysarthria comprising a slow down in speech tempo and a reduction in syllable duration variation (termed isochronous syllable pacing), characteristics that some authors have argued stem from a lengthening of short vocalic elements[87, 88], that is, those elements involving more rapid movements that may rely more on a finer-grained internal feedback control. Indeed, cerebellar dysarthria has been characterized as "compromised execution of single vocal tract gestures in terms of, presumably, an impaired ability to generate adequate muscular forces under time-critical conditions"[87].

Evidence for a sensory–motor hierarchy—Linguistic research over the past several decades has clearly shown that language is hierarchically organized and classic work on speech error analysis has shown that the speech production mechanism reflects this hierarchical organization[2, 4]. More recent behavioural evidence for a hierarchical organization for motor control circuits comes from studies of speech errors in internal (imagined) speech. Research on overt speech errors has shown that errors have a lexical bias (slips of the tongue tend to form words rather than nonwords) and exhibit a phonemic similarity effect (phonemes that share more articulatory features tend to interact more often in errors). Recent work has found that errors do occur and can be detected in internally generated speech[34]. Interestingly, the properties of internal errors vary depending on whether speech is imagined without silent articulation or with silent articulation. When speech is imagined without articulation — that is, when motor programs are not implemented — speech errors exhibit a lexical bias but do not show a phonemic similarity effect[34]. By contrast, when speech is silently articulated, both lexical and phonemic similarity effects are detectable[89]. These findings suggest that at least two levels of a control hierarchy exist, one at the level of phonemes (feature clusters) that is brought into play during actual articulation and the other at a higher phonemic level that functions even without overt motor action[89].

Imagined speech without articulation activates a network that includes posterior portions of Broca's area, dorsal premotor cortex, area Spt, and posterior superior temporal sulcus–superior temporal gyrus[43, 90, 91]. Studies that have directly contrasted fully imagined speech with silently articulated speech have reported greater involvement of primary motor and somatosensory cortex with articulated speech than with unarticulated speech[92, 93], consistent with the notion of hierarchically organized control circuits.

Interaction of auditory and somatosensory systems—The present suggestion that the sensory targets at the higher and lower hierarchical levels are auditory and somatosensory in nature, respectively, implies that these two sensory systems interact. Direct neurophysiological evidence for such an interaction has been found in both monkeys and humans. The caudal medial area of monkey auditory 'belt' cortex has been found to be a site

of auditory and somatosensory convergence[94, 95]. Electrophysiological [96, 97] and fMRI[98] data have confirmed similar auditory–somatosensory interaction in human auditory cortex in both hemispheres.

Most of the discussion regarding the functional role of auditory–somatosensory interaction has focused on perceptual modulation arising from phase resetting of neural oscillations in auditory cortex by somatosensory inputs[99]. Perceptual modulation is an important aspect of control circuits — forward predictions can be viewed as a form of perceptual modulation[41, 100–102] — and within the HSFC framework such a mechanism may allow somatosensory inputs to fine tune temporal aspects of forward auditory predictions. For example, activation of a syllable target in auditory cortex does not necessarily provide information about the timing (onset and rate) of articulation of that syllable. However, given that somatosensory targets correspond to vocal tract gesture endpoints, which define the phase of articulation, somatosensory-driven phase resetting in the auditory system may provide critical temporal information to auditory prediction. Auditory–somatosensory interaction presumably operates in the other direction as well, such as in the process of activating the appropriate somatosensory targets for a given auditory target. It is unclear whether the auditory regions described above that have been argued to support somatosensory influence on auditory perception also support auditory-to-somatosensory information flow. Nonetheless, the present model, as well as others such as the DIVA model[40, 53], predict an auditory–somatosensory interaction; such an interaction is consistent with available evidence.

Computational considerations—It is typically assumed that forward prediction is enabled via an efference copy of the motor command. In this conceptualization, the efference signal is ‘after the fact’ in the sense that it is a copy of a completed motor plan, implying that forward prediction plays no major role in initial motor planning. It is only when an error is detected that the efference copy results in a modulation of the motor command.

Here, I offer a different perspective in which efference signals and the resulting forward predictions are part of the motor planning process from the start. The concept is as follows. The auditory phonological system defines the motor target, which is activated via input from the lexical (lemma) level. The lemma also activates the associated motor phonological representation. The sensory and motor phonological systems then interact to ensure that the activated motor representation will indeed hit the auditory target. The activated auditory target activates the associated motor representation further reinforcing the motor activation. At the same time, the activated motor representation sends an inhibitory signal to the auditory target; the HSFC model’s equivalent of an efference signal. The assumption that this signal is inhibitory is consistent with other feedback control models[40, 48], and the logic here is that when there is a match between prediction and detection (that is, no corrections are necessary), the signals will roughly cancel each other out. Thus, in the present model, one can think of the excitatory sensory target-to-motor signal as a ‘correction’ signal that is turned on from the start. If no corrections are needed, the inhibitory motor-to-sensory ‘efference’ signal turns off the ‘correction’ signal. If, however, the wrong motor program is activated, it will then inhibit a non-target in the sensory system and therefore leave the

correction signal that is coming from the sensory target fully activated, which in turn will continue to work toward activating the correct motor representation. Thus, forward prediction and error correction in the HSFC model is part of the motor planning process. A small-scale simulation was carried out to assess the feasibility of both the basic architecture and the broad computational assumptions (FIG. 5).

Work in both human[100, 103, 104] and non-human primates[105] has shown that cortical auditory responses to self-vocalizations are attenuated compared with those responses to hearing a playback recording of the same sounds. This motor-induced suppression effect is consistent with the idea that a forward sensory prediction is instantiated as an inhibitory signal. The present simulation result that the auditory target is suppressed relative to baseline suggests that the architecture proposed here may provide a computational explanation for motor-induced suppression. This may also provide an explanation for why modulation of the motor system (for example, by transcranial magnetic stimulation) may affect speech perception[106], sometimes in highly specific ways [107]: motor activation can result in a modulation of sensory systems, thus potentially affecting perception[41, 108, 109].

Suppression of sensory target activity makes sense computationally for two reasons. One is to prevent interference with the next sensory target. In the context of connected speech, auditory phonological targets (syllables) need to be activated in a rapid series. Residual activation of a preceding phonological target may interfere with activation of a subsequent target if the former is not quickly suppressed. An inhibitory motor-to-sensory input provides a mechanism for achieving this.

The second benefit of target suppression is that it can enhance detection of off-target sensory feedback. Detection of deviation from the predicted sensory consequence of an action is a critical function of forward prediction mechanisms, as it allows the system to update the internal model. Recent work on selective attention has suggested that attentional gain signals that are applied to flanking or ‘off-target’ sensory features comprise a computationally effective and empirically supported mechanism to detect differences between targets and non-targets[110–113]. In the present context, target suppression would have the same functional consequence on detection as increasing the gain on flanking non-targets, namely, to increase the detectability of deviations from expectation.

The target suppression mechanism also resolves a noted problem in psycholinguistics concerning simultaneously monitoring both inner and external feedback by the same system given the time delay between the two[39, 114]. In the HSFC model, internal and external monitoring are just early and later phases, respectively, of the same mechanism. In the early, internal phase, errors in motor planning fail to inhibit the driving activation of the sensory representation, which acts as a ‘correction’ signal; in the later, external monitoring phase, the sensory representation is suppressed, consistent with some models of top-down sensory prediction[115, 116], which enhances detection of deviation from expectation; that is, the detection of errors.

In summary, the computational and architectural approach adopted here, specifically the idea that forward prediction is instantiated as an inhibitory input to sensory systems, achieves

several things with essentially one mechanism. First, it serves as part of a mechanism for internal error correction in cases where the wrong motor program is activated. Second, it serves to minimize interference between one target and the next during the production of a movement sequence. Third, it enhances the detection of deviation in overt sensory feedback from the predicted sensory consequences. Fourth, it provides an explanation for the motor-induced suppression effect. Last, it provides a mechanism for explaining the influence of the motor system on the perception of others' speech.

Conclusions

The goal of this article was to formulate a model of speech production that integrates theoretical constructs from linguistic and motor control perspectives and to link the model to a sketch of the underlying neural circuits. Although recognizable features exist in the model from the two research traditions, the framework is not merely a cut and paste job. Integration of the various ideas and data has led to some novel features (or at least novel combinations of ideas) including parallel activation of 'phonological' forms; a computational architecture that integrates motor selection, forward prediction, error detection, and error correction into one mechanism; and the idea that there is a rough correspondence between linguistic notions such as phoneme and syllable and motor control circuits involving somatosensory and auditory systems.

