The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model

Gregory Hickok

Department of Cognitive Sciences, University of California, Irvine, Irvine, CA 92697, Greg.hickok@uci.edu

Abstract

Speech recognition is an active process that involves some form of predictive coding. This statement is relatively uncontroversial. What is less clear is the source of the prediction. The dual-stream model of speech processing suggests that there are two possible sources of predictive coding in speech perception: the motor speech system and the lexical-conceptual system. Here I provide an overview of the dual-stream model of speech processing and then discuss evidence concerning the source of predictive coding during speech recognition. I conclude that, in contrast to recent theoretical trends, the dorsal sensory-motor stream is not a source of forward prediction that can facilitate speech recognition. Rather, it is forward prediction coming out of the ventral stream that serves this function.

Learning outcomes—Readers will (1) be able to explain the dual route model of speech processing including the function of the dorsal and ventral streams in language processing, (2) understand how disruptions to certain components of the dorsal stream can cause conduction aphasia, (3) be able to explain the fundamental principles of state feedback control in motor behavior, and (4) understand the role of predictive coding in motor control and in perception and how predictive coding coming out of the two streams may have different functional consequences.

1. Introduction

The dual route model of the organization of speech processing is grounded in the fact that the brain has to do two kinds of things with acoustic speech information. On the one hand, acoustic speech input must be linked to conceptual-semantic representations; that is, speech must be understood. On the other hand, the brain must to link acoustic speech information to the motor speech system; that is, speech sounds must be reproduced with the vocal tract. These are different computational tasks. The set of operations involved in translating a sound pattern, [kæt], into a distributed representation corresponding to the meaning of that word [domesticated animal, carnivore, furry, purrs] must be non-identical to the set of operations involved in translating that same sound pattern into a set of motor commands [velar, plosive, unvoiced; front, voiced, etc.]. The endpoint representations are radically different; therefore the set of computations in the two types of transformations must be different. The neural pathways involved must also, therefore, be non-identical. The separability of these two pathways is demonstrated by the fact that (i) we can easily translate sound into motor speech without linking to the conceptual system as when we repeat a

pseudoword and (ii) inability to produce speech as a result of acquired or congenital neurological disease or temporary deactivation of the motor speech system does not preclude the ability to comprehension spoken language (Bishop, Brown, & Robson, 1990; Hickok, in press; Hickok, Costanzo, Capasso, & Miceli, 2011; Hickok, Houde, & Rong, 2011; Hickok, et al., 2008; Rogalsky, Love, Driscoll, Anderson, & Hickok, 2011).

These facts are not new to language science – they were the foundations of Wernicke's model of the neurology of language (Wernicke, 1874/1977) – nor are they unique to language. In the last two decades, a similar dual-stream model has been developed in the visual processing domain (Milner & Goodale, 1995). The idea has thus stood the test of time and has general applicability in understanding higher brain function.

In this article I have two goals: One is to summarize the dual route model for speech processing as it is currently understood, and the other is to consider how this framework might be useful for thinking about predictive coding in neural processing. Predictive coding is a notion that has received a lot of attention both in motor control and in perceptual processing, i.e., in the two processing streams, suggesting that a dual-stream framework for predictive coding may be useful in understanding possibly different forms of predictive coding.

2. The dual route model of speech processing

The dual-route model (Fig. 1) holds that a ventral stream, which involves structures in the superior and middle portions of the temporal lobe, is involved in processing speech signals for comprehension. A dorsal stream, which involves structures in the posterior planum temporale region and posterior frontal lobe, is involved in translating acoustic speech signals into articulatory representations, which are essential for speech production. In contrast to the typical view that speech processing is mainly left hemisphere dependent, a wide range of evidence suggests that the ventral stream is bilaterally organized (although with important computational differences between the two hemispheres). The dorsal stream, on the other hand, is strongly leftdominant.

