Nature Reviews Neuroscience | AOP, published online 5 january 2012; doi:10.1038/nrn3158

PERSPECTIVES

OPINION

Computational neuroanatomy of speech production

Gregory Hickok

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: whereas 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 level 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 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 (HSFC) 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 the 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 via a region in the Sylvian fissure at the parieto-temporal boundary (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 crucial component of motor control, but 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 successfully negotiate the track 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 negotiate 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 practised 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 brain, 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[9] (FIG. 1). 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 of 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 crucial 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 crucial 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 by 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 (SFC) 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 SFC provides a solution to the engineering problem outlined above, and the existence of such internal models in SFC has been supported experimentally[10–12]. For these reasons, the SFC approach has been highly influential and is 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 the motor control tradition, strong evidence can be found for the existence of internal SFC 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 SFC, 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.

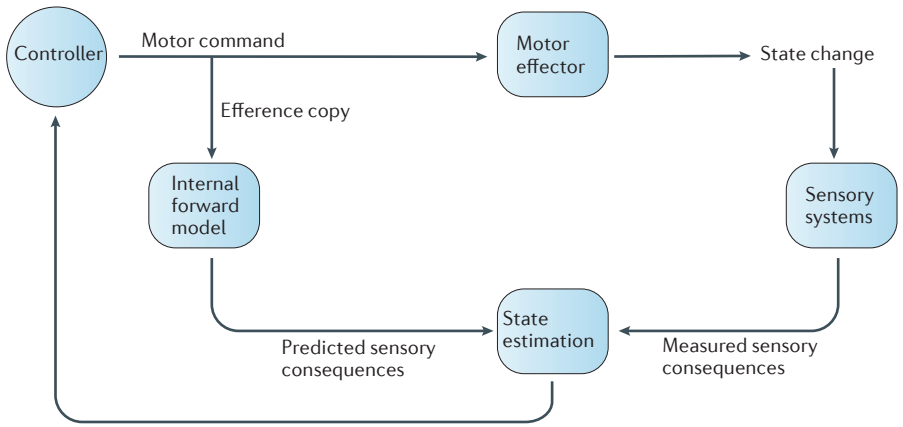

Figure 1 | **State feedback control.** State feedback control models typically include a motor controller that sends commands to a motor effector, which in turn results in a change of state (such as a change in the position of an arm). State changes are detected by sensory systems. Most state feedback control models also include an internal forward model that receives a copy of the motor command that is 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 difference between the predicted and measured sensory consequences is used as a motor correction signal that relays to the controller. Figure is adapted, with permission, from REF. 9 © (2008) Springer.

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 to be 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 network 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 various 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[35–39], including the suggestion that error detection and correction in speech may not rely on sensory systems[34], 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, for example, 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 into 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 briefly review that model as a starting point.

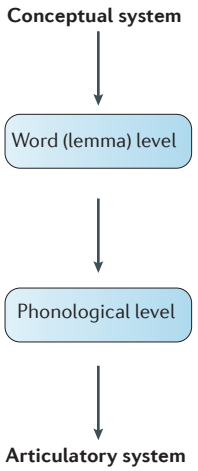

Figure 2 | **Two-stage psycholinguistic model of speech production.** Psycholinguistic models of speech production typically identify two major linguistic stages of processing: the word (or lemma) stage, in which an abstract word form without phonological specification is coded, and the phonological 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 to be outside the domain of linguistic analysis of speech production.

An integrated state feedback control model. The integrated SFC model of speech production[41] (FIG. 3) builds on models proposed by Guenther *et al.*[5], Tian and Poeppel[12], and Houde and Nagarajan[42]. Consistent with SFC models generally, the integrated 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, labelled 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, which is 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[5,42,48]. 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 nineteenth century models of the neural organization of language[49,50] and provides an explanation for certain forms of language disruption following brain injury (BOX 1).

Extending the model. Here, I outline an extension of the integrated SFC model, which I refer to simply as the HSFC model (FIG. 4). In the HSFC model there are two hierarchically organized levels of SFC, which are similar to those proposed by Gracco and Lofqvist[24,51]. 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 the auditory cortex and motor programs coded in the Brodmann area 44 (BA44) portion of Broca's area and/or lower BA6, with the area Spt computing a coordinate transform between the sensory and motor areas. This is the loop described in the 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[24,51] and involve a sensory–motor loop that includes sensory targets coded primarily in the somatosensory cortex (as suggested by V. Gracco, personal communication) and motor programs coded in the lower primary motor cortex (M1), with 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[5,15,52]. Here, I 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[53,54] and that the speech of others, even if it is ambient, can be unintentionally imitated by a listener or speaker[55–57]. The existence of echolalia, the tendency of individuals with certain acquired or developmental speech disorders to repeat heard speech[58,59], 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)[60–62]. For these reasons, motor–somatosensory loops are prominent components of motor control models for speech[5,40,63].

