



Published in final edited form as:

*Curr Opin Neurobiol.* 2015 August ; 33: 174–181. doi:10.1016/j.conb.2015.04.006.

## The Cortical Computations Underlying Feedback Control in Vocal Production

John F. Houde<sup>1</sup> and Edward F. Chang<sup>2</sup>

<sup>1</sup>Department of Otolaryngology, University of California, San Francisco

<sup>2</sup>Department of Neurological Surgery, University of California, San Francisco

### Abstract

Recent neurophysiological studies of speaking are beginning to elucidate the neural mechanisms underlying auditory feedback processing during vocalizations. Here we review how research findings impact our state feedback control (SFC) model of speech motor control. We will discuss the evidence for cortical computations that compare incoming feedback with predictions derived from motor efference copy. We will also review observations from auditory feedback perturbation studies that demonstrate clear evidence for a state estimate correction process, which drives compensatory motor behavioral responses. While there is compelling support for cortical computations in the SFC model, there are still several outstanding questions that await resolution by future neural investigations.

### INTRODUCTION

When we speak we also hear ourselves. This auditory feedback is not only critical for speech learning and maintenance, but also for the online control of everyday speech. When sensory feedback is altered, we make immediate corrective adjustments to our speech to compensate for those changes. A speaker moves the articulators of his/her vocal tract (i.e., the lungs, larynx, tongue, jaw, and lips) so that an acoustic output is generated that is interpreted by a listener as the words the speaker intended to convey. In this review, we will focus on the prominent role of auditory feedback in speaking. For a number of years, we have explained this role using a model of speech motor control based on state feedback control (SFC) [1–11]

The SFC model explains a range of behavioral phenomena concerning speaking [9, 12], and other proposed models of speech production [13–20] can be described as special cases of SFC [21, 22]. Since its development, considerable new discoveries have been made about the neural substrate of auditory feedback processing during speaking. In this article, we consider how the findings from these recent studies impact our model.

---

*Corresponding author:* John Houde, houde@phy.ucsf.edu; Edward Chang, edward.chang@ucsf.edu, University of California, San Francisco, 505 Parnassus Ave, Room M779, San Francisco, CA 94143-0112, phone (415) 353-3904, fax (415) 353-3907.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

In our model (Fig 1), when a speaker is prompted to produce a speech sound, higher frontal cortex (IFG) responds by activating several speech control networks, including activating a speech motor control network (blue arrow in fig 1). This cortical network operates via state feedback control (SFC): During articulation, vPMC maintains a running estimate of the current articulatory state (orange in fig 1); this state carries multimodal information about current lip position, tongue body position, formant 1 (F1), formant 2 (F2), and any other parameter the CNS has learned is important to monitor for achieving correct production of the speech sound. M1 generates articulatory controls based on this state estimate, using a state feedback control law (state fb ctrl law in fig 1) that keeps the vocal tract tracking a desired state trajectory (e.g., one that produces the desired speech sound). The estimate of articulatory state is continually refined as articulation proceeds, with incoming sensory feedback from the vocal tract (both somatosensory and auditory feedback) being compared with feedforward sensory predictions (green arrows), generating feedback corrections (red arrows) to the state estimate. In turn, M1 makes use of the updated state estimate to generate further controls that move the estimated state closer to the desired articulatory state trajectory. This process continues until state trajectory generating the speech sound has been fully produced.

Our SFC model is derived from the general state feedback control framework used in optimal feedback control (OFC) models of motor behavior [5, 6, 10, 23, 24]. In this framework, control relies on state estimates furnished by recursive Bayesian filtering: motor efference copy and the previous state estimate determine a prior distribution of predicted next states, and this prior is then updated via Bayes rule using the likelihood of the current sensory feedback. This general form of Bayesian filtering lacks a direct comparison between incoming and predicted sensory feedback, which is notable because feedback comparison is the part of our SFC model's state correction process that allows our model to account for many of our empirical findings. Under linear Gaussian assumptions, however, the Bayesian filtering process reduces to exactly the feedback-comparison-based state correction process found in our SFC model [25].

In the sections that follow we consider what recent neural investigations tell us about how speaking is controlled, and how they impact our SFC model of speaking. We will conclude with brief discussion of some questions about our model that remain unresolved.

### **Neural evidence for auditory feedback processing during speaking**

A crucial window onto the control of speaking (or indeed any motor task) is found in examinations of the role of sensory feedback in the process. For speaking, the processing of auditory feedback is particularly important, because the most proximal goal of speaking is to create sounds. If done correctly, a listener will interpret meaning from these sounds. Thus, the feedback that a speaker can use to most directly monitor the correctness of his/her speech output comes via auditory input. It is not surprising, therefore, that studies have found a variety of neural phenomena indicating that speakers actively monitor their auditory feedback and modulate their ongoing speech motor output based on this feedback [26–43].

**Speaking-Induced Suppression (SIS)**—One of the first of the neurophysiological phenomena found associated with auditory feedback processing during speech was the phenomenon of Speaking-induced suppression (SIS): the response of a subject's auditory cortices to his/her own self-produced speech is significantly smaller than the response seen when the subject passively re-listens to playback of the same speech (see Figure 2). This effect, which we call speaking-induced suppression (SIS), has been seen using positron emission tomography (PET) [44–46], electroencephalography (EEG) [47, 48], and magnetoencephalography (MEG) [49–54]. An analog of the SIS effect has also been seen in non-human primates [55–57]. MEG experiments have shown that the SIS effect is only minimally explained by a general suppression of auditory cortex during speaking and that this suppression is not happening in the more peripheral parts of the CNS [52]. They have also shown that the observed suppression goes away if the subject's feedback is altered to mismatch his/her expectations [52, 53], as is consistent with some of the PET study findings.