2.1. Ventral Stream: Mapping from sound to meaning

2.1.1. Bilateral organization and parallel computation—The ventral stream is bilaterally organized, although not computationally redundant in the two hemispheres. Evidence for bilateral organization comes from the observation that chronic or acute disruption of the left hemisphere due to stroke (Baker, Blumsteim, & Goodglass, 1981; Rogalsky, et al., 2011; Rogalsky, Pitz, Hillis, & Hickok, 2008) or functional deactivation in Wada procedures (Hickok, et al., 2008) does not result in a dramatic decline in the ability of patients to processes speech sound information during comprehension (Hickok & Poeppel, 2000, 2004, 2007). Bilateral lesions involving the superior temporal lobe do, however, result in severe speech perception deficits (Buchman, Garron, Trost-Cardamone, Wichter, & Schwartz, 1986; Poeppel, 2001).

Data from neuroimaging has been more controversial. One consistent and uncontroversial finding is that, when contrasted with a resting baseline, listening to speech activates the superior temporal gyrus (STG) bilaterally including the dorsal STG and superior temporal sulcus (STS). However, when listening to speech is contrasted against various acoustic baselines, some studies have reported left-dominant activation patterns that particularly implicate anterior temporal regions (Narain, et al., 2003; S.K. Scott, Blank, Rosen, & Wise, 2000), leading some authors to argue for a left-lateralized anterior temporal network for speech recognition (Rauschecker & Scott, 2009; S.K. Scott, et al., 2000). This claim is tempered by several facts, however. First, the studies that reported the unilateral anterior

activation foci were severely underpowered. More recent studies with larger sample sizes have reported more posterior temporal lobe activations (Leff, et al., 2008; Spitsyna, Warren, Scott, Turkheimer, & Wise, 2006) and bilateral activation (Okada, et al., 2010). Second, the involvement of the anterior temporal lobe has been reported on the basis of studies that contrast listening to sentences with listening to unintelligible speech. Given such a course-grained comparison, it is impossible to tell what level of linguistic analysis is driving the activation. This is particularly problematic because the anterior temporal lobe has been specifically implicated in higher-level processes such as syntax- and sentence-level semantic integration (Friederici, Meyer, & von Cramon, 2000; Humphries, Binder, Medler, & Liebenthal, 2006; Humphries, Love, Swinney, & Hickok, 2005; Humphries, Willard, Buchsbaum, & Hickok, 2001; Vandenberghe, Nobre, & Price, 2002). Given the ambiguity in functional imaging work, lesion data is particularly critical in adjudicating between the theoretical possibilities and as noted above, lesion data favor a more bilateral model of speech recognition with posterior regions being the most critical for speech recognition (Bates, et al., 2003).

2.1.2. Computational asymmetries—The hypothesis that phoneme-level processes in speech recognition are bilaterally organized does not imply that the two hemispheres are computationally identical. In fact, there is strong evidence for hemispheric differences in the processing of acoustic/speech information (Abrams, Nicol, Zecker, & Kraus, 2008; Boemio, Fromm, Braun, & Poeppel, 2005; Giraud, et al., 2007; Hickok & Poeppel, 2007; Zatorre, Belin, & Penhune, 2002). The basis of these differences is less clear. One view is that the difference turns on selectivity for temporal (left hemisphere) versus spectral (right hemisphere) resolution (Zatorre, et al., 2002). Another proposal is that the two hemispheres differ in terms of their sampling rate, with the left hemisphere operating at a faster rate (25–50 Hz) and the right hemisphere at a slower rate (4–8 Hz) (Poeppel, 2003). These two proposals are not incompatible as there is a relation between sampling rate and spectral vs. temporal resolution (Zatorre, et al., 2002). Further research is needed to sort out these details. For present purposes, the central point is that this asymmetry of function indicates that spoken word recognition involves parallel pathways -- at least one in each hemisphere -- in the mapping from sound to meaning (Hickok & Poeppel, 2007). Although this conclusion differs from standard models of speech recognition (Luce & Pisoni, 1998; Marslen-Wilson, 1987; McClelland & Elman, 1986), it is consistent with the fact that speech contains redundant cues to phonemic information and with behavioral evidence suggesting that the speech system can take advantage of these different cues (Remez, Rubin, Pisoni, & Carrell, 1981; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995).