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 articulators, we can then view phonemes as the articulatory configurations that are defined by the end points (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 positions (consonants) or open positions (vowels). Owing to co-articulation, however, the acoustic consequences of the articulatory configurations that define phonemes are not restricted to — and indeed are 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[64] (BOX 2). However, the vocal tract configurations that define phonemes, the end points of an articulatory half cycle, do have somatosensory consequences. Lip closure or tongue raising, for example, have detectable somatosensory consequences at the end-point vocal tract configurations — the point in time that defines the phoneme — even in the

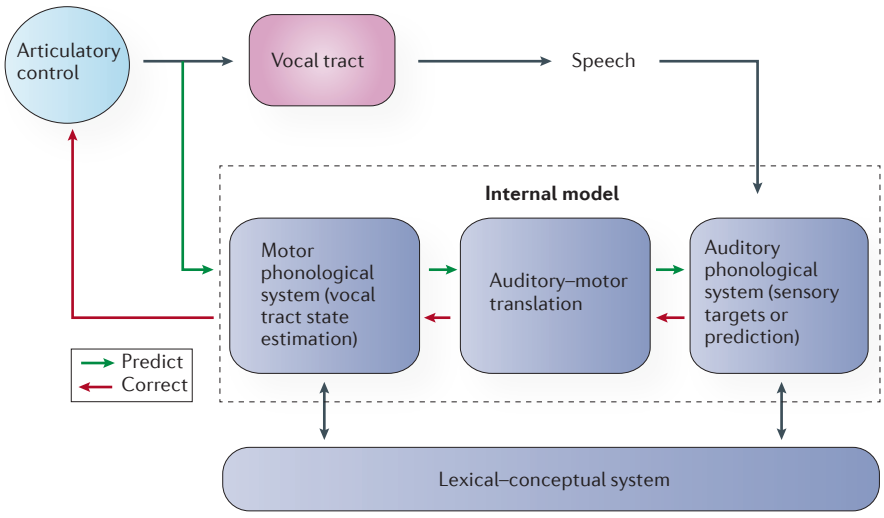

Figure 3 | **The state feedback control model.** The architecture of the state feedback control (SFC) model is derived from state feedback models of motor control, but it incorporates processing levels that have been identified in psycholinguistic research (particularly those in the two-stage psycholinguistic 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. Figure is reproduced, with permission, from REF. 41 © (2011) Elsevier.

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 end points of each half cycle.

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[52,63]. Given these considerations, phonemes and syllables may be distributed, in a 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 end point 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[52]. 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 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 not only with the present HSFC model but also with Levelt *et al.*'s 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[65], which is consistent with the HSFC 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'[66] and then fine tuned to particular situations, which can vary in terms of muscle fatigue, mechanical loads, obstacles, and so on[13].

The ventral premotor cortex and motor vocabularies. The ventral premotor cortex has been implicated in motor vocabularies in both speech and manual gestures[13,40,42,67,68]. As noted above, Levelt *et al.*'s notion of a mental syllabary — a repository of gestural scores for the most highly used syllables in a language[3] — has been linked to the ventral premotor cortex in a large-scale meta-analysis of functional imaging studies[65]. 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 the ventral premotor cortex to repeating syllables[69].

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[70,71]. Although this conclusion should be regarded as 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[72]. The ventral premotor cortex has been implicated in the aetiology of AOS[73], as has the nearby anterior insula[74,75]. 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[72]. 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

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[85]. 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[82–84], 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[50,116]: 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 part in speech production once the systems are disconnected (see also REF. 46 for similar arguments). What was lacking from Wernicke's account, however, was a principled explanation for why the sensory system has 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[117–119]. Modern findings instead implicate a cortical system that computes a sensorimotor coordinate transformation[43,44,83,89]. In short, the integrated SFC model improves our understanding of conduction aphasia[41].

components of the circuit (AOS affects access to motor phonological codes and conduction aphasia affects internal SFC).

In the visual–manual domain, physiological evidence from monkeys has suggested the existence of grasping-related motor vocabularies in the ventral premotor cortex[67,68]. 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 by interactions with the primary motor cortex[13]. This conceptualization is 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[18,40,76–79], and the cerebellum has been specifically implicated as being part of a forward model[80,81]. The suggestion here is that parietal and cerebellar circuits are performing a similar sensory–motor coordinate transform function but at different levels in the sensory–motor hierarchy (see REF. 77 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[82–85], cerebellar dysfunction results in a characteristic dysarthria comprising a slowing down of 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[86,87] (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"[86].