These results are well accounted for by the hypothesized feedback comparison operation at the heart of our SFC model. The onset of speech is predicted from efference copy of motor output in the speaking condition, generating small prediction errors and a small auditory response. On the other hand, in the absence of an onset prediction, the same speech onset generates a large prediction error and a large auditory response during passive listening.

Subsequent to these initial studies, more recent studies have refined this SFC account of SIS. First, studies examining high-gamma responses using direct recordings with electrocorticography (ECoG) have found that SIS is not seen across areas of auditory cortex, but instead is localized to specific subsets of auditory responsive electrodes [58–60]. This cortical response heterogeneity contrasts greatly with the clear SIS effects seen in the M100 evoked response. These results suggest the reasonable possibility that not all areas of auditory cortex are devoted to processing auditory feedback for guiding speech motor control, or, equivalently, that SIS may be a marker of what specific areas of auditory cortex do process feedback for speech motor control.

Second, SIS varies with the natural trial-to-trial variability in repeated vowel productions[61]. Vowel productions whose initial formants deviated most from the median production showed the least SIS and those closest to the median production showed the most SIS. This pattern of “SIS falloff” was consistent with a feedback prediction representing the median production but not variations around this median. This suggested there are limits on the precision of the efference copy-derived predictions hypothesized in SFC to account for SIS, either because the mapping from motor commands to auditory expectations is imprecise, or because the sources of the noise generating the observed production variability are further downstream of the motor cortical outputs presumed to drive efference copy-based sensory predictions. Results of another recent study based on ECoG are consistent with this last point. Bouchard et al. found that, after coarticulatory effects were removed from the audio data, activity in sensorimotor cortex was able to predict a significant fraction of the trial-to-trial variance in vowel productions, but this fraction was modest, suggesting that cortical activity variation is not the only influence on vowel production variability [62].

The results of the Niziolek et al. study also provided indirect evidence for the action of auditory feedback control during speaking. In addition to being associated with less SIS, it was also found that those vowel productions deviating most from the median production underwent a process of “centering” after speech onset. That is, after speech onset, the formant tracks of these productions converged towards the formant tracks of the median production. Furthermore, it was found that, across subjects, this centering of deviating productions was significantly related to those productions’ reduced SIS [61].

**Speech Perturbation Response Enhancement (SPRE)**—More direct evidence of a role for auditory feedback processing in speaking can be seen in behavioral experiments where speakers compensate for artificial perturbations in their auditory feedback during speaking. Such compensatory responses have been seen in response to perturbations of speech amplitude [30, 31], and vowel formants [32, 33, 63], but the compensatory responses to perturbations of pitch (the so-called “pitch perturbation reflex”) have been the most thoroughly studied, both behaviorally [29, 42, 64–66] as well as in neurophysiological investigations [58, 67–70]. Like the studies of SIS, many of the neural investigations of pitch compensation have been based on evoked responses. In these studies, subjects phonated a steady pitch, and at some point in this phonation, the pitch of their audio feedback was suddenly shifted up or down. Using either EEG [67] or MEG [69], the auditory cortical response to this sudden pitch perturbation was recorded and compared with the auditory response recorded during passive listening to playback of the pitch-perturbed feedback. These studies found that, compared with passive listening, the auditory response to the perturbation during active phonation (speaking) was enhanced, which we call Speech Perturbation Response Enhancement (SPRE) (see Figure 2). This effect differed from SIS in several important ways. It had the opposite polarity of the SIS response (i.e., it was a response enhancement, not a response suppression), and it was principally seen not in the 100ms post onset peak of the evoked response (M100/N1), but instead in the later part of the response (M200/P2).

These characteristics can be accounted for in the SFC model. First, the model posits that SPRE does not arise from any speak/listen difference in the generation of auditory prediction errors. The model assumes that during speaking, the CNS predicts what it will hear based on efference copy of vocal motor commands which, in this case, are maintaining a steady pitch. The model assumes that during passive listening, the CNS predicts it will continue to hear what it has been hearing: a steady pitch. Thus, in both the speak and listen conditions, the prediction is the same (a steady pitch). As a result, onset of the perturbation generates the same size auditory prediction error in both conditions.

But the model also posits that auditory prediction errors are then passed back to higher auditory cortex (red arrow from A1 to vSMG/pSTS in Figure 1), where they are used to correct the current state estimate (red box labeled “state corr auditory” in Figure 1). Here is where the model posits a speak/listen difference. There is a gain associated with this state correction process (called the Kalman gain [1, 2, 71, 72]) that determines how strongly auditory prediction errors drive state corrections. This gain is set to reflect how correlated auditory feedback variations are with changes in the true articulatory state. During speaking, the CNS can indirectly estimate this correlation from the correlation between auditory

feedback and somatosensory feedback (i.e., it can use somatosensory feedback as a noisy measure of the true articulatory state). The CNS then sets the state correction gain using this estimate. But during passive listening, there is no measure of the true articulatory state to correlate with auditory input. In this case, we posit that, without any way to estimate it, the CNS conservatively assumes a low value for the correlation, and sets the state correction gain correspondingly low. This lower gain means that, for the same size auditory prediction error, smaller state corrections are generated during listening than during speaking. In other words, during speaking, there is an enhanced state correction response to the perturbation (SPRE), which we would see reflected not in early (100ms post-perturbation) activity related to auditory prediction errors, but instead in the later activity of higher auditory cortex where state corrections are generated. It is therefore not surprising that the SPRE effect is seen not in M100/N1 responses, but instead in the later M200/P2 responses [67, 69]. Also consistent with this is evidence from the MEG study of SPRE showing that the effect is strongest in higher levels of temporal cortex [69].