2.1.3. Phonological processing and the STS—Beyond the earliest stages of speech recognition, there is accumulating evidence that portions of the STS are important for representing and/or processing phonological information (Binder, et al., 2000; Hickok & Poeppel, 2004, 2007; Indefrey & Levelt, 2004; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Price, et al., 1996). The STS is activated by language tasks that require access to phonological information - including both the perception and production of speech (Indefrey & Levelt, 2004), and during active maintenance of phonemic information (B. Buchsbaum, Hickok, & Humphries, 2001; Hickok, Buchsbaum, Humphries, & Muftuler, 2003). Portions of the STS seem to be relatively selective for acoustic signals that contain phonemic information when compared to complex non-speech signals (Hickok & Poeppel, 2007; Liebenthal, et al., 2005; Narain, et al., 2003; Okada, et al., 2010). STS activation can be modulated by the manipulation of psycholinguistic variables that tap phonological networks (Okada & Hickok, 2006), such as phonological neighborhood density (the number of words that sound similar to a target word) and this region shows neural adaptation effects to phonological level information (Vaden, Muftuler, & Hickok, 2009).

2.1.4. Lexical-semantic access—During auditory comprehension, the goal of speech processing is to use phonological information to access conceptual-semantic representations that are critical to comprehension. The dual-stream model holds that while conceptual-semantic representations are widely distributed throughout cortex, a more focal system serves as a computational interface that maps between phonological level representations and distributed conceptual representations (Hickok & Poeppel, 2000, 2004, 2007). This interface is not the site for storage of conceptual information. Instead, it is hypothesized to store information regarding the *relation* (or correspondences) between phonological information on the one hand and conceptual information on the other. Most authors agree that the temporal lobe(s) play a critical role in this process, but again there is disagreement regarding the role of anterior versus posterior regions. The evidence for both of these viewpoints is presented below.

Damage to posterior temporal lobe regions, particularly along the middle temporal gyrus, has long been associated with auditory comprehension deficits (Bates, et al., 2003; H. Damasio, 1991; N.F. Dronkers, Redfern, & Knight, 2000), an effect recently confirmed in a large-scale study involving 101 patients (Bates, et al., 2003). We can infer that these deficits are primarily post-phonemic in nature as phonemic deficits following unilateral lesions to the areas are mild (Hickok & Poeppel, 2004). Data from direct cortical stimulation studies corroborate the involvement of the middle temporal gyrus in auditory comprehension but also indicate the involvement of a much broader network involving most of the superior temporal lobe (including anterior portions) and the inferior frontal lobe(Miglioretti & Boatman, 2003). Functional imaging studies have also implicated posterior middle temporal regions in lexical-semantic processing (Binder, et al., 1997; Rissman, Eliassen, & Blumstein, 2003; Rodd, Davis, & Johnsrude, 2005). These findings do not preclude the involvement of more anterior regions in lexical-semantic access, but they do make a strong case for significant involvement of posterior regions.

Anterior temporal lobe (ATL) regions have been implicated both in lexical-semantic and sentence-level processing (syntactic and/or semantic integration processes). Patients with semantic dementia, who have been used to argue for a lexical-semantic function (S.K. Scott, et al., 2000; Spitsyna, et al., 2006), have atrophy involving the ATL bilaterally, along with deficits on lexical tasks such as naming, semantic association, and single-word comprehension (Gorno-Tempini, et al., 2004). However, these deficits are not specific to the mapping between phonological and conceptual representations and indeed appear to involve more general semantic integration (Patterson, Nestor, & Rogers, 2007). Further, given that atrophy in semantic dementia involves a number of regions in addition to the lateral ATL, including bilateral inferior and medial temporal lobe, bilateral caudate nucleus, and right posterior thalamus, among others (Gorno-Tempini, et al., 2004), linking the deficits specifically to the ATL is difficult.

Higher-level syntactic and compositional semantic processing might involve the ATL. Functional imaging studies have found portions of the ATL to be more active while subjects listen to or read sentences rather than unstructured lists of words or sounds (Friederici, et al., 2000; Humphries, et al., 2005; Humphries, et al., 2001; Mazoyer, et al., 1993; Vandenberghe, et al., 2002). This structured-versus-unstructured effect is independent of the semantic content of the stimuli, although semantic manipulations can modulate the ATL response somewhat (Vandenberghe, et al., 2002). Damage to the ATL has also been linked to deficits in comprehending complex syntactic structures (N. F. Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004). However, data from semantic dementia is contradictory, as these patients are reported to have good sentence-level comprehension (Gorno-Tempini, et al., 2004).