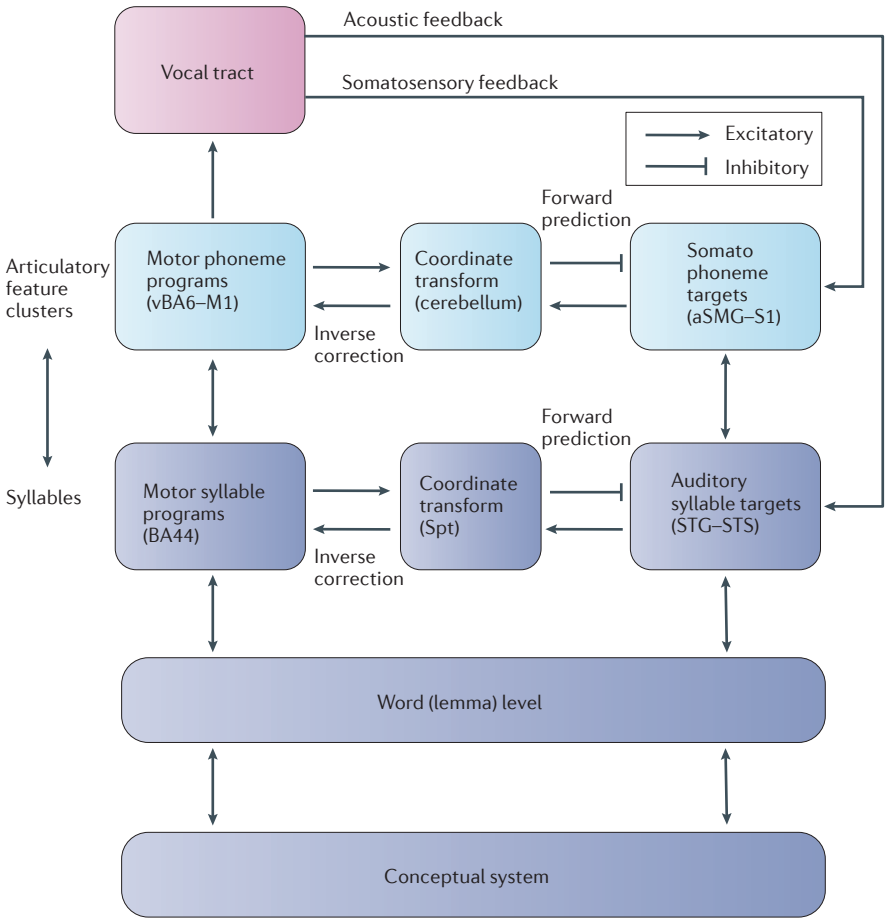

Figure 4 | **The hierarchical state feedback control model.** The hierarchical state feedback control (HSFC) model includes two hierarchical levels of feedback control, each with its own internal and external sensory feedback loops. As in psycholinguistic models, the input to the HSFC model starts with 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 (in which Spt stands for Sylvian fissure at the parietotemporal boundary and BA44 stands for Brodmann area 44). 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. aSMG, anterior supramarginal gyrus; M1, primary motor cortex; S1, primary somatosensory cortex; STG, superior temporal gyrus; STS, superior temporal sulcus; vBA6, ventral BA6.

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 non-words) 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[35]. 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[35]. By contrast, when speech is silently articulated, both lexical and phonemic similarity effects are detectable[88].

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[88].

Imagined speech without articulation activates a network that includes posterior portions of Broca's area, the dorsal premotor cortex, area Spt, and the posterior superior temporal sulcus–superior temporal gyrus[43,89,90]. Studies that have directly contrasted fully imagined speech with silently articulated speech have reported greater involvement of the primary motor and somatosensory cortex with articulated speech than with unarticulated speech[91,92], which is 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[93,94]. Electrophysiological[95,96] and fMRI[97] data have confirmed similar auditory–somatosensory interaction in the 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 the auditory cortex by somatosensory inputs[98]. Perceptual modulation is an important aspect of control circuits — forward predictions can be viewed as a form of perceptual modulation[41,99–101] — 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 the 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 end points, which define the phase of articulation, somatosensory-driven phase resetting in the auditory system may provide crucial 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 that have been argued to support somatosensory influence on auditory perception (see above) also support auditory-to-somatosensory information flow. Nonetheless, the present model, as well as others such as the DIVA model[40,52], predict an auditory–somatosensory interaction; such an interaction is consistent with available evidence.

Computational considerations. It is typically assumed that forward prediction is enabled by 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 part 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 target of a speech act, which is activated by input from the lexical (lemma) level. The lemma also activates the associated motor phonological representation. The sensory and motor phonological systems then interact in the following way 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[40,47], 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 towards activating the correct motor representation. Thus, forward prediction and error correction in the HSFC model is part

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, for example, /d/, can have different acoustic patterns in different syllable contexts, such as in /di/ and /da/[64]. This lack of invariance between acoustics and perception is arguably the fundamental problem in speech perception[64,120,121]. An early solution to this problem was the motor theory, which held that the target of speech perception is not acoustic representations but motor gestures[64,121]. However, the idea that low-level articulatory plans form the basis of perception has been rejected on empirical grounds[122–124]. In response, variants of the model have been proposed in which the objects of speech perception are more abstract gestural goals[120,125], but this idea is functionally indistinguishable from an auditory theory that assumes that the goals of speech gestures are sensory states.

To resolve the lack of invariance problem, several other approaches have been taken, including the search for possibly overlooked acoustic features that hold an invariant relation to phonemic categories[126]. In addition, a range of approaches that accept that a variable acoustic–phonemic relationship exists, but these approaches use various active processes, such as motor prediction[127], normalization[128] or top-down lexical constraint[129], to circumvent the problem.

Another class of solutions that is 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[124,130–133]; 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[134–136]. 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[129,137,138]. 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 that is less involved in speech recognition.

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 humans[99,102,103] and non-human primates[104] 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.