In these initial investigations of SPRE, the effect is seen as a dominant feature of the evoked response to the pitch perturbation, but in subsequent investigations based on ECoG [58, 70], the reflection of the effect in high gamma power changes was seen to be more complicated. Not all electrode sites exhibited SPRE, and of those sites that did exhibit SPRE, many of them did not express SIS. In addition, several of the sites expressing SIS did not also express SPRE. The fact that not all sites showed SPRE is easily explained as reflecting the fact that not all areas of auditory cortex are devoted to feedback control of speaking, and the sites showing SIS but not SPRE can also be accounted for in the SFC model where the feedback comparison operation that generates SIS is separate from the state correction operation that generates SPRE. However, the model does predict that since the feedback comparison operation feeds into the state correction operation, all sites expressing SPRE should also express SIS. That this is not the case is a challenge for our original SFC formulation (and for the many models that are variants of SFC), and suggests the possibility that some state corrections could be based on feedback comparison operations that use predictions not derived from efference copy, thus don't express SIS.

Regardless of these variations of SPRE/SIS characteristics, our SFC model predicts that the SPRE should be most directly associated with feedback control of speaking, and indeed the Chang et al. study found evidence that this is the case. SPRE (as measured by the difference between speaking and listening responses to the pitch perturbation) was correlated with compensation across trials and SPRE significantly predicted the amount of compensation, whereas SIS did not [58]. SFC model predicts this: The SPRE expressing part of the model (state correction) is dependent on the state correction gain, while the SIS part (feedback comparison) is not. The state correction gain is postulated to be dynamically estimated on-line, so it varies a bit from trial to trial in the experiment. That trial-to-trial variability is expressed not in the SIS part of the model, but instead in the activity of the SPRE-expressing part, as well as in the downstream compensatory motor responses driven by the state correction.

More recent studies have elaborated our picture of the neural correlates of the pitch perturbation reflex, and these have, in turn, helped to elaborate further the details of our SFC

model. A recent study used magnetoencephalographic imaging (MEGI) to look at the high gamma responses to the pitch perturbation, and found that the dominant response to the perturbation was in the right hemisphere, in premotor cortex and the supramarginal gyrus [73]. Further, the study found that this right hemisphere activity was linked to the left via a dynamically changing pattern of functional connectivity. This result is consistent with that seen in a prior study of responses to formant perturbations [33] that also found right hemisphere involvement in perturbation responses. Taken together, these results imply that our SFC model has a neural substrate that is distributed between the two hemispheres, with the MEGI data suggesting that the early responding left hemisphere primarily detects feedback prediction mismatches, while the later responding right hemisphere is more involved in generating state corrections from the feedback prediction errors.

**Unresolved questions about the SFC model**—In sum, our SFC model accounts for much of what has been recently learned about the neural substrate of auditory feedback processing in speaking. Nevertheless, there remain several unresolved issues concerning the structure of the model and its neural instantiation. Central to the model are the motor-to-sensory mappings used to generate feedback predictions from speech motor activity. However little is known about the nature or neural substrate of these mappings, or the neural mechanisms by which their outputs are compared with incoming feedback. Are the mappings represented in frontal areas like vPMC, or instead in sensory areas like pSTG and SII? For each sense modality, is there a single mapping that's shared in the production of all speech sounds, or are there separate mappings for each speech sound's production? Recent ECoG studies show great differences in how speech features are organized in sensorimotor cortex [74] and auditory cortex [75] which may be difficult to reconcile with a single shared mapping. There are also studies showing that altering the audiomotor mapping in one word's production doesn't generalize to other words [39], suggesting individual mappings. If this were the case, then different words' productions might be controlled by separate SFC-based speech motor control networks (see Fig 1). Another issue concerns the structure of the SFC model. Recent ECoG studies and stimulation studies have also found a blurring of the distinction between primary motor (M1) and sensory (S1) areas around the central sulcus [74], suggesting a tight coupling between the two areas. This suggests the possibility of a control hierarchy, where M1 and S1 function as a low-level controller of articulatory movements, which in turn is controlled by a higher level SFC-based speech motor control network integrating both somatosensory and auditory feedback [20, 76]. These issues must be resolved in future investigations of the neural substrate of speech motor control.

## Acknowledgements

This work was supported by NIH Grants R01RDC010145 (JH), DP2OD00862 (EC), and NSF Grant BCS-0926196 (JH). EFC was also supported by Bowes Foundation, Curci Foundation, and McKnight Foundation. Edward Chang is a New York Stem Cell Foundation - Robertson Investigator. This research was supported by The New York Stem Cell Foundation.