In summary, there is strong evidence that lexical-semantic access from auditory input involves the posterior lateral temporal lobe. In terms of syntactic and compositional semantic operations, neuroimaging evidence is converging on the ATL as an important component of the computational network (Humphries, et al., 2006; Humphries, et al., 2005; Vandenberghe, et al., 2002), however, the neuropsychological evidence remains equivocal.

2.2. Dorsal stream: Mapping from sound to action

The earliest proposals regarding the dorsal auditory stream argued that this system was involved in spatial hearing - a "where" function (Rauschecker, 1998) - similar to the dorsal "where" stream proposal in the cortical visual system (Ungerleider & Mishkin, 1982). More recently, there has been a convergence on the idea that the dorsal stream supports auditory-motor integration (Hickok & Poeppel, 2000, 2004, 2007; Rauschecker & Scott, 2009; S. K. Scott & Wise, 2004; Wise, et al., 2001). Specifically, the idea is that the auditory dorsal stream supports an interface between auditory and motor representations of speech, a proposal similar to the claim that the dorsal visual stream has a sensory-motor integration function (Andersen, 1997; Milner & Goodale, 1995).

2.2.1. The need for auditory-motor integration—The idea of auditory-motor interaction in speech is not new. Wernicke's classic model of the neural circuitry of language incorporated a link between sensory and motor representations of speech and argued explicitly that sensory systems participated in speech production (Wernicke, 1874/1969). More recently, research on motor control has revealed why this sensory-motor link is so critical. Motor acts aim to hit sensory targets. In the visual-manual domain, we identify the location and shape of a cup visually (the sensory target) and generate a motor command that allows us to move our limb toward that location and shape the hand to match the shape of the object. In the speech domain, the targets are not external objects but internal representations of the sound pattern (phonological form) of a word. We know that the targets are auditory in nature because manipulating one's auditory feedback in speech production results in compensatory changes in motor speech acts (Houde & Jordan, 1998; Larson, Burnett, Bauer, Kiran, & Hain, 2001; Purcell & Munhall, 2006). For example, if a subject is asked to produce one vowel and the feedback that he or she hears is manipulated so that it sounds like another, the subject will change the vocal tract configuration so that the feedback sounds like the original vowel. In other words, talkers will readily modify their motor articulations to hit an auditory target indicating that the goal of speech production is not a particular motor configuration but rather a speech *sound* (Guenther, Hampson, & Johnson, 1998). The role of auditory input is nowhere more apparent than in development where the child must use acoustic information in his/her linguistic environment to shape vocal tract movements that must reproduced those sounds.

During the last decade, a great deal of progress has been made in mapping the neural organization of sensorimotor integration for speech. This work has identified a network of regions that include auditory areas in the superior temporal sulcus, motor areas in the left inferior frontal gyrus (parts of Broca's area), a more dorsal left premotor site, and an area in the left planum temporale region, referred to as area Spt (Fig. 1) (B. Buchsbaum, et al., 2001; B. R. Buchsbaum, et al., 2005; Hickok, et al., 2003; Hickok, Houde, et al., 2011; Hickok, Okada, & Serences, 2009; Wise, et al., 2001). One current hypothesis is that the STS regions code sensory-based representations of speech, the motor regions code motor-based representations of speech, and area Spt serves as a sensory-motor integration system, computing a transform between sensory and motor speech representations (Hickok, 2012; Hickok, et al., 2003; Hickok, Houde, et al., 2011; Hickok, et al., 2009).

Lesion evidence is consistent with the functional imaging data implicating Spt as part of a sensorimotor integration circuit. Damage to auditory-related regions in the left hemisphere often results in speech production deficits (A. R. Damasio, 1992; H. Damasio, 1991), demonstrating that sensory systems participate in motor speech. More specifically, damage to the left temporal-parietal junction is associated with conduction aphasia, a syndrome that is characterized by good comprehension but frequent phonemic errors in speech production (Baldo, Klostermann, & Dronkers, 2008; H. Damasio & Damasio, 1980; H. Goodglass, 1992). Conduction aphasia has classically been considered to be a disconnection syndrome involving damage to the arcuate fasciculus. However, there is now good evidence that this syndrome results from cortical dysfunction (Anderson, et al., 1999; Hickok, et al., 2000). The production deficit is load-sensitive: Errors are more likely on longer, lower-frequency words, and on verbatim repetition of strings of speech with little semantic constraint (H. Goodglass, 1992; H Goodglass, 1993). In the context of the above discussion, the effects of such lesions can be understood as an interruption of the system that serves at the interface between auditory target and the motor speech actions that can achieve them (Hickok & Poeppel, 2000, 2004, 2007). The lesion distribution of conduction aphasia has been shown to overlap the location of auditory-motor integration area Spt (B. R. Buchsbaum, et al., 2011), consistent with idea that conduction aphasia results from damage to this interface system.