Despite whatever virtues the framework has, no doubt exists that it is an oversimplification and many important facts and ideas from all traditions have not been considered. For example, although I have presented the somatosensory and auditory systems as neatly separable hierarchical levels, the nature of their interaction between levels may be dramatically more complex. Correspondingly, the mapping between these levels and linguistic units such as phonemes and syllables is sure to be a nuanced one. Furthermore, it is clear that speech planning is not restricted to phoneme- and syllable-sized units and also includes words, phrases, intonation patterns and complexities such as morphological processes and syllabification. In addition, important circuits and brain regions — including the basal ganglia, supplementary motor area and right hemisphere motor-related areas — have been completely ignored despite the fact that they are surely involved in speech motor control. Nonetheless, I suggest that the exercise of attempting an integrated approach to modeling the dorsal stream speech production system has resulted in some novel, testable ideas that are worth pursuing in more detail and in this sense, the proposed framework hopefully serves its purpose.

Acknowledgments

I would like to thank J. Houde, H. Nusbaum, and D. Poeppel for comments on earlier drafts and sections of this paper, and also V. Gracco who inspired some of the key ideas that are fleshed out here. This work was supported by a grant (DC009659) from the National Institutes of Health.

References

1. Dell GS. A spreading activation theory of retrieval in language production. Psychological Review. 1986; 93:283–321. [PubMed: 3749399]
2. Fromkin V. The non-anomalous nature of anomalous utterances. Language. 1971; 47:27–52.

3. Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. Behavioral & Brain Sciences. 1999; 22:1–75. [PubMed: 11301520]
4. Garrett, MF. The psychology of learning and motivation. In: Bower, GH., editor. Volume 9: advances in research and theory. Academic Press; New York: 1975. p. 133-177.
5. Guenther FH, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements. Psychological Review. 1998; 105:611–633. [PubMed: 9830375]
6. Houde JF, Jordan MI. Sensorimotor adaptation in speech production. Science. 1998; 279:1213–1216. [PubMed: 9469813]
7. Fairbanks G. Systematic research in experimental phonetics: 1. A theory of the speech mechanism as a servosystem. Journal of Speech and Heearing Disorders. 1954; 19:133–139.
8. Kawato M. Internal models for motor control and trajectory planning. Current Opinion in Neurobiology. 1999; 9:718–27. [PubMed: 10607637]
9. Shadmehr R, Krakauer JW. A computational neuroanatomy for motor control. Exp Brain Res. 2008; 185:359–81. [PubMed: 18251019]
10. Shadmehr R, Mussa-Ivaldi FA. Adaptive representation of dynamics during learning of a motor task. Journal of Neuroscience. 1994; 14:3208–24. [PubMed: 8182467]
11. Wolpert DM, Ghahramani Z, Jordan MI. An internal model for sensorimotor integration. Science. 1995; 269:1880–2. [PubMed: 7569931]
12. Tian X, Poeppel D. Mental imagery of speech and movement implicates the dynamics of internal forward models. Frontiers in Psychology. 2010; 1:166. [PubMed: 21897822]
13. Grafton ST. The cognitive neuroscience of prehension: recent developments. Exp Brain Res. 2010; 204:475–91. [PubMed: 20532487]
14. Wolpert DM, Doya K, Kawato M. A unifying computational framework for motor control and social interaction. Philos Trans R Soc Lond B Biol Sci. 2003; 358:593–602. [PubMed: 12689384]
15. Perkell J, et al. Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. Speech Communication. 1997; 22:227–250.
16. Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F0 responses to manipulations in pitch feedback. J Acoust Soc Am. 1998; 103:3153–61. [PubMed: 9637026]
17. Larson CR, Burnett TA, Bauer JJ, Kiran S, Hain TC. Comparison of voice F0 responses to pitch-shift onset and offset conditions. J Acoust Soc Am. 2001; 110:2845–8. [PubMed: 11785786]
18. Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. Neuroimage. 2008; 39:1429–43. [PubMed: 18035557]
19. Tremblay S, Shiller DM, Ostry DJ. Somatosensory basis of speech production. Nature. 2003; 423:866–9. [PubMed: 12815431]
20. Grafton, ST., Aziz-Zadeh, L., Ivry, RB. The cognitive neurosciences. Gazzaniga, MS., editor. MIT Press; Cambridge, MA: 2009. p. 641-652.
21. Grafton ST, Hamilton AF. Evidence for a distributed hierarchy of action representation in the brain. Hum Mov Sci. 2007; 26:590–616. [PubMed: 17706312]
22. Diedrichsen J, Shadmehr R, Ivry RB. The coordination of movement: optimal feedback control and beyond. Trends Cogn Sci. 2010; 14:31–9. [PubMed: 20005767]
23. Jackson JH. Remarks on Evolution and Dissolution of the Nervous System. Journal of Mental Science. 1887; 33:25–48.
24. Gracco VL. Some organizational characteristics of speech movement control. J Speech Hear Res. 1994; 37:4–27. [PubMed: 8170129]
25. Browman CP, Goldstein L. Articulatory phonology: an overview. Phonetica. 1992; 49:155–80. [PubMed: 1488456]
26. Plaut, DC., Kello, CT. The emergence of language. MacWhinney, B., editor. Lawrence Erlbaum Associates; Mahwah, NJ: 1999. p. 381-416.
27. Bock, K. The MIT Encyclopedia of the Cognitive Sciences. Wilson, RA., Keil, FC., editors. MIT Press; Cambridge, MA: 1999. p. 453-456.
28. Dell, GS. An invitation to cognitive science: Language. Glietman, LR., Liberman, M., editors. MIT Press; Cambridge, MA: 1995. p. 183-208.
29. Levelt, WJM. Speaking: From intention to articulation. MIT Press; Cambridge, MA: 1989.

30. Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. Lexical access in aphasic and nonaphasic speakers. Psychological Review. 1997; 104:801–838. [PubMed: 9337631]
31. Levelt WJM. Models of word production. Trends in Cognitive Sciences. 1999; 3:223–232. [PubMed: 10354575]
32. Vigliocco G, Antonini T, Garrett MF. Grammatical gender is on the tip of Italian tongues. Psychological Science. 1998; 8:314–317.
33. Levelt WJ. Monitoring and self-repair in speech. Cognition. 1983; 14:41–104. [PubMed: 6685011]
34. Oppenheim GM, Dell GS. Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. Cognition. 2008; 106:528–37. [PubMed: 17407776]
35. Postma A. Detection of errors during speech production: a review of speech monitoring models. Cognition. 2000; 77:97–132. [PubMed: 10986364]
36. Huettig F, Hartsuiker RJ. Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. Language and Cognitive Processes. 2010; 25:347–374.
37. Nickels L, Howard D. Phonological errors in aphasic naming: comprehension, monitoring and lexicality. Cortex. 1995; 31:209–37. [PubMed: 7555004]
38. Ozdemir R, Roelofs A, Levelt WJ. Perceptual uniqueness point effects in monitoring internal speech. Cognition. 2007; 105:457–65. [PubMed: 17156770]
39. Nozari N, Dell GS, Schwartz MF. Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. Cognitive Psychology. 2011; 63:1–33. [PubMed: 21652015]
40. Golfinopoulos E, Tourville JA, Guenther FH. The integration of large-scale neural network modeling and functional brain imaging in speech motor control. Neuroimage. 2010; 52:862–74. [PubMed: 19837177]
41. Hickok G, Houde J, Rong F. Sensorimotor integration in speech processing: computational basis and neural organization. Neuron. 2011; 69:407–22. [PubMed: 21315253]
42. Houde JF, Nagarajan SS. Speech production as state feedback control. Frontiers in Human Neuroscience. 2011; 5
43. Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. Journal of Cognitive Neuroscience. 2003; 15:673–682. [PubMed: 12965041]
44. Hickok G, Okada K, Serences JT. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. J Neurophysiol. 2009; 101:2725–32. [PubMed: 19225172]
45. Howard D, Nickels L. Separating input and output phonology: Semantic, phonological, and orthographic effects in short-term memory impairment. Cognitive Neuropsychology. 2005; 22:42–77. [PubMed: 21038240]
46. Jacquemot C, Dupoux E, Bachoud-Levi AC. Breaking the mirror: Asymmetrical disconnection between the phonological input and output codes. Cognitive Neuropsychology. 2007; 24:3–22. [PubMed: 18416481]
47. Shelton JR, Caramazza A. Deficits in lexical and semantic processing: Implications for models of normal language. Psychonomic Bulletin & Review. 1999; 6:5–27. [PubMed: 12199314]
48. Ventura MI, Nagarajan SS, Houde JF. Speech target modulates speaking induced suppression in auditory cortex. BMC Neurosci. 2009; 10:58. [PubMed: 19523234]
49. Houde JF, Nagarajan SS. Speech production as state feedback control. Frontiers in Human Neuroscience. (in press).
50. Lichtheim L. On aphasia. Brain. 1885; 7:433–484.
51. Wernicke, C. Wernicke's works on aphasia: A sourcebook and review. Eggert, GH., editor. Mouton; The Hague: 1874/1977. p. 91-145.
52. Gracco VL, Lofqvist A. Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. Journal of Neuroscience. 1994; 14:6585–97. [PubMed: 7965062]
53. Perkell JS. Movement goals and feedback and feedback control mechanisms in speech production. Journal of Neurolinguistics. (in press).