## REFERENCES

1. Stengel, RF. Optimal Control and Estimation. Mineola, NY: Dover Publications, Inc; 1994.
2. Jacobs, OLR. Introduction to Control Theory. edn 2nd. Oxford, UK: Oxford University Press; 1993.

3. Tin C, Poon C-S. Internal models in sensorimotor integration: perspectives from adaptive control theory. *Journal of Neural Engineering*. 2005; 2:S147–S163. [PubMed: 16135881]
4. Wolpert DM. Computational approaches to motor control. *Trends in Cognitive Sciences*. 1997; 1:209. [PubMed: 21223909]
5. Todorov E. Optimality principles in sensorimotor control. *Nature Neuroscience*. 2004; 7:907–915. [PubMed: 15332089] \*\* In this often-cited review article, the main principles of optimal state feedback control are described, along with how it explains some key characteristics of movements not easily explained by other modelling frameworks.
6. Todorov E, Jordan MI. Optimal feedback control as a theory of motor coordination. *Nat Neurosci*. 2002; 5:1226–1235. [PubMed: 12404008]
7. Shadmehr R, Krakauer JW. A computational neuroanatomy for motor control. *Experimental Brain Research*. 2008; 185:359–381. [PubMed: 18251019]
8. Guigon E, Baraduc P, Desmurget M. Optimality, stochasticity, and variability in motor behavior. *J Comput Neurosci*. 2008; 24:57–68. [PubMed: 18202922]
9. Houde JF, Nagarajan SS. Speech production as state feedback control. *Frontiers in Human Neuroscience*. 2011; 5:82. [PubMed: 22046152] \* In this article, the state feedback control (SFC) model of speech motor control is described in detail. The motivation for SFC is developed from consideration of the role and limitations of the higher levels of the central nervous system (cortex, basal ganglia, and cerebellum) in sensory feedback control of speaking.
10. Scott SH. The computational and neural basis of voluntary motor control and planning. *Trends In Cognitive Sciences*. 2012; 16:541–549. [PubMed: 23031541]
11. Wolpert DM, Ghahramani Z, Jordan MI. An internal model for sensorimotor integration. *Science*. 1995; 269:1880–1882. [PubMed: 7569931] \*\* This article reports one of the first studies providing evidence that the CNS maintains and updates an estimate of the state of the system being controlled (in this case, the hand) during movement.
12. Hickok G, Houde JF, Rong F. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*. 2011; 69:407–422. [PubMed: 21315253]
13. Guenther FH. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*. 1995; 102:594–621. [PubMed: 7624456]
14. 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] \* This article is one of the earliest descriptions of the DIVA (directions into velocities of articulators) model of speech motor control, which also makes the case for speech goals being represented in acoustic dimensions.
15. Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain And Language*. 2006; 96:280–301. [PubMed: 16040108] \*\* This article describes in detail the DIVA model of speech motor control as it is currently realized, with parallel speech motor control systems (feedforward and feedback control subsystems) converging in motor cortex to generate desired articulatory trajectories.
16. Guenther FH, Vladusich T. A Neural Theory of Speech Acquisition and Production. *J Neurolinguistics*. 2012; 25:408–422. [PubMed: 22711978]
17. Tian X, Poeppel D. Mental imagery of speech and movement implicates the dynamics of internal forward models. *Front Psychol*. 2010; 1:166. [PubMed: 21897822]
18. Price CJ, Crinion JT, Macsweeney M. A Generative Model of Speech Production in Broca's and Wernicke's Areas. *Frontiers in Psychology*. 2011; 2:237. [PubMed: 21954392]
19. Hickok G. The architecture of speech production and the role of the phoneme in speech processing. *Lang Cogn Process*. 2014; 29:2–20. [PubMed: 24489420]
20. Hickok G. Computational neuroanatomy of speech production. *Nat Rev Neurosci*. 2012; 13:135–145. [PubMed: 22218206] \* In this article, the case is made for a hierarchy in the control of speaking, with a lower phonemic level control system being in turn controlled by a higher-level syllable control system. A key feature of the model is the claim that auditory feedback interacts primarily with the syllable-level controller.
21. Houde JF, Kort NS, Niziolek CA, Chang EF, Nagarajan SS. Neural evidence for state feedback control of speaking. *Proceedings of Meetings on Acoustics*. 2013; 19