Recent theoretical work has clarified the computational details underlying auditory-motor integration in the dorsal stream. Drawing on recent advances in understanding motor control generally, speech researchers have emphasized the role of internal forward models in speech motor control (Aliu, Houde, & Nagarajan, 2009; Golfinopoulos, Tourville, & Guenther, 2010; Hickok, Houde, et al., 2011). The basic idea is that the nervous system makes forward predictions about the future state of the motor articulators and the sensory consequences of the predicted actions to control action. The predictions are assumed to be generated by an internal model that receives efference copies of motor commands and integrates them with information about the current state of the system and past experience (learning) of the relation between particular motor commands and their sensory consequences. This internal model affords a mechanism for detecting and correcting motor errors, i.e., motor actions that fail to hit their sensory targets.

Several models have been proposed with similar basic assumptions, but slightly different architectures (Aliu, et al., 2009; Golfinopoulos, et al., 2010; Guenther, et al., 1998; Hickok, Houde, et al., 2011). One such model is shown in Figure 2 (Hickok, Houde, et al., 2011). Input to the system comes from a lexical-conceptual network as assumed by psycholinguistic models of speech production (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Levelt, Roelofs, & Meyer, 1999). In between the input/output system is a phonological system that is split into two components, corresponding to sensory input and motor output subsystems, and is mediated by a sensorimotor translation system that corresponds to area Spt (B. Buchsbaum, et al., 2001; Hickok, et al., 2003; Hickok, et al., 2009). Parallel inputs to sensory and motor systems are needed to explain neuropsychological observations (Jacquemot, Dupoux, & Bachoud-Levi, 2007), such as conduction aphasia, as we will see immediately below. Inputs to the auditory-phonological network define the auditory targets of speech acts. As a motor speech unit (ensemble) begins to be activated, its predicted auditory consequences can be checked against the auditory target. If they match, then that unit will continue to be activated, resulting in an articulation that will hit the target. If there is a mismatch, then a correction signal can be generated to activate the correct motor unit.

This model provides a natural explanation of the symptom pattern in conduction aphasia. A lesion to Spt would disrupt the ability to generate forward predictions in the auditory cortex

and thereby the ability to perform internal feedback monitoring, making errors more frequent than in an unimpaired system (Fig. 2B). However, this would not disrupt the activation of auditory targets via the lexical semantic system, thus leaving the patient capable of detecting errors in their own speech, a characteristic of conduction aphasia. Once an error is detected, however, the correction signal will not be accurately translated to the internal model of the vocal tract due to disruption of Spt. The ability to detect but not accurately correct speech errors should result in repeated unsuccessful self-correction attempts, again a characteristic of conduction aphasia.

3. Forward prediction in speech perception

An examination of current discussions of speech perception in the literature gives the impression that forward prediction in speech perception is virtually axiomatic. It is demonstrably true that knowing what to listen for enhances our ability to perceive speech. What is less clear is the source of these predictions.