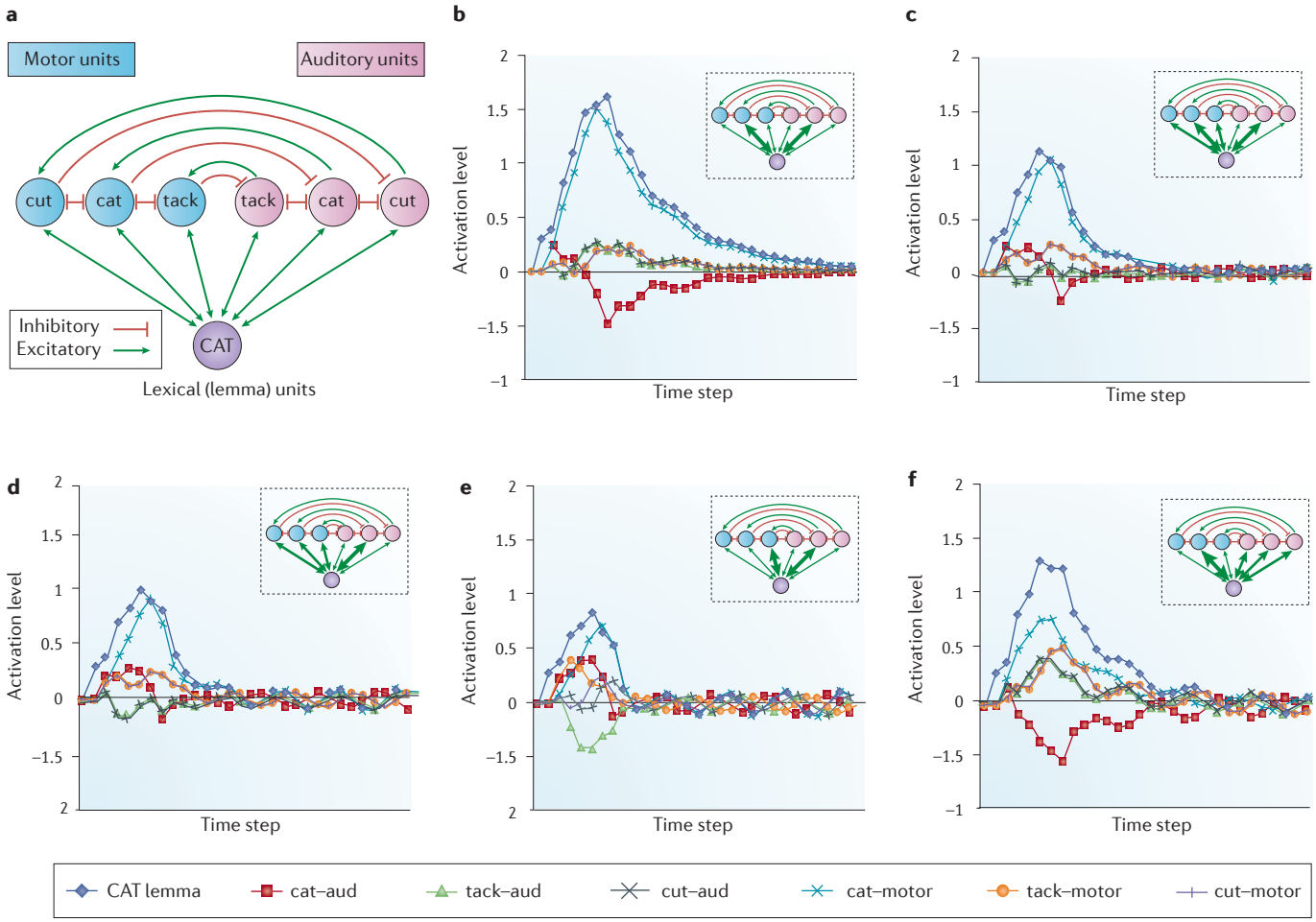

Figure 5 | **Internal feedback control simulation.** The simulation aims to model a small component of the proposed hierarchical state feedback control (HSFC) model of speech production. **a** | The modelled fragment comprises one node in the lemma-level network and a small phonological neighbourhood at the auditory level (pink nodes) and motor level (blue nodes). The lines represent excitatory and inhibitory connections. Specifically, it was assumed that the target lemma ('CAT') projects reciprocally to all nodes in the motor and sensory neighbourhood (that is, to the target, 'cat', as well as to the 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: $A(j, t) = A(j, t-1)(1-q) + \Sigma\, pA(i, t-1)$, 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 modelled. The following parameters were used for all simulations. Input activation to the lemma node was provided for five 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 were 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 was activated and the auditory target was suppressed after an initial brief activation. The entire network then returned to baseline. **c** | Simulated behaviour of the model when connection weights to the auditory target were strong and selective but such weights to motor targets were weaker and less selective (the weights to the target motor nodes were 0.6 and the weights to non-target motor nodes were 0.4). **d** | Simulated behaviour of the model when connection weights to the auditory target were strong and selective but there was no selectivity for the motor targets (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 were strong and selective but there was a strong and selective activation of the wrong motor target. Activation of the auditory target overcame 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 were strong and selective but those to the auditory target were not selective. Correct motor activation is also possible under this scenario but is not as robust as when the auditory target is also activated (as in **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 that is worth further investigation.

The present simulation result — that is, 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[105], sometimes in highly specific ways[106]: motor activation can result in a modulation of sensory systems, thus potentially affecting perception[41,107,108].

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 for detecting differences between targets and non-targets[109–112]. 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[34,113]. 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. However, in the later, external monitoring phase, the sensory representation is suppressed, which is consistent with some models of top-down sensory prediction[114,115], and this enhances the 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 in which 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.

Glossary

Fricatives
Speech sounds produced by forcing air through a small constriction in the vocal tract, creating turbulent air flow. Examples from English include [v], [f] and [s].

Liquids
Speech sounds produced by a constriction of the vocal tract, but not enough to cause the turbulent airflow associated with fricatives. Examples from English include [l] and [r].

Morphemes
Morphemes are the smallest units of meaning in a language. They can be 'free' (that is, they can exist as a free-standing unit, as in the word 'cook') or 'bound' (that is, they must be tied to another morpheme, as in 'pre' and 'ed' in the word 'precooked').

Phonemes
Phonemes are the minimal units of speech that distinguish between two words in a language. Thus, the onset sound in 'bit' versus that in 'pit' are different phonemes, as are the final sounds in 'bit' versus 'bid'.

Phonology
Phonology is the study of the representation and organization of phonemes and phoneme patterns in a language.

Phrasal level units
Phrasal level units are hierarchically structured clusters of words. For example, the sentence, 'the cat chased the mouse', can be decomposed into at least three phrasal units — 'the cat', 'chased the mouse' and 'the mouse' — that cluster together in a particular hierarchical arrangement.

Psycholinguistics
Psycholinguistics typically refers to the study of how language information is processed in real time during either comprehension or production. By contrast, linguistics typically refers to the study of the principles or representations that characterize all human languages.

Sibilants
A subtype of fricatives in which airflow is directed towards the sharp edges of the teeth, which are held close together. Examples from English include [s] and [z].

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 modelling 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.