54. Wilson SM, Saygin AP, Sereno MI, Iacoboni M. Listening to speech activates motor areas involved in speech production. Nat Neurosci. 2004; 7:701–702. [PubMed: 15184903]
55. Fadiga L, Craighero L, Buccino G, Rizzolatti G. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. Eur J Neurosci. 2002; 15:399–402. [PubMed: 11849307]
56. Cooper WE, Lauritsen MR. Feature processing in the perception and production of speech. Nature. 1974; 252:121–3. [PubMed: 4425217]
57. Delvaux V, Soquet A. The influence of ambient speech on adult speech productions through unintentional imitation. Phonetica. 2007; 64:145–73. [PubMed: 17914281]
58. Kappes J, Baumgaertner A, Peschke C, Ziegler W. Unintended imitation in nonword repetition. Brain Lang. 2009; 111:140–51. [PubMed: 19811813]
59. Christman SS, Boutsen FR, Buckingham HW. Perseveration and other repetitive verbal behaviors: functional dissociations. Seminars in Speech and Language. 2004; 25:295–307. [PubMed: 15599820]
60. Duffy, JR. Motor speech disorders: Substrates, Differntial diagnosis, and Management. Mosby; St. Louis: 1995.
61. Niemi M, Laaksonen JP, Ojala S, Aaltonen O, Happonen RP. Effects of transitory lingual nerve impairment on speech: an acoustic study of sibilant sound /s. Int J Oral Maxillofac Surg. 2006; 35:920–3. [PubMed: 16889939]
62. Niemi M, Laaksonen JP, Aaltonen O, Happonen RP. Effects of transitory lingual nerve impairment on speech: an acoustic study of diphthong sounds. J Oral Maxillofac Surg. 2004; 62:44–51. [PubMed: 14699548]
63. Niemi M, et al. Acoustic and neurophysiologic observations related to lingual nerve impairment. Int J Oral Maxillofac Surg. 2009; 38:758–65. [PubMed: 19369034]
64. Perkell JS, et al. The distinctness of speakers' /s/-/S/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. Journal of Speech Language and Hearing Research. 2004; 47:1259–69.
65. Liberman AM. Some results of research on speech perception. Journal of the Acoustical Society of America. 1957; 29:117–123.
66. Indefrey P, Levelt WJ. The spatial and temporal signatures of word production components. Cognition. 2004; 92:101–44. [PubMed: 15037128]
67. Rizzolatti G, et al. Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. Exp Brain Res. 1988; 71:491–507. [PubMed: 3416965]
68. Rizzolatti G, et al. Neurons related to reaching-grasping arm movements in the rostral part of area 6 (area 6a beta). Experimental Brain Research. 1990; 82:337–50. [PubMed: 2286236]
69. Rizzolatti G, et al. Neurons related to goal-directed motor acts in inferior area 6 of the macaque monkey. Experimental Brain Research. 1987; 67:220–4. [PubMed: 3622679]
70. Peeva MG, et al. Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. Neuroimage. 2010; 50:626–38. [PubMed: 20035884]
71. Aichert I, Ziegler W. Syllable frequency and syllable structure in apraxia of speech. Brain and Language. 2004; 88:148–59. [PubMed: 14698739]
72. Laganaro M, Croisier M, Bagou O, Assal F. Progressive apraxia of speech as a window into the study of speech planning processes. Cortex. 2011
73. Ogar J, Slama H, Dronkers N, Amici S, Gorno-Tempini ML. Apraxia of speech: an overview. Neurocase. 2005; 11:427–32. [PubMed: 16393756]
74. Hillis AE, et al. Re-examining the brain regions crucial for orchestrating speech articulation. Brain. 2004; 127:1479–87. [PubMed: 15090478]
75. Dronkers NF. A new brain region for coordinating speech articulation. Nature. 1996; 384:159–161. [PubMed: 8906789]
76. Ogar J, et al. Clinical and anatomical correlates of apraxia of speech. Brain and Language. 2006; 97:343–50. [PubMed: 16516956]
77. Ito M. Control of mental activities by internal models in the cerebellum. Nat Rev Neurosci. 2008; 9:304–13. [PubMed: 18319727]

78. Wolpert DM, Miall RC, Kawato M. Internal models in the cerebellum. Trends in Cognitive Sciences. 1998; 9:338–347.
79. Nowak DA, Topka H, Timmann D, Boecker H, Hermsdorfer J. The role of the cerebellum for predictive control of grasping. Cerebellum. 2007; 6:7–17. [PubMed: 17366262]
80. Desmurget M, Grafton S. Forward modeling allows feedback control for fast reaching movements. Trends Cogn Sci. 2000; 4:423–431. [PubMed: 11058820]
81. Pasalar S, Roitman AV, Durfee WK, Ebner TJ. Force field effects on cerebellar Purkinje cell discharge with implications for internal models. Nat Neurosci. 2006; 9:1404–11. [PubMed: 17028585]
82. Shadmehr R, Smith MA, Krakauer JW. Error correction, sensory prediction, and adaptation in motor control. Annu Rev Neurosci. 2010; 33:89–108. [PubMed: 20367317]
83. Baldo JV, Klostermann EC, Dronkers NF. It's either a cook or a baker: patients with conduction aphasia get the gist but lose the trace. Brain Lang. 2008; 105:134–40. [PubMed: 18243294]
84. Buchsbaum BR, et al. Conduction Aphasia and Phonological Short-term Memory: A Meta-Analysis of Lesion and fMRI data. Brain and Language. (e-pub 2011).
85. Damasio H, Damasio AR. The anatomical basis of conduction aphasia. Brain. 1980; 103:337–350. [PubMed: 7397481]
86. Goodglass, H. Conduction aphasia. Kohn, SE., editor. Lawrence Erlbaum Associates; Hillsdale, N.J: 1992. p. 39-49.
87. Ackermann H, Mathiak K, Riecker A. The contribution of the cerebellum to speech production and speech perception: clinical and functional imaging data. Cerebellum. 2007; 6:202–13. [PubMed: 17786816]
88. Ackermann H, Vogel M, Petersen D, Poremba M. Speech deficits in ischaemic cerebellar lesions. J Neurol. 1992; 239:223–7. [PubMed: 1597689]
89. Oppenheim GM, Dell GS. Motor movement matters: the flexible abstractness of inner speech. Mem Cognit. 2010; 38:1147–60.
90. Buchsbaum B, Hickok G, Humphries C. Role of Left Posterior Superior Temporal Gyrus in Phonological Processing for Speech Perception and Production. Cognitive Science. 2001; 25:663–678.
91. Buchsbaum BR, Olsen RK, Koch P, Berman KF. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. Neuron. 2005; 48:687–97. [PubMed: 16301183]
92. Murphy K, et al. Cerebral areas associated with motor control of speech in humans. J Appl Physiol. 1997; 83:1438–47. [PubMed: 9375303]
93. Shuster LI, Lemieux SK. An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. Brain Lang. 2005; 93:20–31. [PubMed: 15766765]
94. Smiley JF, et al. Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. J Comp Neurol. 2007; 502:894–923. [PubMed: 17447261]
95. Schroeder CE, et al. Somatosensory input to auditory association cortex in the macaque monkey. Journal of Neurophysiology. 2001; 85:1322–7. [PubMed: 11248001]
96. Foxe JJ, et al. Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Brain Res Cogn Brain Res. 2000; 10:77–83. [PubMed: 10978694]
97. Murray MM, et al. Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cereb Cortex. 2005; 15:963–74. [PubMed: 15537674]
98. Foxe JJ, et al. Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. Journal of Neurophysiology. 2002; 88:540–3. [PubMed: 12091578]
99. Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE. Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron. 2007; 53:279–92. [PubMed: 17224408]