The idea that forward prediction plays a critical role in motor control, including speech, is well-established. Recently, several groups have suggested that forward prediction from the motor system may facilitate speech *perception* (Friston, Daunizeau, Kilner, & Kiebel, 2010; Hickok, Houde, et al., 2011; Rauschecker & Scott, 2009; Sams, Mottonen, & Sihvonen, 2005; Skipper, Nusbaum, & Small, 2005; van Wassenhove, Grant, & Poeppel, 2005; Wilson & Iacoboni, 2006). The logic behind this idea is the following: If the motor system can generate predictions for the sensory consequences of one's own speech actions, then maybe this system can be used to make predictions about the sensory consequences of others' speech and thereby facilitate perception. Research using transcranial magnetic stimulation (TMS) has been argued to provide evidence for this view by showing that stimulation of motor speech regions can modulate perception of speech under some circumstances (D'Ausilio, et al., 2009; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Mottonen & Watkins, 2009). There are both conceptual and empirical problems with this idea, however. Conceptually, the goal of forward prediction in the context of motor control is to detect *deviation* from prediction, i.e., prediction *error*, which can then be used to correct movement. If the prediction is correct, i.e., the movement has hit its intended sensory target and no correction is needed, the system can ignore the sensory input. This is not the kind of system one would want to engineer for enhancing perception. One would want a system designed to *enhance* the perception of predicted elements, not ignore the occurrence of predicted elements. Empirically, motor prediction in general tends to lead to a decrease in the perceptual response and indeed a decrease in perceptual sensitivity. The classic example is saccadic suppression: Self-generated eye movements do not result in the percept of visual motion even though a visual image is sweeping across the retina. The mechanism underlying this fact appears to be motor-induced neural suppression of motion signals (Bremmer, Kubischik, Hoffmann, & Krekelberg, 2009). A similar motor-induced suppression effect has been reported during self-generated speech (Aliu, et al., 2009; Ventura, Nagarajan, & Houde, 2009).

But even if there are general concerns about the feasibility of a motor-based prediction enhancing speech perception, one could point to the TMS literature as empirical evidence in favor of a facilitatory effect of motor prediction on speech perception. There are problems with this literature, however. One problem is that recent lesion-based work contradicts the TMS-based findings. Damage to the motor speech system does not cause corresponding deficits in speech perception as one would expect if motor prediction were critically important (Hickok, Costanzo, et al., 2011; Rogalsky, et al., 2011). A second problem is that the TMS studies used measures that are susceptible to response bias. It is therefore unclear whether TMS is modulating perception, as is typically assumed, or just modulating response

bias. Raising concern in this respect is a recent study that used a motor fatigue paradigm to modulate the motor speech system and then examine the effects on speech perception. Using signal detection methods, this study reports that the motor manipulation affected responses bias but not perceptual discrimination (d') (Sato, et al., 2011). A third problem is that all of the TMS studies use tasks that require participants to consciously attend to phonemic information. Such tasks may draw on motor resources (e.g., phonological working memory) that are not normally engaged in speech perception during auditory comprehension (Hickok & Poeppel, 2000, 2004, 2007).

While sensory prediction coming out of the motor system (dorsal stream) is getting a lot of attention currently, there is another source of prediction: namely the ventral stream. Ventral stream prediction has a long history in perceptual processing and has been studied under several terminological guises such as priming, context effects, and top-down expectation. Behaviorally, increasing the predictability of a sensory event via context (G. A. Miller, Heise, & Lichten, 1951) or priming (Ellis, 1982; Jackson & Morton, 1984) correspondingly *increases* the detectability of the predicted stimulus, unlike the behavioral effects of motor-based prediction. A straightforward view of these observations is that forward prediction is useful for perceptual recognition only when it is coming out of the ventral stream.

One possible exception to this conclusion comes from research on audiovisual speech integration. Visual speech has a significant effect on speech recognition (Dodd, 1977; Sumby & Pollack, 1954), so much so that mismatched visual speech information can alter auditory speech percepts under some conditions (McGurk & MacDonald, 1976). Because visual speech provides information about the articulatory gestures generating speech sounds, the influence of visual speech has been linked to motor-speech function (Liberman & Mattingly, 1985) and more recently to motor-generated forward prediction (Sams, et al., 2005; Skipper, et al., 2005). Is this evidence that motor prediction can actually facilitate auditory speech recognition? Probably not. Recent work has suggested that auditory-related cortical areas in the superior temporal lobe, but not motor areas, exhibit the neurophysiological signatures of multisensory integration, namely an additive or supra-additive response to audiovisual speech (AV > A or V alone) and/or greater activity when audiovisual information is synchronized compared to when it is not (Calvert, Campbell, & Brammer, 2000; L. M. Miller & D'Esposito, 2005). Further behavioral responses to audiovisual speech stimuli have been linked to modulations of activity in these superior temporal lobe regions (Beauchamp, Nath, & Pasalar, 2010; Nath & Beauchamp, 2011, 2012; Nath, Fava, & Beauchamp, 2011). And finally, there is evidence that motor-speech experience is not necessary for audiovisual integration (Burnham & Dodd, 2004; Rosenblum, Schmuckler, & Johnson, 1997; Siva, Stevens, Kuhl, & Meltzoff, 1995). The weight of the evidence suggests that the influence of visual speech information is not motor speech-mediated but rather is a function of cross-sensory integration in the ventral stream.