Gregory Hickok is at the Department of Cognitive Sciences, University of California, Irvine, California 92697, USA.

e-mail: gshickok@uci.edu

doi:10.1038/nrn2158
Published online 5 January 2012

1. Dell, G. S. A spreading activation theory of retrieval in language production. *Psychol. Rev.* **93**, 283–321 (1986).
2. Fromkin, V. The non-anomalous nature of anomalous utterances. *Language* **47**, 27–52 (1971).
3. Levelt, W. J. Roelofs, A. & Meyer, A. S. A theory of lexical access in speech production. *Behav. Brain Sci.* **22**, 1–75 (1999).

4. Garrett, M. F. in *The Psychology of Learning and Motivation* Vol. 9 (ed. Bower, G. H.) 133–177 (Academic Press, New York, 1975).
5. Guenther, F. H., Hampson, M. & Johnson, D. A theoretical investigation of reference frames for the planning of speech movements. *Psychol. Rev.* **105**, 611–633 (1998).
6. Houde, J. F. & Jordan, M. I. Sensorimotor adaptation in speech production. *Science* **279**, 1213–1216 (1998).
7. Fairbanks, G. Systematic research in experimental phonetics. I. A theory of the speech mechanism as a servosystem. *J. Speech Hear. Disord.* **19**, 133–139 (1954).
8. Kawato, M. Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* **9**, 718–727 (1999).
9. Shadmehr, R. & Krakauer, J. W. A computational neuroanatomy for motor control. *Exp. Brain Res.* **185**, 359–381 (2008).
10. Shadmehr, R. & Mussa-Ivaldi, F. A. Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* **14**, 3208–3224 (1994).
11. Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. An internal model for sensorimotor integration. *Science* **269**, 1880–1882 (1995).
12. Tian, X. & Poeppel, D. Mental imagery of speech and movement implicates the dynamics of internal forward models. *Front. Psychol.* **1**, 166 (2010).
13. Grafton, S. T. The cognitive neuroscience of prehension: recent developments. *Exp. Brain Res.* **204**, 475–491 (2010).
14. Wolpert, D. M., Doya, K. & Kawato, M. A unifying computational framework for motor control and social interaction. *Phil. Trans. R. Soc. Lond. B* **358**, 593–602 (2003).
15. Perkell, J. S. *et al.* Speech motor control: acoustic goals, saturation effects, auditory feedback and internal models. *Speech Commun.* **22**, 227–250 (1997).
16. Burnett, T. A., Freedland, M. B., Larson, C. R. & Hain, T. C. Voice F0 responses to manipulations in pitch feedback. *J. Acoust. Soc. Am.* **103**, 3153–3161 (1998).
17. Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S. & Hain, T. C. Comparison of voice F0 responses to pitch-shift onset and offset conditions. *J. Acoust. Soc. Am.* **110**, 2845–2848 (2001).
18. Tourville, J. A., Reilly, K. J. & Guenther, F. H. Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* **39**, 1429–1443 (2008).
19. Tremblay, S., Shiller, D. M. & Ostry, D. J. Somatosensory basis of speech production. *Nature* **423**, 866–869 (2003).
20. Grafton, S. T., Aziz-Zadeh, L. & Ivry, R. B. in *The Cognitive Neurosciences* Ch. 44 (ed. Gazzaniga, M. S.) 641–652 (MIT Press, Cambridge, Massachusetts, USA, 2009).
21. Grafton, S. T. & Hamilton, A. F. Evidence for a distributed hierarchy of action representation in the brain. *Hum. Mov. Sci.* **26**, 590–616 (2007).
22. Diedrichsen, J., Shadmehr, R. & Ivry, R. B. The coordination of movement: optimal feedback control and beyond. *Trends Cogn. Sci.* **14**, 31–39 (2010).
23. Jackson, J. H. Remarks on evolution and dissolution of the nervous system. *J. Ment. Sci.* **33**, 25–48 (1887).
24. Gracco, V. L. Some organizational characteristics of speech movement control. *J. Speech Hear. Res.* **37**, 4–27 (1994).
25. Browman, C. P. & Goldstein, L. Articulatory phonology: an overview. *Phonetica* **49**, 155–180 (1992).
26. Plaut, D. C. & Kello, C. T. in *The Emergence of Language* Ch. 14 (ed. MacWhinney, B.) 381–416 (Lawrence Erlbaum Associates, Mahwah, New Jersey, USA, 1999).
27. Bock, K. in *The MIT Encyclopedia of the Cognitive Sciences* (eds Wilson, R. A. & Keil, F. C.) 453–456 (MIT Press, Cambridge, Massachusetts, USA, 1999).
28. Dell, G. S. in *An Invitation to Cognitive Science: Language* Ch. 7 (eds Glietman, L. R. & Liberman, M.) 183–208 (MIT Press, Cambridge, Massachusetts, USA, 1995).