100. Aliu SO, Houde JF, Nagarajan SS. Motor-induced suppression of the auditory cortex. J Cogn Neurosci. 2009; 21:791–802. [PubMed: 18593265]
101. Heinks-Maldonado TH, et al. Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. Arch Gen Psychiatry. 2007; 64:286–96. [PubMed: 17339517]
102. Frith CD, Blakemore S, Wolpert DM. Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. Brain Res Brain Res Rev. 2000; 31:357–63. [PubMed: 10719163]
103. Paus T, Perry DW, Zatorre RJ, Worsley KJ, Evans AC. Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. European Journal of Neuroscience. 1996; 8:2236–2246. [PubMed: 8950088]
104. Christoffels IK, van de Ven V, Waldorp LJ, Formisano E, Schiller NO. The sensory consequences of speaking: parametric neural cancellation during speech in auditory cortex. PLoS One. 2011; 6:e18307. [PubMed: 21625532]
105. Eliades SJ, Wang X. Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. Journal of Neurophysiology. 2003; 89:2194–207. [PubMed: 12612021]
106. Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M. The essential role of premotor cortex in speech perception. Curr Biol. 2007; 17:1692–6. [PubMed: 17900904]
107. D'Ausilio A, et al. The motor somatotopy of speech perception. Curr Biol. 2009; 19:381–5. [PubMed: 19217297]
108. Callan DE, Jones JA, Callan AM, Akahane-Yamada R. Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. Neuroimage. 2004; 22:1182–94. [PubMed: 15219590]
109. Wilson SM, Iacoboni M. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. Neuroimage. 2006; 33:316–25. [PubMed: 16919478]
110. Jazayeri M, Movshon JA. Optimal representation of sensory information by neural populations. Nat Neurosci. 2006; 9:690–6. [PubMed: 16617339]
111. Jazayeri M, Movshon JA. A new perceptual illusion reveals mechanisms of sensory decoding. Nature. 2007; 446:912–5. [PubMed: 17410125]
112. Regan D, Beverley KI. Postadaptation orientation discrimination. J Opt Soc Am A. 1985; 2:147–55. [PubMed: 3973752]
113. Scolari M, Serences JT. Adaptive allocation of attentional gain. J Neurosci. 2009; 29:11933–42. [PubMed: 19776279]
114. Vigliocco G, Hartsuiker RJ. The interplay of meaning, sound, and syntax in sentence production. Psychological Bulletin. 2002; 128:442–72. [PubMed: 12002697]
115. Friston K. The free-energy principle: a unified brain theory? Nat Rev Neurosci. 2010; 11:127–38. [PubMed: 20068583]
116. Summerfield C, Egner T. Expectation (and attention) in visual cognition. Trends Cogn Sci. 2009; 13:403–9. [PubMed: 19716752]
117. Hickok G, et al. A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. Neuroscience Letters. 2000; 287:156–160. [PubMed: 10854735]
118. Anderson JM, et al. Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model. Brain and Language. 1999; 70:1–12. [PubMed: 10534369]
119. Dronkers, N., Baldo, J. Encyclopedia of Neuroscience. Squire, LR., editor. Academic Press; Oxford: 2009. p. 343-348.
120. Hickok, G. Language and the brain. Grodzinsky, Y.Shapiro, L., Swinney, D., editors. Academic Press; San Diego: 2000. p. 87-104.
121. Galantucci B, Fowler CA, Turvey MT. The motor theory of speech perception reviewed. Psychon Bull Rev. 2006; 13:361–77. [PubMed: 17048719]
122. Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychol Rev. 1967; 74:431–61. [PubMed: 4170865]

123. Hickok G. The role of mirror neurons in speech perception and action word semantics. Language and Cognitive Processes. 2010; 25:749–776.

124. Lotto AJ, Hickok GS, Holt LL. Reflections on Mirror Neurons and Speech Perception. Trends Cogn Sci. 2009; 13:110–114. [PubMed: 19223222]

125. Massaro DW, Chen TH. The motor theory of speech perception revisited. Psychon Bull Rev. 2008; 15:453–7. discussion 458–62. [PubMed: 18488668]

126. Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition. 1985; 21:1–36. [PubMed: 4075760]

127. Stevens KN, Blumstein SE. Invariant cues for place of articulation in stop consonants. Journal of the Acoustical Society of America. 1978; 64:1358–68. [PubMed: 744836]

128. Stevens, KN., Halle, M. Models for the perception of speech and visual form. Walthen-Dunn, W., editor. MIT Press; Cambridge, MA: 1967. p. 88-102.

129. Nusbaum, HC., Magnuson, JS. Talker variability in speech processing. Johnson, K., Mullennix, JW., editors. Academic Press; San Diego: 1997. p. 109-132.

130. McClelland JL, Elman JL. The TRACE model of speech perception. Cognitive Psychology. 1986; 18:1–86. [PubMed: 3753912]

131. Massaro, DW. Handbook of psycholinguistics. Gernsbacher, MA., editor. Academic Press; San Diego: 1994. p. 219-263.

132. Vaden KI, Piquado T, Hickok G. Sublexical Properties of Spoken Words Modulate Activity in Broca's Area but Not Superior Temporal Cortex: Implications for Models of Speech Recognition. J Cogn Neurosci. 2011

133. Greenberg, S. Listening to speech: an auditory perspective. Greenberg, S., Ainsworth, WA., editors. Erlbaum; Mahwah, NJ: 2005. p. 411-433.

134. Klatt DH. Speech perception: a model of acoustic-phonetic analysis and lexical access. Journal of Phonetics. 1979; 7:279–312.

135. Johnson K. The auditory/perceptual basis for speech segmentation. OSU Working Papers in Linguistics. 1997; 50

136. Johnson K. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. Journal of Phonetics. 2006; 34:485–499.

137. Goldinger SD. Echoes of echoes? An episodic theory of lexical access. Psychol Rev. 1998; 105:251–79. [PubMed: 9577239]

138. Stevens KN. Toward a model for lexical access based on acoustic landmarks and distinctive features. Journal of the Acoustic Society of America. 2002; 111:1872–1891.

139. Marslen-Wilson WD. Functional parallelism in spoken word-recognition. Cognition. 1987; 25:71–102. [PubMed: 3581730]

Box 3

Feedback control simulation

The simulation was intended to model a small phonological neighborhood at one level in the hierarchy in both motor and sensory space with the connectivity pattern shown in Figure 5. Specifically, it was assumed that the target lemma — 'CAT' in the example — projected reciprocally to all nodes in the motor and sensory neighborhood, corresponding sensory and motor nodes are reciprocally connected to each other, and each node with the sensory and motor phonological space reciprocally inhibits the other nodes within that sensory or motor space. Activity at each node was calculated by summing all of a node's weighted inputs and adding this to its existing activation level as described in the following equation.

$$A(j,t)=A(j,t-1)(1-q)+\sum pA(i,t-1)$$

where $A(j, t)$ is the activation level of node j at time t, q is a decay rate, and p is connection weight. The model is fully linear in that negative activation values are added in rather preventing negative activations from spreading.

Learning was not simulated, nor was a sensorimotor transformation layer because only a small representational space was modeled.

The following parameters were used for all simulations. Input activation to the lemma node was provided for 5 time steps at a level of 0.3 then dropped to zero for all remaining time steps. Decay rate = 0.7, motor-to-sensory (forward prediction) inhibitory weight = 1.0, sensory-to-motor excitatory weight = 1.0, lateral inhibition weight = 0.25.