22. Houde, JF.; Niziolek, C.; Kort, N.; Agnew, Z.; Nagarajan, SS. 10th International Seminar on Speech Production May 5–8, 2014. Cologne, Germany: 2014. Simulating a state feedback model of speaking; p. 202-205.
23. Scott SH. Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*. 2004; 5:534–546.
24. Kording KP, Wolpert DM. Bayesian decision theory in sensorimotor control. *Trends Cogn Sci*. 2006; 10:319–326. [PubMed: 16807063]
25. Särkkä, S. Bayesian filtering and smoothing. Cambridge, U.K.; New York: Cambridge University Press; 2013.
26. Chang-Yit R, Pick J, Herbert L, Siegel GM. Reliability of sidetone amplification effect in vocal intensity. *Journal of Communication Disorders*. 1975; 8:317–324. [PubMed: 802981]
27. Elman JL. Effects of frequency-shifted feedback on the pitch of vocal productions. *J Acoust Soc Am*. 1981; 70:45–50. [PubMed: 7264071]
28. Kawahara H. Transformed auditory feedback: Effects of fundamental frequency perturbation. *J Acoust Soc Am*. 1993; 94:1883.
29. Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America*. 1998; 103:3153–3161. [PubMed: 9637026] \*\* In this article, the basic characteristics of the pitch perturbation reflex (speakers' compensatory responses to perturbations of the perceived pitch of their ongoing speech) are first described in detail.
30. Bauer JJ, Mittal J, Larson CR, Hain TC. Vocal responses to unanticipated perturbations in voice loudness feedback: an automatic mechanism for stabilizing voice amplitude. *J Acoust Soc Am*. 2006; 119:2363–2371. [PubMed: 16642849]
31. Heinks-Maldonado TH, Houde JF. Compensatory responses to brief perturbations of speech amplitude. *Acoustics Research Letters Online*. 2005; 6:131–137.
32. Purcell DW, Munhall KG. Compensation following real-time manipulation of formants in isolated vowels. *Journal of the Acoustical Society of America*. 2006; 119:2288–2297. [PubMed: 16642842] \* This article describes the first study showing that speakers quickly compensate for perturbations of the formants of their auditory feedback of their ongoing speech, in a manner similar to the pitch perturbation reflex.
33. Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*. 2008; 39:1429–1443. [PubMed: 18035557] \*\* This article reports on an fMRI-based study of the neural activity associated with speakers' compensatory responses to formant feedback perturbations. It was one of the first studies to reveal the network of areas in the CNS associated with auditory feedback processing during speaking.
34. Houde JF, Jordan MI. Sensorimotor adaptation in speech production. *Science*. 1998; 279:1213–1216. [PubMed: 9469813] \*\* This is the original study showing that speaking exhibits sensorimotor adaptation, where repeated exposure to formant-altered auditory feedback causes speakers to make long-term compensatory adjustments to their speech.
35. Purcell DW, Munhall KG. Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *Journal of the Acoustical Society of America*. 2006; 120:966–977. [PubMed: 16938984]
36. Katseff S, Houde J, Johnson K. Partial compensation for altered auditory feedback: a tradeoff with somatosensory feedback? *Language and Speech*. 2012; 55:295–308. [PubMed: 22783636]
37. Lametti DR, Nasir SM, Ostry DJ. Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*. 2012; 32:9351–9358. [PubMed: 22764242] \*\* This study showed that, across subjects, sensitivity to somatosensory feedback perturbations was inversely related to sensitivity to auditory feedback perturbations. The result is a fundamental step forward in explaining the variability across speakers seen in responses to altered sensory feedback.
38. MacDonald EN, Johnson EK, Forsythe J, Plante P, Munhall KG. Children's development of self-regulation in speech production. *Curr Biol*. 2012; 22:113–117. [PubMed: 22197241] \* This study showed that even very young children exhibit sensorimotor adaptation in their speech.



39. Rochet-Capellan A, Richer L, Ostry DJ. Nonhomogeneous transfer reveals specificity in speech motor learning. *J Neurophysiol.* 2012; 107:1711–1717. [PubMed: 22190628] \* This extensive study employing many subjects, supported the claim that sensorimotor adaptation of speech exhibits very little generalization to novel contexts. The finding is consistent with the hypothesis that the representations controlling the production of different utterances are largely independent of each other.
40. Villacorta VM, Perkell JS, Guenther FH. Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *Journal of The Acoustical Society of America.* 2007; 122:2306–2319. [PubMed: 17902866]
41. Jones JA, Munhall KG. Perceptual calibration of F0 production: evidence from feedback perturbation. *Journal of the Acoustical Society of America.* 2000; 108:1246–1251. [PubMed: 11008824]
42. Liu P, Chen Z, Jones JA, Huang D, Liu H. Auditory feedback control of vocal pitch during sustained vocalization: a cross-sectional study of adult aging. *PLoS ONE.* 2011; 6:e22791. [PubMed: 21799942] \* This study showed that, contrary to expectations, the magnitude of compensations speakers produce in response to pitch feedback perturbations actually increases with age, up to a point.
43. Shiller DM, Sato M, Gracco VL, Baum SR. Perceptual recalibration of speech sounds following speech motor learning. *Journal of the Acoustical Society of America.* 2009; 125:1103–1113. [PubMed: 19206885] \*\* This study two significant findings: first, that the production of fricatives exhibits sensorimotor adaptation, and second, and more importantly, that exposure to altered sensory feedback during speaking causes changes not only in speech production, but also in speech perception.
44. Hirano S, Naito Y, Okazawa H, Kojima H, Honjo I, Ishizu K, Yenokura Y, Nagahama Y, Fukuyama H, Konishi J. Cortical activation by monaural speech sound stimulation demonstrated by positron emission tomography. *Exp Brain Res.* 1997; 113:75–80. [PubMed: 9028776]
45. Hirano S, Kojima H, Naito Y, Honjo I, Kamoto Y, Okazawa H, Ishizu K, Yonekura Y, Nagahama Y, Fukuyama H, et al. Cortical processing mechanism for vocalization with auditory verbal feedback. *Neuroreport.* 1997; 8:2379–2382. [PubMed: 9243644] \* This is the first neuroimaging study to report an effect consistent with speaking-induced suppression (SIS). The study found auditory cortical activation during speaking only when the speaker's audio feedback was altered.
46. Hirano S, Kojima H, Naito Y, Honjo I, Kamoto Y, Okazawa H, Ishizu K, Yonekura Y, Nagahama Y, Fukuyama H, et al. Cortical speech processing mechanisms while vocalizing visually presented languages. *Neuroreport.* 1996; 8:363–367. [PubMed: 9051811]
47. Ford JM, Mathalon DH. Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal Of Psychiatric Research.* 2004; 38:37–46. [PubMed: 14690769]
48. Ford JM, Mathalon DH, Heinks T, Kalba S, Faustman WO, Roth WT. Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *American Journal of Psychiatry.* 2001; 158:2069–2071. [PubMed: 11729029] \* This study not only demonstrated the speaking-induced suppression (SIS) effect in the auditory N1 response using EEG, but also showed that Schizophrenic patients had an impaired SIS effect.
49. Numminen J, Curio G. Differential effects of overt, covert and replayed speech on vowel- evoked responses of the human auditory cortex. *Neuroscience Letters.* 1999; 272:29–32. [PubMed: 10507535]
50. Numminen J, Salmelin R, Hari R. Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters.* 1999; 265:119–122. [PubMed: 10327183]
51. Curio G, Neuloh G, Numminen J, Jousmaki V, Hari R. Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping.* 2000; 9:183–191. [PubMed: 10770228] \* This was the first study to demonstrate the SIS effect in the auditory M100 response via magnetoencephalography (MEG).
52. Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the auditory cortex during speech: An MEG study. *J Cogn Neurosci.* 2002; 14:1125–1138. [PubMed: 12495520] \* This article reported on a series of three MEG experiments that supported the claim that SIS arises from auditory feedback being compared with a motor-efference-derived feedback prediction. The