29. Levelt, W. J. *Speaking: From Intention to Articulation* (MIT Press, Cambridge, Massachusetts, USA, 1989).
30. Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M. & Gagnon, D. A. Lexical access in aphasic and nonaphasic speakers. *Psychol. Rev.* **104**, 801–838 (1997).
31. Levelt, W. J. Models of word production. *Trends Cogn. Sci.* **3**, 223–232 (1999).
32. Vigliocco, G., Antonini, T. & Garrett, M. F. Grammatical gender is on the tip of Italian tongues. *Psychol. Sci.* **8**, 314–317 (1998).
33. Levelt, W. J. Monitoring and self-repair in speech. *Cognition* **14**, 41–104 (1983).
34. Nozari, N., Dell, G. S. & Schwartz, M. F. Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cogn. Psychol.* **63**, 1–33 (2011).
35. Oppenheim, G. M. & Dell, G. S. Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition* **106**, 528–537 (2008).
36. Postma, A. Detection of errors during speech production: a review of speech monitoring models. *Cognition* **77**, 97–132 (2000).
37. Huettig, F. & Hartsuiker, R. J. Listening to yourself is like listening to others: external, but not internal, verbal self-monitoring is based on speech perception. *Lang. Cognitive Proc.* **25**, 347–374 (2010).
38. Nickels, L. & Howard, D. Phonological errors in aphasic naming: comprehension, monitoring and lexicality. *Cortex* **31**, 209–237 (1995).
39. Ozdemir, R., Roelofs, A. & Levelt, W. J. Perceptual uniqueness point effects in monitoring internal speech. *Cognition* **105**, 457–465 (2007).
40. Golfinopoulos, E., Tourville, J. A. & Guenther, F. H. The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *Neuroimage* **52**, 862–874 (2010).
41. Hickok, G., Houde, J. & Rong, F. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* **69**, 407–422 (2011).
42. Houde, J. F. & Nagarajan, S. S. Speech production as state feedback control. *Front. Hum. Neurosci.* **5**, 82 (2011).
43. Hickok, G., Buchsbaum, B., Humphries, C. & Muftuler, T. Auditory–motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J. Cognitive Neurosci.* **15**, 673–682 (2003).
44. Hickok, G., Okada, K. & Serences, J. T. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J. Neurophysiol.* **101**, 2725–2732 (2009).
45. Howard, D. & Nickels, L. Separating input and output phonology: semantic, phonological, and orthographic effects in short-term memory impairment. *Cogn. Neuropsychol.* **22**, 42–77 (2005).
46. Jacquemot, C., Dupoux, E. & Bachoud-Levi, A. C. Breaking the mirror: asymmetrical disconnection between the phonological input and output codes. *Cogn. Neuropsychol.* **24**, 3–22 (2007).
47. Shelton, J. R. & Caramazza, A. Deficits in lexical and semantic processing: implications for models of normal language. *Psychon. Bull. Rev.* **6**, 5–27 (1999).
48. Ventura, M. I., Nagarajan, S. S. & Houde, J. F. Speech target modulates speaking induced suppression in auditory cortex. *BMC Neurosci.* **10**, 58 (2009).
49. Lichtheim, L. On aphasia. *Brain* **7**, 433–484 (1885).
50. Wernicke, C. in *Wernicke's Works on Aphasia: A Sourcebook and Review* (ed. Eggert, G. H.) 91–145 (Mouton, The Hague, The Netherlands, 1874/1977).
51. Gracco, V. L. & Lofqvist, A. Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. *J. Neurosci.* **14**, 6585–6597 (1994).
52. Perkell, J. S. Movement goals and feedback and feedback control mechanisms in speech production. *J. Neurolinguist.* 26 Mar 2010 (doi:10.1016/j.jneuroling.2010.02.011).
53. Wilson, S. M., Saygin, A. P., Sereno, M. I. & Iacoboni, M. Listening to speech activates motor areas involved in speech production. *Nature Neurosci.* **7**, 701–702 (2004).
54. Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* **15**, 399–402 (2002).
55. Cooper, W. E. & Lauritsen, M. R. Feature processing in the perception and production of speech. *Nature* **252**, 121–123 (1974).
56. Delvaux, V. & Soquet, A. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* **64**, 145–173 (2007).
57. Kappes, J., Baumgaertner, A., Peschke, C. & Ziegler, W. Unintended imitation in nonword repetition. *Brain Lang.* **111**, 140–151 (2009).
58. Christman, S. S., Boutsen, F. R. & Buckingham, H. W. Perseveration and other repetitive verbal behaviors: functional dissociations. *Semin. Speech Lang.* **25**, 295–307 (2004).