4. Conclusions

The dual stream model has proven to be a useful framework for understanding aspects of the neural organization of language (Hickok & Poeppel, 2000, 2004, 2007) and much progress has been made in understanding the neural architecture and computations of both the ventral and dorsal streams. I suggested here that it might also be fruitful to consider the notion of forward prediction in the context of a dual stream model; specifically that the dorsal and ventral streams represent two sources of forward prediction. Based on the evidence reviewed and consistent with their primary computational roles, I conclude that dorsal stream forward prediction primarily serves motor control functions and does not facilitate recognition of others' speech, whereas ventral stream forward prediction functions to enhance speech recognition.

Appendix A. Continuing education

1. The dorsal speech stream primarily supports

  A. Sensory-motor integration

  B. Speech perception

  C. Auditory comprehension

  D. Conceptual-semantic analysis

  E. Lexical access

2. The ventral speech stream primarily supports

  A. Sensory-motor integration

  B. Auditory comprehension

  C. Motor control

  D. State feedback control

  E. Speech production

3. Conduction aphasia results from

  A. Damage to the arcuate fasciculus

  B. Damage to Wernicke's area

  C. Bilateral damage to receptive speech centers

  D. Dysfunction involving the dorsal stream

  E. Dysfunction of the ventral stream

4. The ability to recognize speech is supported by

  A. The left hemisphere alone

  B. The right hemisphere alone

  C. Both hemispheres symmetrically

  D. Both hemispheres but with some asymmetry in the way the signal is analyzed

  E. Broca's area

5. Predictive coding or forward prediction

  A. refers to the brains ability to generate predictions about motor states and sensory events.

  B. is an important part of motor control circuits

  C. is possible within both the dorsal and ventral streams

  D. is a well accepted notion in neural computation

  E. All of the above

Answer key: 1:A, 2: B, 3: D; 4: D, 5: E

References

Abrams DA, Nicol T, Zecker S, Kraus N. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. Journal of Neuroscience. 2008; 28:3958–3965. [PubMed: 18400895]

Aliu SO, Houde JF, Nagarajan SS. Motor-induced suppression of the auditory cortex. Journal of Cognitive Neuroscience. 2009; 21:791–802. [PubMed: 18593265]

Andersen R. Multimodal integration for the representation of space in the posterior parietal cortex. Philosophical Transactions of the Royal Society of London B Biological Sciences. 1997; 352:1421–1428.

Anderson JM, Gilmore R, Roper S, Crosson B, Bauer RM, Nadeau S, Beversdorf DQ, Cibula J, Rogish M III, Kortencamp S, Hughes JD, Gonzalez Rothi LJ, Heilman KM. Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model. Brain and Language. 1999; 70:1–12. [PubMed: 10534369]

Baker E, Blumsteim SE, Goodglass H. Interaction between phonological and semantic factors in auditory comprehension. Neuropsychologia. 1981; 19:1–15. [PubMed: 7231654]

Baldo JV, Klostermann EC, Dronkers NF. It's either a cook or a baker: patients with conduction aphasia get the gist but lose the trace. Brain and Language. 2008; 105:134–140. [PubMed: 18243294]

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF. Voxel-based lesion-symptom mapping. Nature Neuroscience. 2003; 6:448–450.

Beauchamp MS, Nath AR, Pasalar S. fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. Journal of Neuroscience. 2010; 30:2414–2417. [PubMed: 20164324]

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. Human temporal lobe activation by speech and nonspeech sounds. Cerebral Cortex. 2000; 10:512–528. [PubMed: 10847601]

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T. Human brain language areas identified by functional magnetic resonance imaging. Journal of Neuroscience. 1997; 17:353–362. [PubMed: 8987760]

Bishop DV, Brown BB, Robson J. The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals. Journal of Speech and Hearing Research. 1990; 33:210–219. [PubMed: 2359262]

Boemio A, Fromm S, Braun A, Poeppel D. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nature Neuroscience. 2005; 8:389–395.