experiments confirmed the SIS effect, showed SIS is not completely explained by general auditory suppression, and showed SIS is abolished when feedback is greatly altered.

53. Heinks-Maldonado TH, Nagarajan SS, Houde JF. Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport*. 2006; 17:1375–1379. [PubMed: 16932142]
54. Ventura MI, Nagarajan SS, Houde JF. Speech target modulates speaking induced suppression in auditory cortex. *BMC Neuroscience*. 2009; 10:58. [PubMed: 19523234]
55. Eliades SJ, Wang X. Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cereb Cortex*. 2005; 15:1510–1523. [PubMed: 15689521]
56. Eliades SJ, Wang X. Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J Neurophysiol*. 2003; 89:2194–2207. [PubMed: 12612021] \*\* This is the first study to show the SIS effect in non-human primate vocalizations.
57. Eliades SJ, Wang X. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*. 2008; 453:1102–1106. [PubMed: 18454135] \*\* This is the first study to show in a non-human primate what is one element of the speech perturbation response enhancement (SPRE) effect: auditory neurons that respond only when the monkey's vocal feedback was altered.
58. Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF. Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences*. 2013; 110:2653–2658. \*\* This first ECoG study of the speech perturbation response enhancement (SPRE) effect had a number of important findings relevant to modeling the control of speech, chief among them being the significant close relationship between electrodes exhibiting SPRE and their correlation with vocal compensation responses to the pitch feedback perturbations.
59. Greenlee JD, Jackson AW, Chen F, Larson CR, Oya H, Kawasaki H, Chen H, Howard MA. Human Auditory Cortical Activation during Self-Vocalization. *PLoS One*. 2011; 6:e14744. [PubMed: 21390228]
60. Flinker A, Chang EF, Kirsch HE, Barbaro NM, Crone NE, Knight RT. Single-trial speech suppression of auditory cortex activity in humans. *J Neurosci*. 2010; 30:16643–16650. [PubMed: 21148003]
61. Niziolek CA, Nagarajan SS, Houde JF. What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience*. 2013; 33:16110–16116. [PubMed: 24107944] \*\* This article reports on a study of the relationship between SIS and trial-to-trial variability in produced vowel formants. The main finding was that SIS was lower for trials whose formants deviated from the median production, thus demonstrating limits to the precision of the putative motor-efference-derived feedback predictions.
62. Bouchard KE, Chang EF. Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex. *J Neurosci*. 2014; 34:12662–12677. [PubMed: 25232105] \* This study correlated ECoG electrode activity with vowel pitch and formants, and found evidence for coarticulation at the level of sensorimotor cortex (SMC), as well as possible limitations to how tightly SMC controls trial-to-trial variability in vowel productions.
63. Niziolek CA, Guenther FH. Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *J Neurosci*. 2013; 33:12090–12098. [PubMed: 23864694]
64. Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK. Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Exp Brain Res*. 2000; 130:133–141. [PubMed: 10672466]
65. 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; 110:2845–2848. [PubMed: 11785786]
66. Larson CR, Altman KW, Liu HJ, Hain TC. Interactions between auditory and somatosensory feedback for voice F0 control. *Experimental Brain Research*. 2008; 187:613–621. [PubMed: 18340440] \* This study showed that the auditory pitch perturbation response was enhanced when somatosensory feedback was reduced, providing support for the hypothesis that mismatch with unaltered somatosensory feedback limits the size of the compensatory response to pitch feedback perturbations.
67. Behroozmand R, Karvelis L, Liu H, Larson CR. Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. *Clin Neurophysiol*. 2009; 120:1303–1312. [PubMed: 19520602] \* This EEG study was the first to demonstrate the speech perturbation enhancement (SPRE) effect in humans.