The first simulation assumed a strong and selective activation of both the auditory and motor phonological targets. Weights to target nodes = 0.8 and weights to non-target nodes = 0.2. The correct motor node is activated while the corresponding auditory node is initially activated and then quickly inhibited (Figure 5B). The entire network then returns to baseline. The remaining simulations show the effects of weaker motor selectivity (0.6 target, 0.4 non-targets; Figure 5C), no motor selectivity (0.5 target, 0.5 non-targets; Figure 5D), and incorrect lemma-to-motor node activation (target = 0.2, non-target$_1$ = 0.8, non-target$_2$ = 0.2; Figure 5E), while maintaining a strong and selective lemma-to-auditory activation, as in the first simulation. In other words, these simulations assess the ability of a strong and selective auditory target activation to overcome weak or incorrect motor node activation; this may be part of the mechanism for learning motor correlates of auditory (and at a higher level, word/lemma) targets, e.g., via Hebbian learning. It is clear from Figures 5B–E that an accurate auditory node activation can correct a weak or incorrect motor node activation in a network with this type of architecture. In the absence of auditory target selection (no selectivity for auditory targets) a strongly activated motor target alone is sufficient to yield activation of a motor node (Figure 5F) as would be the case in conduction aphasia if the present account is correct.

Large-scale simulations will be needed to determine how successfully this kind of model will scale up, but the present simulation suggests that the architecture presented here is at least a viable possibility worth further investigation. In addition, it is worth noting explicitly that this is an abstract model that could be implemented neurally in a number of ways.

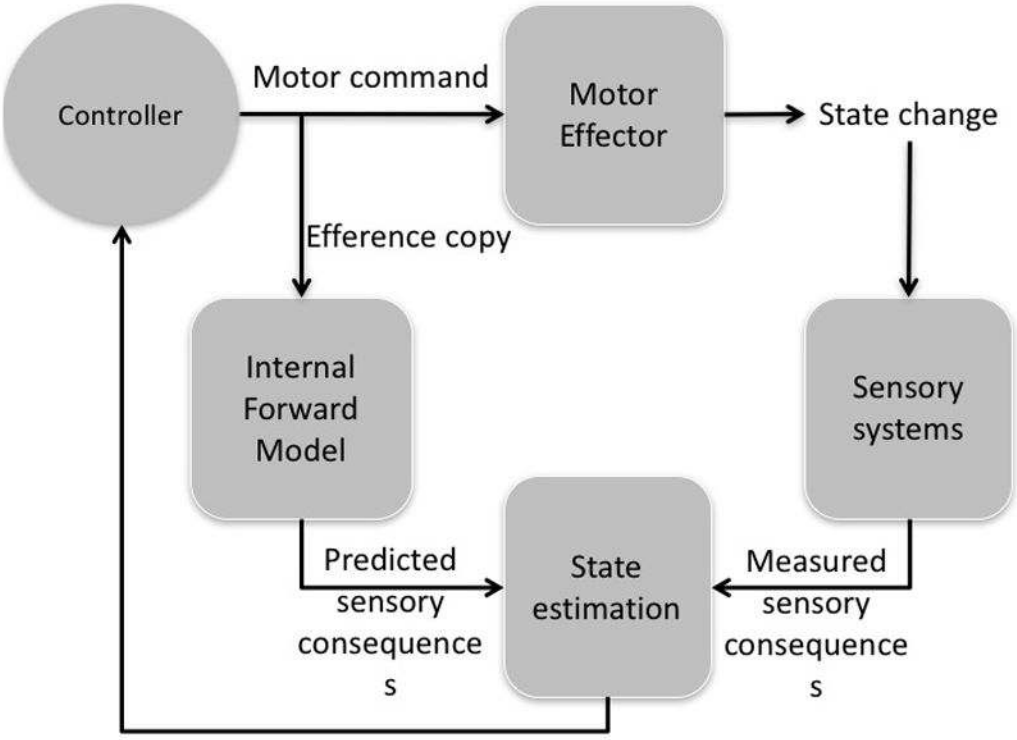

Figure 1. State feedback control

State feedback control models typically include a motor controller that sends commands to a motor effector, which in turn result in a change of state, such as a change in the position of an arm. State changes are detected by sensory systems. Most models also include an internal forward model that receives a copy of the motor command issued by the controller and generates a prediction of the sensory consequences of the command that can be compared against the measured sensory consequences. The error between the predicted and measured sensory consequences is used as a correction signal to correct such error. Image adapted from Shadmehr and Krakauer 2008.

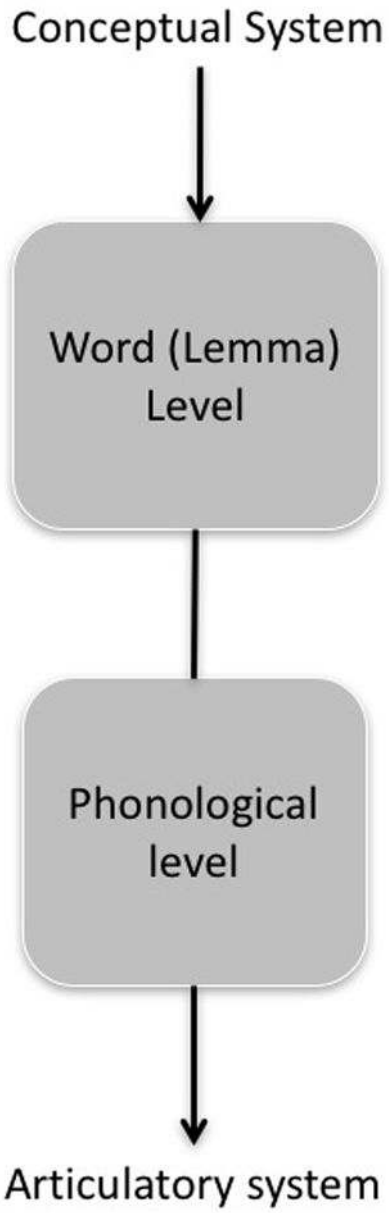

Figure 2. Two-stage psycholinguistic model of speech production

Psycholinguistic models of speech production typically identify two major linguistic stages of processing, the word (or lemma) stage in which an abstract word form without phonological specification is coded and the phonological stage in which the phonological form of the word is coded. The distinction between these stages can be intuitively understood when considering tip-of-the-tongue states in which we know the word we want to use (that is, we have accessed the lemma) but we cannot retrieve the phonological form. These linguistic stages of processing receive input from the conceptual system and send output to the motor articulatory system. Conceptual and articulatory processes are typically considered outside the domain of linguistic analysis of speech production.

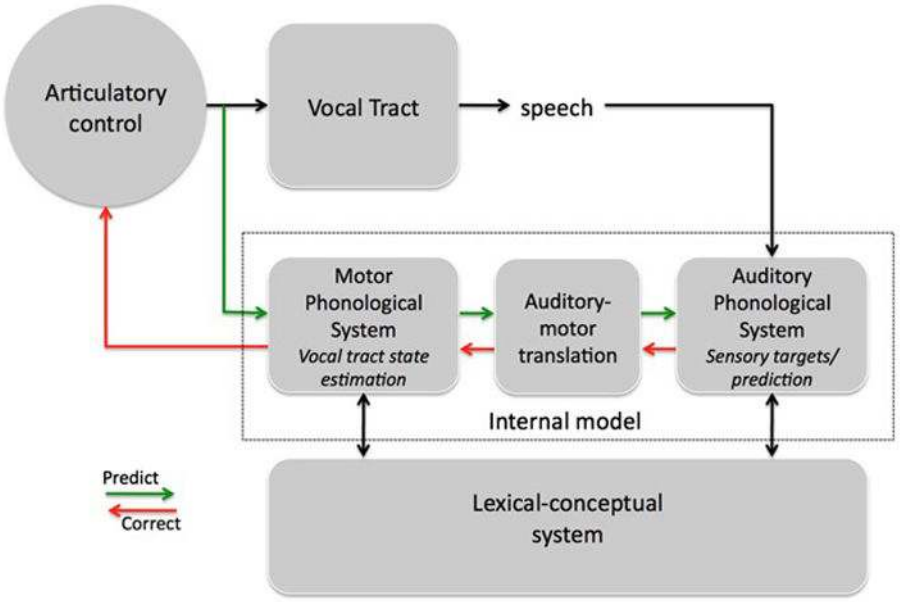

Figure 3. The state feedback control (SFC) model

The architecture of the SFC model is derived from state feedback models of motor control but it incorporates processing levels that have been identified in psycholinguistic research (particularly those in the two-stage psycholinguistic model). The SFC model includes a motor controller that sends an efference copy to the internal model (dashed box), which generates predictions as to the state of the vocal tract in the motor phonological system as well as predictions of the sensory consequences of an action in the auditory phonological system. This division of labour is supported by neuropsychological findings. Communication between the auditory and motor systems is achieved by an auditory–motor translation system. The two stages of the psycholinguistic model are evident in the lexical-conceptual system, which is intended to represent, in part, the lemma level, and the motor–auditory phonological systems, which correspond to the phonological level. Reprinted with permission.