59. Duffy, J. R. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management* (Mosby, St. Louis, Missouri, USA, 1995).
60. Niemi, M., Laaksonen, J. P., Ojala, S., Aaltonen, O. & Happonen, R. P. Effects of transitory lingual nerve impairment on speech: an acoustic study of sibilant sound /s/. *Int. J. Oral Maxillofac. Surg.* **35**, 920–923 (2006).
61. Niemi, M., Laaksonen, J. P., Aaltonen, O. & Happonen, R. P. Effects of transitory lingual nerve impairment on speech: an acoustic study of diphthong sounds. *J. Oral Maxillofac. Surg.* **62**, 44–51 (2004).
62. Niemi, M. *et al.* Acoustic and neurophysiologic observations related to lingual nerve impairment. *Int. J. Oral. Maxillofac. Surg.* **38**, 758–765 (2009).
63. Perkell, J. S. *et al.* The distinctness of speakers' /s/-/S/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *J. Speech Lang. Hear. Res.* **47**, 1259–1269 (2004).
64. Liberman, A. M. Some results of research on speech perception. *J. Acoust. Soc. Am.* **29**, 117–123 (1957).
65. Indefrey, P. & Levelt, W. J. The spatial and temporal signatures of word production components. *Cognition* **92**, 101–144 (2004).
66. 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.* **71**, 491–507 (1988).
67. Rizzolatti, G. *et al.* Neurons related to reaching-grasping arm movements in the rostral part of area 6 (area 6aβ). *Exp. Brain Res.* **82**, 337–350 (1990).
68. Rizzolatti, G. *et al.* Neurons related to goal-directed motor acts in inferior area 6 of the macaque monkey. *Exp. Brain Res.* **67**, 220–224 (1987).
69. Peeva, M. G. *et al.* Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage* **50**, 626–638 (2010).
70. Aichert, I. & Ziegler, W. Syllable frequency and syllable structure in apraxia of speech. *Brain Lang.* **88**, 148–159 (2004).
71. Laganaro, M., Croisier, M., Bagou, O. & Assal, F. Progressive apraxia of speech as a window into the study of speech planning processes. *Cortex* 26 Mar 2011 (doi:10.1016/j.cortex.2011.03.010).
72. Ogar, J., Slama, H., Dronkers, N., Amici, S. & Gorno-Tempini, M. L. Apraxia of speech: an overview. *Neurocase* **11**, 427–432 (2005).
73. Hillis, A. E. *et al.* Re-examining the brain regions crucial for orchestrating speech articulation. *Brain* **127**, 1479–1487 (2004).
74. Dronkers, N. F. A new brain region for coordinating speech articulation. *Nature* **384**, 159–161 (1996).
75. Ogar, J. *et al.* Clinical and anatomical correlates of apraxia of speech. *Brain Lang.* **97**, 343–350 (2006).
76. Ito, M. Control of mental activities by internal models in the cerebellum. *Nature Rev. Neurosci.* **9**, 304–313 (2008).
77. Wolpert, D. M., Miall, R. C. & Kawato, M. Internal models in the cerebellum. *Trends Cogn. Sci.* **9**, 338–347 (1998).
78. Nowak, D. A., Topka, H., Timmann, D., Boecker, H. & Hermsdorfer, J. The role of the cerebellum for predictive control of grasping. *Cerebellum* **6**, 7–17 (2007).
79. Desmurget, M. & Grafton, S. Forward modeling allows feedback control for fast reaching movements. *Trends Cogn. Sci.* **4**, 423–431 (2000).
80. Pasalar, S., Roitman, A. V., Durfee, W. K. & Ebner, T. J. Force field effects on cerebellar Purkinje cell discharge with implications for internal models. *Nature Neurosci.* **9**, 1404–1411 (2006).
81. Shadmehr, R., Smith, M. A. & Krakauer, J. W. Error correction, sensory prediction, and adaptation in motor control. *Annu. Rev. Neurosci.* **33**, 89–108 (2010).
82. Baldo, J. V., Klostermann, E. C. & Dronkers, N. F. It's either a cook or a baker: patients with conduction aphasia get the gist but lose the trace. *Brain Lang.* **105**, 134–140 (2008).
83. Buchsbaum, B. R. *et al.* Conduction aphasia, sensory–motor integration, and phonological short-term memory — an aggregate analysis of lesion and fMRI data. *Brain Lang.* **119**, 119–128 (2011).
84. Damasio, H. & Damasio, A. R. The anatomical basis of conduction aphasia. *Brain* **103**, 337–350 (1980).
85. Goodglass, H. in *Conduction Aphasia* Ch. 3 (ed. Kohn, S. E.) 39–49 (Lawrence Erlbaum Associates, Hillsdale, New Jersey, USA, 1992).
86. Ackermann, H., Mathiak, K. & Riecker, A. The contribution of the cerebellum to speech production and speech perception: clinical and functional imaging data. *Cerebellum* **6**, 202–213 (2007).