68. Parkinson AL, Flagmeier SG, Manes JL, Larson CR, Rogers B, Robin DA. Understanding the neural mechanisms involved in sensory control of voice production. *NeuroImage*. 2012; 61:314–322. [PubMed: 22406500]
69. Kort NS, Nagarajan SS, Houde JF. A bilateral cortical network responds to pitch perturbations in speech feedback. *Neuroimage*. 2014; 86:525–535. [PubMed: 24076223] \* This is the first study to demonstrate the SPRE effect using MEG. It also showed that the largest SPRE effect occurred bilaterally in higher level regions of auditory cortex.
70. Greenlee JD, Behroozmand R, Larson CR, Jackson AW, Chen F, Hansen DR, Oya H, Kawasaki H, Howard MA 3rd. Sensory-motor interactions for vocal pitch monitoring in non-primary human auditory cortex. *PLoS One*. 2013; 8:e60783. [PubMed: 23577157]
71. Todorov, E. Optimal control theory. In: Doya, K.; Ishii, S.; Pouget, A., editors. *Bayesian Brain: Probabilistic Approaches to Neural Coding*. Rao RPN: MIT Press; 2006. p. 269-298.
72. Kalman RE. A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*. 1960; 82:35–45.
73. Kort N, Nagarajan SS, Houde JF. A right-lateralized cortical network drives error correction to voice pitch feedback perturbation. *J Acoust Soc Am*. 2013; 134:4234.
74. Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. *Nature*. 2013; 495:327–332. [PubMed: 23426266]
75. Mesgarani N, Cheung C, Johnson K, Chang EF. Phonetic feature encoding in human superior temporal gyrus. *Science*. 2014; 343:1006–1010. [PubMed: 24482117]
76. Todorov E, Li W, Pan X. From task parameters to motor synergies: A hierarchical framework for approximately optimal control of redundant manipulators. *Journal of Robotic Systems*. 2005; 22:691–710. [PubMed: 17710121]

**BOX 1-MAIN CONCEPTS**

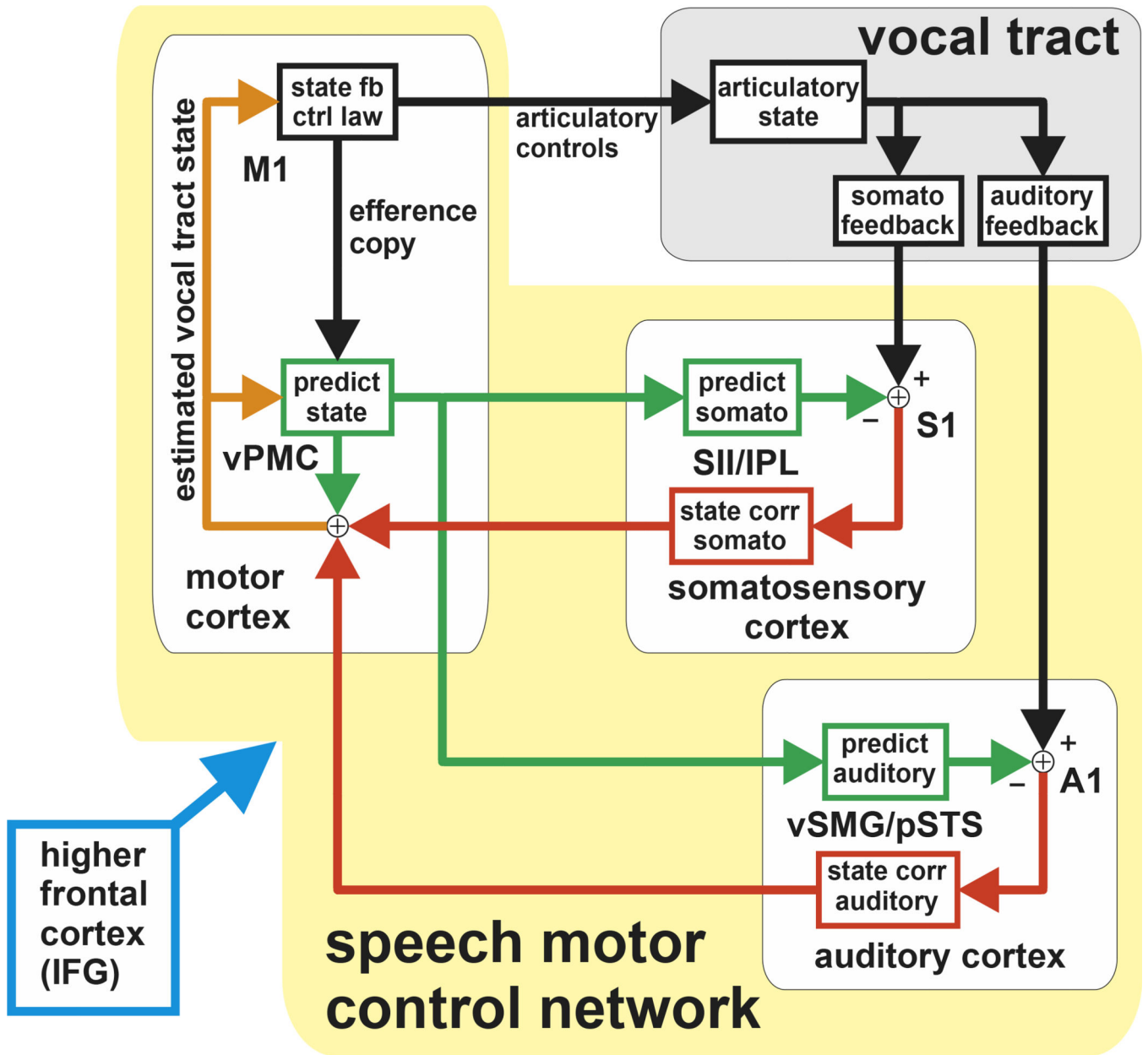
1. In the state feedback control (SFC) model, speaking is controlled using a running estimate of the current vocal tract state; this estimate is updated by comparing incoming sensory feedback with feedback predictions derived from motor efference copy.
2. The SFC model explains why auditory cortex is suppressed when listening to one's own speech (speaking-induced suppression, or SIS) but enhanced when a feedback perturbation is perceived during speaking (speech perturbation response enhancement, or SPRE).
3. The degree of suppression in SIS is reduced in utterance productions that deviate from the median production, implying that the accuracy of feedback predictions derived from motor efference copy is limited.
4. When auditory feedback is perturbed, the activity of areas in auditory cortex that express SPRE is more correlated with behavioral compensation than activity in other auditory areas; this result is predicted by the SFC model.