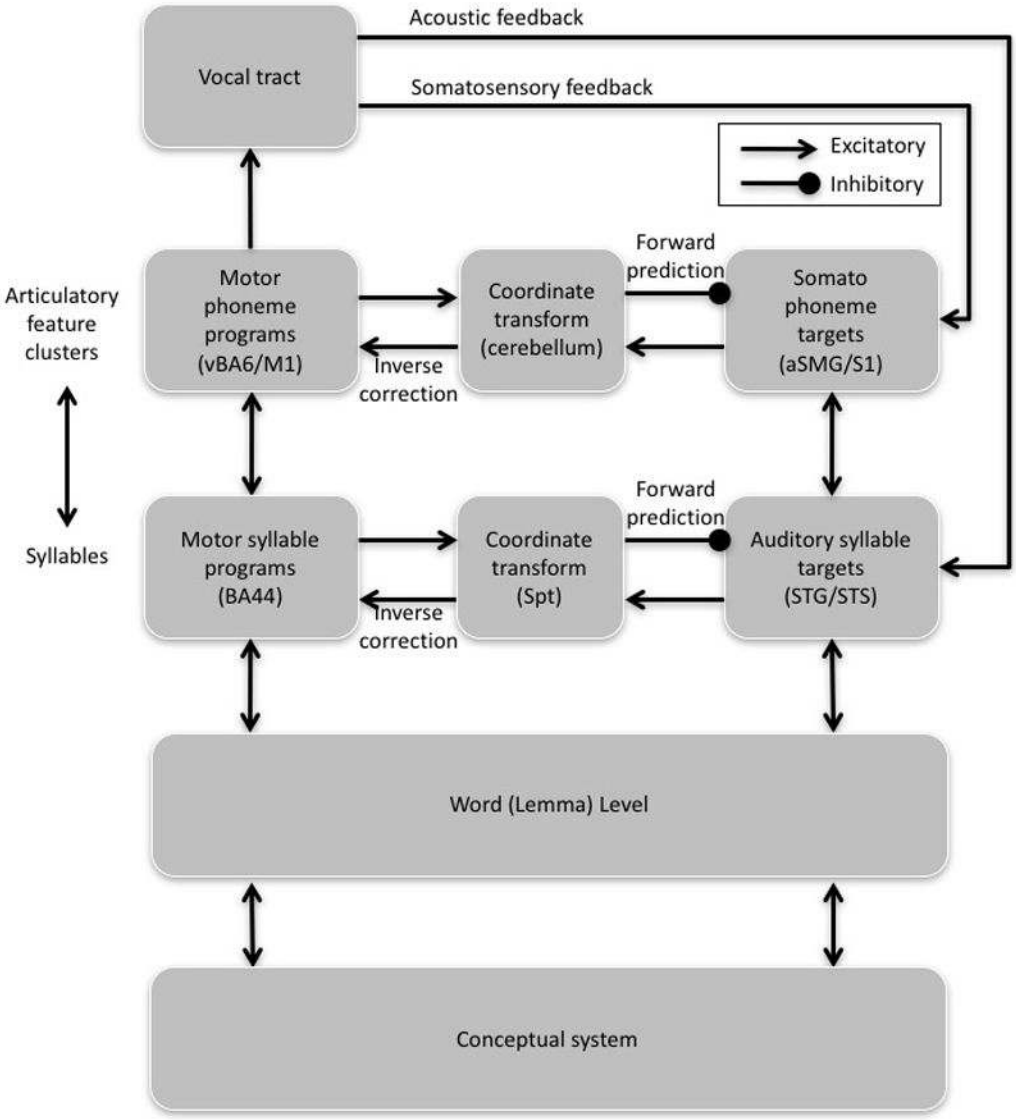

Figure 4. The hierarchical state feedback control (HSFC) model

The HSFC model includes two hierarchical levels of feedback control, each with its own internal and external sensory feedback loops. As in psycholinguistic models, the input to the HSFC model starts with the activation of a conceptual representation that in turn excites a corresponding word (lemma) representation. The word level projects in parallel to sensory and motor sides of the highest, fully cortical level of feedback control, the auditory–Spt–BA44 loop. This higher-level loop in turn projects, also in parallel, to the lower-level somatosensory–cerebellum–motor cortex loop. Direct connections between the word level and the lower-level circuit may also exist, although they are not depicted here. The HSFC model differs from the state feedback control (SFC) model in two main respects. First, 'phonological' processing is distributed over two hierarchically organized levels implicating a higher-level cortical auditory–motor circuit and a lower-level somatosensory–motor circuit, which roughly map onto syllabic and phonemic levels of analysis, respectively. Second, a true efference copy signal is not a component of the model. Instead, the function

served by an efference copy is integrated into the motor planning process. BA, Brodmann area; M1, primary motor cortex; S1, primary somatosensory area; aSMG, anterior supramarginal gyrus; STG, superior temporal gyrus; STS, superior temporal sulcus; vBA6, ventral BA6.