87. Ackermann, H., Vogel, M., Petersen, D. & Poremba, M. Speech deficits in ischaemic cerebellar lesions. *J. Neurol.* **239**, 223–227 (1992).
88. Oppenheim, G. M. & Dell, G. S. Motor movement matters: the flexible abstractness of inner speech. *Mem. Cognit.* **38**, 1147–1160 (2010).
89. Buchsbaum, B., Hickok, G. & Humphries, C. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn. Sci.* **25**, 663–678 (2001).
90. Buchsbaum, B. R., Olsen, R. K., Koch, P. & Berman, K. F. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* **48**, 687–697 (2005).
91. Murphy, K. *et al.* Cerebral areas associated with motor control of speech in humans. *J. Appl. Physiol.* **83**, 1438–1447 (1997).
92. Shuster, L. I. & Lemieux, S. K. An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain Lang.* **93**, 20–31 (2005).
93. Smiley, J. F. *et al.* Multisensory convergence in auditory cortex. I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J. Comp. Neurol.* **502**, 894–923 (2007).
94. Schroeder, C. E. *et al.* Somatosensory input to auditory association cortex in the macaque monkey. *J. Neurophysiol.* **85**, 1322–1327 (2001).
95. Foxe, J. J. *et al.* Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res. Cogn. Brain Res.* **10**, 77–83 (2000).
96. Murray, M. M. *et al.* Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb. Cortex* **15**, 963–974 (2005).
97. Foxe, J. J. *et al.* Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *J. Neurophysiol.* **88**, 540–543 (2002).
98. Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A. & Schroeder, C. E. Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* **53**, 279–292 (2007).
99. Aliu, S. O., Houde, J. F. & Nagarajan, S. S. Motor-induced suppression of the auditory cortex. *J. Cogn. Neurosci.* **21**, 791–802 (2009).
100. Heinks-Maldonado, T. H. *et al.* Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Arch. Gen. Psychiatry* **64**, 286–296 (2007).
101. Frith, C. D., Blakemore, S. & Wolpert, D. M. Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Res. Brain Res. Rev.* **31**, 357–363 (2000).
102. Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J. & Evans, A. C. Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur. J. Neurosci.* **8**, 2236–2246 (1996).
103. Christoffels, I. K., van de Ven, V., Waldorp, L. J., Formisano, E. & Schiller, N. O. The sensory consequences of speaking: parametric neural cancellation during speech in auditory cortex. *PLoS ONE* **6**, e18307 (2011).
104. Eliades, S. J. & Wang, X. Sensory–motor interaction in the primate auditory cortex during self-initiated vocalizations. *J. Neurophysiol.* **89**, 2194–2207 (2003).
105. Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D. & Iacoboni, M. The essential role of premotor cortex in speech perception. *Curr. Biol.* **17**, 1692–1696 (2007).
106. D'Ausilio, A. *et al.* The motor somatotopy of speech perception. *Curr. Biol.* **19**, 381–385 (2009).
107. Callan, D. E., Jones, J. A., Callan, A. M. & 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* **22**, 1182–1194 (2004).
108. Wilson, S. M. & Iacoboni, M. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *Neuroimage* **33**, 316–325 (2006).
109. Jazayeri, M. & Movshon, J. A. Optimal representation of sensory information by neural populations. *Nature Neurosci.* **9**, 690–696 (2006).
110. Jazayeri, M. & Movshon, J. A. A new perceptual illusion reveals mechanisms of sensory decoding. *Nature* **446**, 912–915 (2007).
111. Regan, D. & Beverley, K. I. Postadaptation orientation discrimination. *J. Opt. Soc. Am. A* **2**, 147–155 (1985).
112. Scolari, M. & Serences, J. T. Adaptive allocation of attentional gain. *J. Neurosci.* **29**, 11933–11942 (2009).
113. Vigliocco, G. & Hartsuiker, R. J. The interplay of meaning, sound, and syntax in sentence production. *Psychol. Bull.* **128**, 442–472 (2002).
114. Friston, K. The free-energy principle: a unified brain theory? *Nature Rev. Neurosci.* **11**, 127–138 (2010).
115. Summerfield, C. & Egner, T. Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* **13**, 403–409 (2009).
116. 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. *Neurosci. Lett.* **287**, 156–160 (2000).
117. Anderson, J. M. *et al.* Conduction aphasia and the arcuate fasciculus: a reexamination of the Wernicke–Geschwind model. *Brain Lang.* **70**, 1–12 (1999).
118. Dronkers, N. & Baldo, J. in *Encyclopedia of Neuroscience* (ed. Squire, L. R.) 343–348 (Academic Press, Oxford, 2009).
119. Hickok, G. in *Language and the Brain* Ch. 4 (eds Grodzinsky, Y., Shapiro, L. & Swinney, D.) 87–104 (Academic Press, San Diego, California, USA, 2000).
120. Galantucci, B., Fowler, C. A. & Turvey, M. T. The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* **13**, 361–377 (2006).
121. Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. Perception of the speech code. *Psychol. Rev.* **74**, 431–461 (1967).
122. Hickok, G. The role of mirror neurons in speech perception and action word semantics. *Lang. Cognitive Proc.* **25**, 749–776 (2010).
123. Lotto, A. J., Hickok, G. S. & Holt, L. L. Reflections on mirror neurons and speech perception. *Trends Cogn. Sci.* **13**, 110–114 (2009).
124. Massaro, D. W. & Chen, T. H. The motor theory of speech perception revisited. *Psychon. Bull. Rev.* **15**, 453–462 (2008).
125. Liberman, A. M. & Mattingly, I. G. The motor theory of speech perception revised. *Cognition* **21**, 1–36 (1985).
126. Stevens, K. N. & Blumstein, S. E. Invariant cues for place of articulation in stop consonants. *J. Acoust. Soc. Am.* **64**, 1358–1368 (1978).
127. Stevens, K. N. & Halle, M. in *Models for the Perception of Speech and Visual Form* (ed. Walthen-Dunn, W.) 88–102 (MIT Press, Cambridge, Massachusetts, USA, 1967).
128. Nusbaum, H. C. & Magnuson, J. S. in *Talker Variability in Speech Processing* Ch. 6 (eds Johnson, K. & Mullennix, J. W.) 109–132 (Academic Press, San Diego, California, USA, 1997).
129. McClelland, J. L. & Elman, J. L. The TRACE model of speech perception. *Cogn. Psychol.* **18**, 1–86 (1986).
130. Massaro, D. W. in *Handbook of Psycholinguistics* Ch. 7 (ed. Gernsbacher, M. A.) 219–263 (Academic Press, San Diego, California, USA, 1994).
131. Vaden, K. I., 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.* **23**, 2665–2674 (2011).
132. Greenberg, S. in *Listening to Speech: An Auditory Perspective* Ch. 25 (eds Greenberg, S. & Ainsworth, W. A.) 411–433 (Erlbaum, Mahwah, New Jersey, USA, 2005).
133. Klatt, D. H. Speech perception: a model of acoustic-phonetic analysis and lexical access. *J. Phonetics* **7**, 279–312 (1979).
134. Johnson, K. The auditory/perceptual basis for speech segmentation. *OSU Work. Pap. Ling.* **50**, 101–113 (1997).
135. Johnson, K. Resonance in an exemplar-based lexicon: the emergence of social identity and phonology. *J. Phonetics* **34**, 485–499 (2006).
136. Goldinger, S. D. Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* **105**, 251–279 (1998).
137. Stevens, K. N. Toward a model for lexical access based on acoustic landmarks and distinctive features. *J. Acoust. Soc. Am.* **111**, 1872–1891 (2002).
138. Marslen-Wilson, W. D. Functional parallelism in spoken word-recognition. *Cognition* **25**, 71–102 (1987).

Competing interests statement
The author declares no competing interests.

Acknowledgements
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 US National Institutes of Health.

FURTHER INFORMATION
Gregory Hickok's homepage 1: http://alns.ss.uci.edu
Gregory Hickok's homepage 2: http://www.talkingbrains.org
ALL LINKS ARE ACTIVE IN THE ONLINE PDF