**UNRESOLVED QUESTIONS**

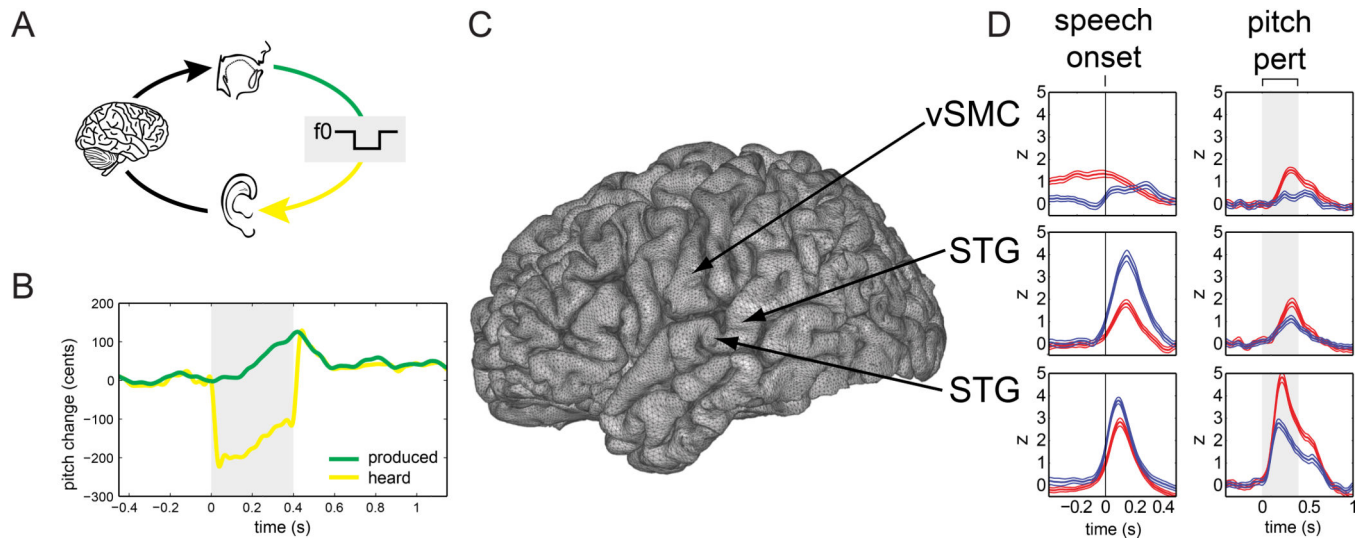
1. What are the neural mechanisms by which incoming sensory feedback is compared with feedback predictions?
2. What is the nature and neural substrate of the motor-to-sensory mappings required for SFC?
3. Is there neural evidence for a hierarchical organization of the speech motor control networks postulated by SFC?

**HIGHLIGHTS (Houde and Chang)**

1. State feedback control (SFC) model is described for speech motor control.
2. SFC explains auditory suppression/enhancement depending on motor predictions.
3. SFC confirmed by auditory-driven vocal compensation during feedback perturbation.



**Figure 1.** A model of speech motor control based on state feedback control (SFC). In the model, articulatory controls sent to the vocal tract from M1 are based on an estimate of the current vocal tract state (orange arrows) that is maintained by an interaction between vPMC and the sensory cortices. In this interaction, feedback predictions (green arrows) are compared with incoming feedback (black arrows), generating corrections to the state estimate (red arrows). See text for details.



**Figure 2.**

Examples of SIS and SPRE during pitch perturbation of vocalization. (a) A DSP shifted the pitch of subjects' vocalizations (green line) and delivered this auditory feedback (yellow line) to subjects' earphones. (b) Pitch track of an example trial. The green line shows the pitch recorded by the microphone (produced) and the yellow line shows the pitch delivered to the earphones (heard). Shaded region shows time interval when DSP shifted pitch by 200 cents (1/6 octave). (c) Location of three electrodes on the cortical surface. (d) High-gamma line plots for each electrode in the speak (red) and listen (blue) conditions, with vertical lines in the left column of plots representing speech onset (where SIS [speak response < listen response] is observed) and shaded regions in the right column of plots representing perturbation onset and offset (where SPRE [speak response > listen response] is observed). Adapted from Chang et al., PNAS 2011.