Figure 5A

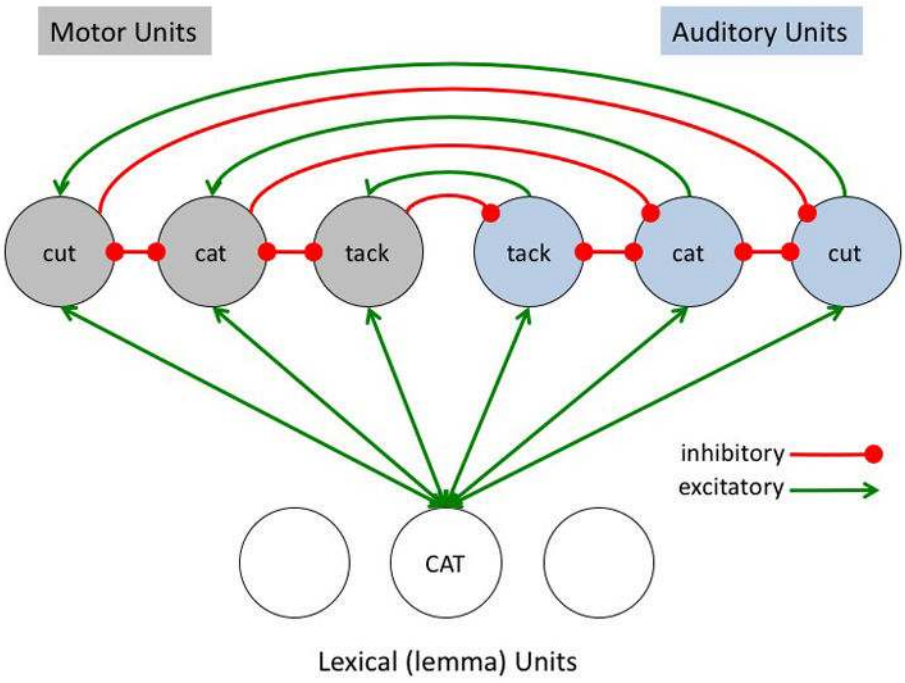

Figure 5B

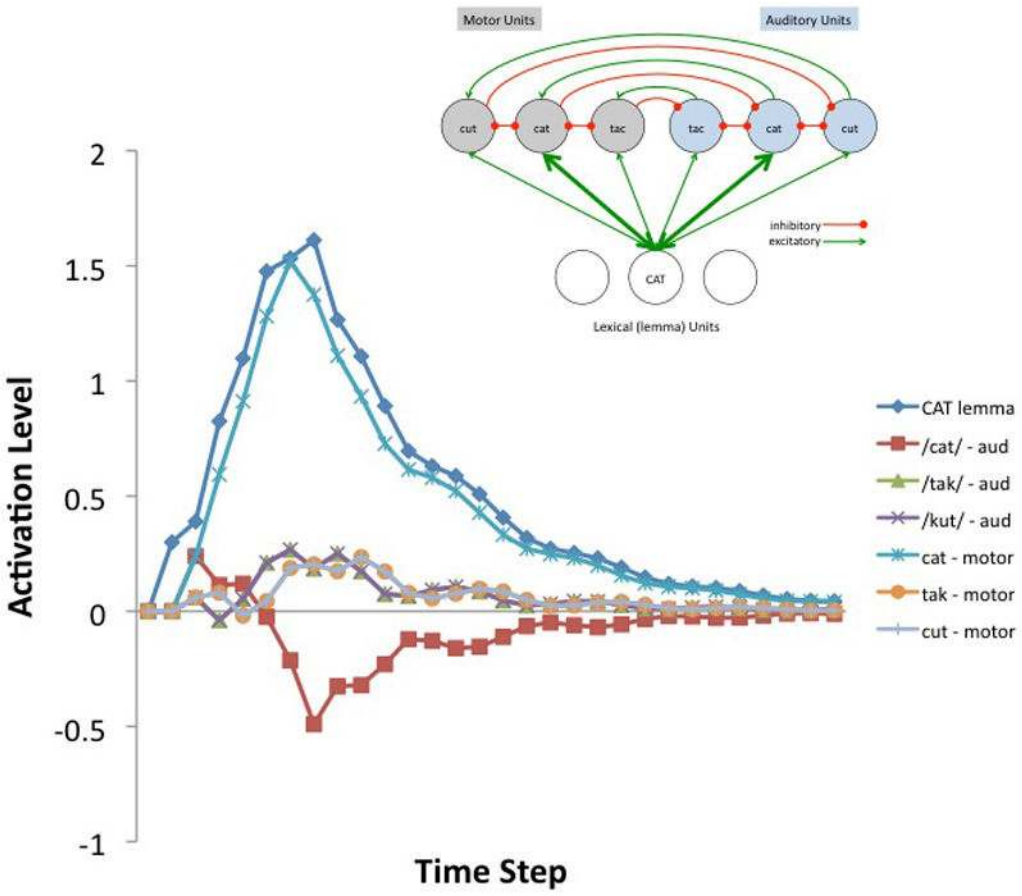

Figure 5C

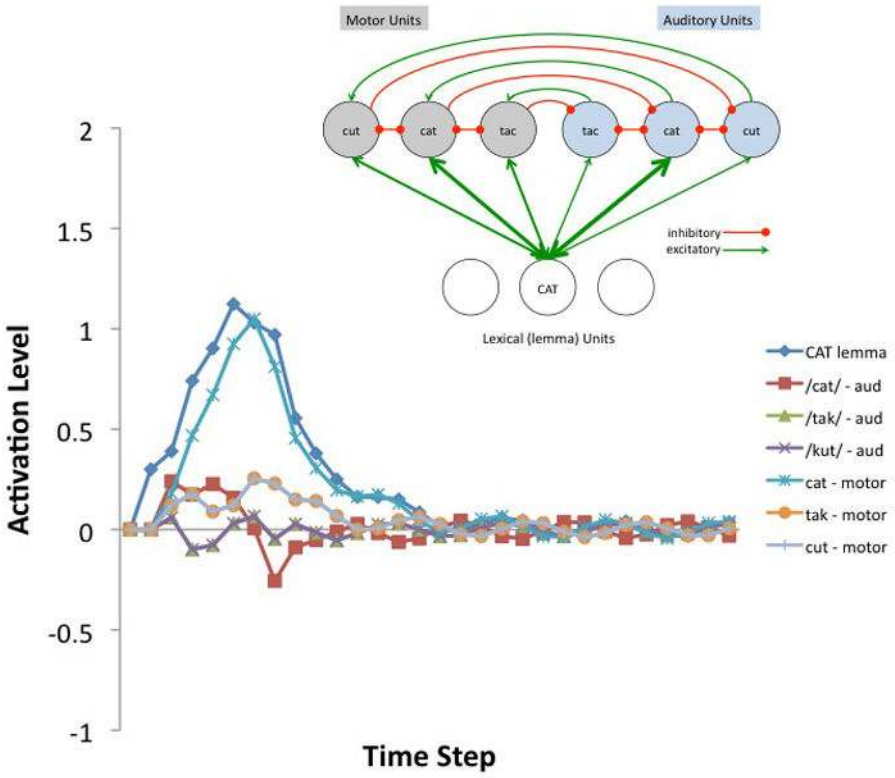

Figure 5D

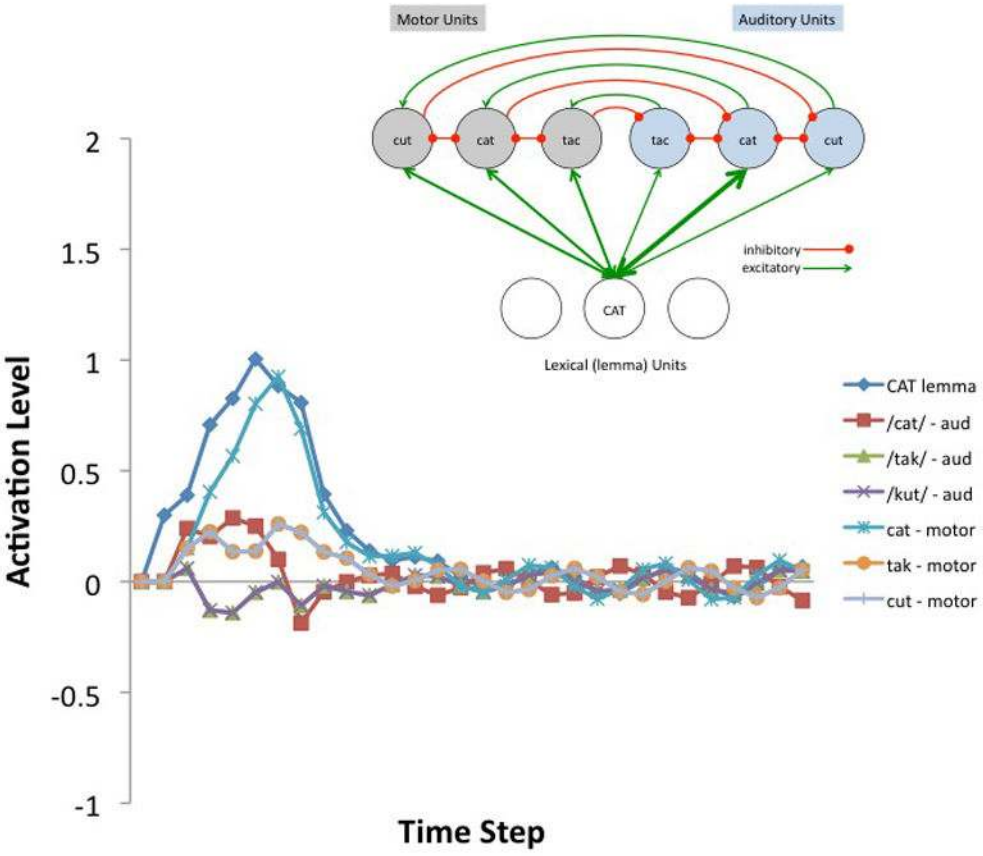

Figure 5E

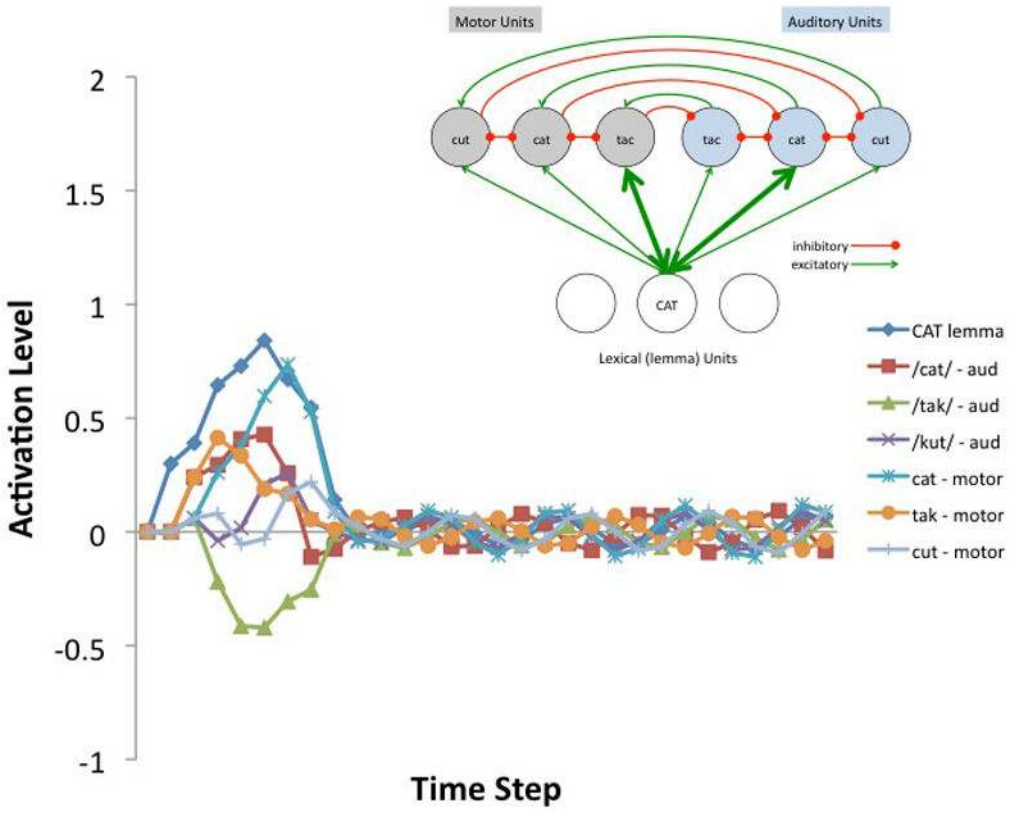

Figure 5F

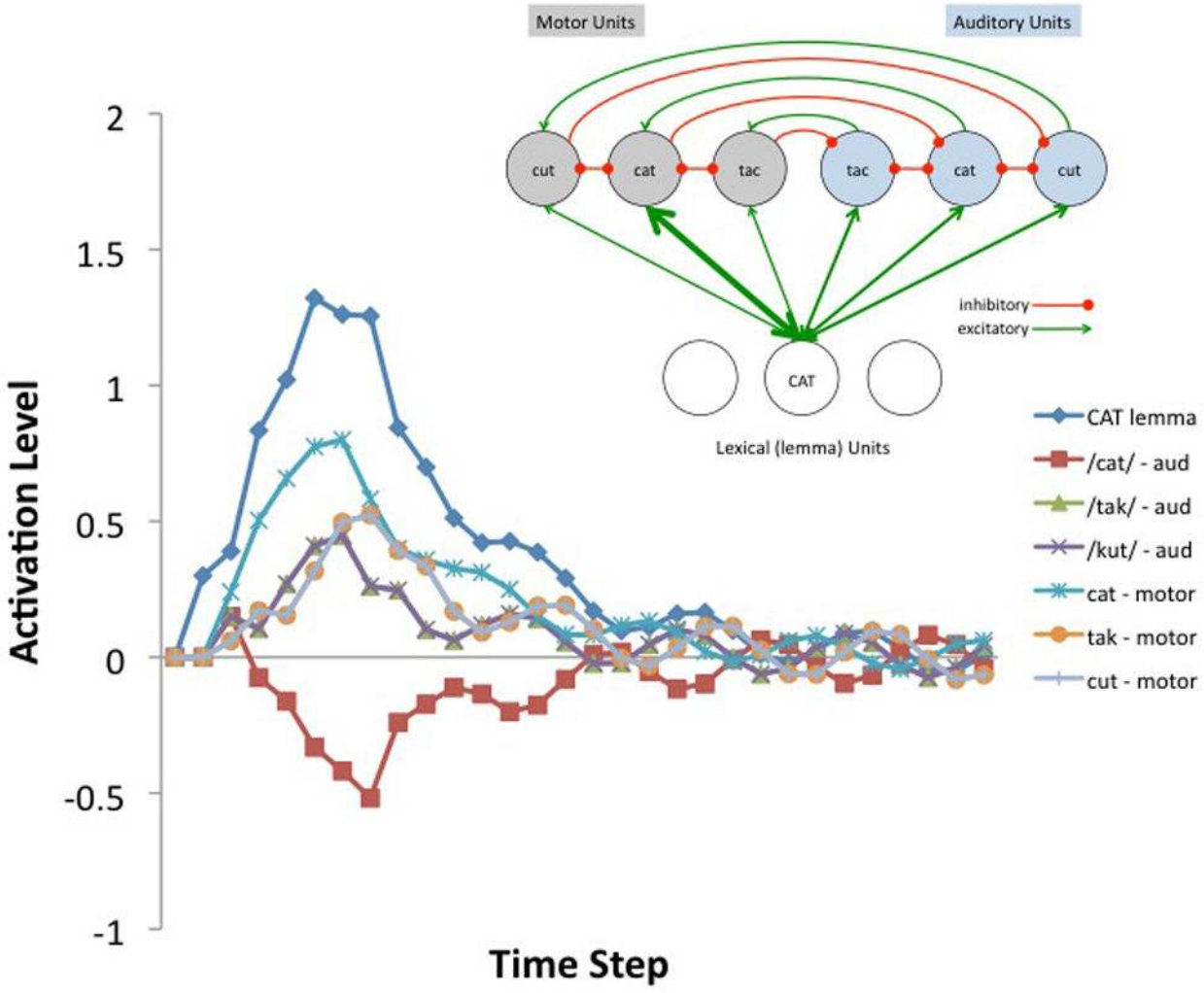

Figure 5. Internal feedback control simulation

The simulation aims to model a small component of the proposed hierarchical state feedback control (HSFC) model of speech production. **a** | The modelled fragment comprises one node in the lemma level network and a small phonological neighbourhood at the auditory (blue nodes) and motor (grey nodes) levels. The lines represent excitatory (arrows) and inhibitory (filled circles) connections. Specifically, it was assumed that the target lemma — 'CAT' — projects reciprocally to all nodes in the motor and sensory neighborhood (that is, to the target, 'cat' as well as non-targets 'tack' and 'cut'), that corresponding sensory and motor nodes are reciprocally connected to each other, and that each node within the sensory and motor phonological space reciprocally inhibits the other nodes within that sensory or motor space. Activity at each node was calculated by summing all of a node's weighted inputs and adding this to its existing activation level as described in the following equation.

$$A(j,t)=A(j,t-1)(1-q)+\sum pA(i,t-1)$$

where $A(j, t)$ is the activation level of node j at time t, q is a decay rate, and p is connection weight. The model is fully linear in that negative activation values are included in the sum. Learning was not simulated, nor was a sensorimotor transformation layer because only a small representational space was modeled. The following parameters were used for all simulations. Input activation to the lemma node was provided for 5 time steps at a level of 0.3 then dropped to zero for all remaining time steps. The decay rate was 0.7, the motor-to-sensory (forward prediction) inhibitory weight was 1.0, the sensory-to-motor excitatory weight was 1.0, and the lateral inhibition weight was 0.25. **b |** Simulated behaviour of the model when connection weights to the auditory and motor targets are strong and selective (the weights to the target nodes were 0.8 and the weights to non-target nodes were 0.2). Note that the correct motor target is activated and the auditory target is suppressed after an initial brief activation. The entire network then returns to baseline. **c |** Simulated behaviour of the model when connection weights to the auditory target are strong and selective (as above) but such weights to motor targets are weaker and less selective (the weights to the target nodes were 0.6 and the weights to non-target nodes were 0.4). **d |** Simulated behaviour of the model when connection weights to the auditory target are strong and selective but there is no selectivity for the motor target (the weights to the motor target and non-target nodes were 0.5). This scenario represents auditory guided motor selection. **e |** Simulated behaviour of the model when connection weights to the auditory target are strong and selective but there is a strong and selective activation of the wrong motor target. Activation of the auditory target overcomes the initial incorrect motor activation. This scenario represents internal error correction in motor selection. **f |** Simulated behaviour of the model when the connection weights to the motor target are strong and selective but those to the auditory target are not selective. Correct motor activation is also possible under this scenario but is not as robust compared to when the auditory target is also activated (as in panel **b).** Large-scale simulations will be needed to determine how successfully this kind of model will scale up, but the present simulation suggests that the architecture presented here is at least a viable possibility worth further investigation. In addition, it is worth noting explicitly that this is an abstract model that could be implemented neurally in a number of ways.