Bremmer F, Kubischik M, Hoffmann KP, Krekelberg B. Neural dynamics of saccadic suppression. Journal of Neuroscience. 2009; 29:12374–12383. [PubMed: 19812313]

Buchman AS, Garron DC, Trost-Cardamone JE, Wichter MD, Schwartz M. Word deafness: One hundred years later. Journal of Neurology, Neurosurgury, and Psychiatry. 1986; 49:489–499.

Buchsbaum B, Hickok G, Humphries C. Role of Left Posterior Superior Temporal Gyrus in Phonological Processing for Speech Perception and Production. Cognitive Science. 2001; 25:663–678.

Buchsbaum BR, Baldo J, Okada K, Berman KF, Dronkers N, D'Esposito M, Hickok G. Conduction aphasia, sensory-motor integration, and phonological short-term memory - an aggregate analysis of lesion and fMRI data. Brain and Language. 2011; 119:119–128. [PubMed: 21256582]

Buchsbaum BR, Olsen RK, Koch PF, Kohn P, Kippenhan JS, Berman KF. Reading, hearing, and the planum temporale. Neuroimage. 2005; 24:444–454. [PubMed: 15627586]

Burnham D, Dodd B. Auditory-visual speech integration by prelinguistic infants: perception of an emergent consonant in the McGurk effect. Developmental Psychobiology. 2004; 45:204–220. [PubMed: 15549685]

Calvert GA, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Current Biology. 2000; 10:649–657. [PubMed: 10837246]

D'Ausilio A, Pulvermuller F, Salmas P, Bufalari I, Begliomini C, Fadiga L. The motor somatotopy of speech perception. Current Biology. 2009; 19:381–385. [PubMed: 19217297]

Damasio AR. Aphasia. New England Journal of Medicine. 1992; 326:531–539. [PubMed: 1732792]

Damasio, H. Neuroanatomical correlates of the aphasias. In: Sarno, M., editor. Acquired aphasia. 2nd ed. San Diego: Academic Press; 1991. p. 45-71.

Damasio H, Damasio AR. The anatomical basis of conduction aphasia. Brain. 1980; 103:337–350. [PubMed: 7397481]

Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. Lexical access in aphasic and nonaphasic speakers. Psychological Review. 1997; 104:801–838. [PubMed: 9337631]

Dodd B. The role of vision in the perception of speech. Perception. 1977; 6:31–40. [PubMed: 840618]

Dronkers, NF.; Redfern, BB.; Knight, RT. The neural architecture of language disorders. In: Gazzaniga, MS., editor. The new cognitive neurosciences. Cambridge, MA: MIT Press; 2000. p. 949-958.

Dronkers NF, Wilkins DP, Van Valin RD Jr, Redfern BB, Jaeger JJ. Lesion analysis of the brain areas involved in language comprehension. Cognition. 2004; 92:145–177. [PubMed: 15037129]

Ellis AW. Modality-specific repetition priming of auditory word recognition. Current Psychological Research. 1982; 2:123–128.

Friederici AD, Meyer M, von Cramon DY. Auditory langue comprehension: An event-related fMRI study on the processing of syntactic and lexical information. Brain and Language. 2000; 74:289–300. [PubMed: 10950920]

Friston KJ, Daunizeau J, Kilner J, Kiebel SJ. Action and behavior: a free-energy formulation. Biological Cybernetics. 2010; 102:227–260. [PubMed: 20148260]

Giraud AL, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RS, Laufs H. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron. 2007; 56:1127–1134. [PubMed: 18093532]

Golfinopoulos E, Tourville JA, Guenther FH. The integration of large-scale neural network modeling and functional brain imaging in speech motor control. Neuroimage. 2010; 52:862–874. [PubMed: 19837177]

Goodglass, H. Diagnosis of conduction aphasia. In: Kohn, SE., editor. Conduction aphasia. Hillsdale, NJ: Lawrence Erlbaum Associates; 1992. p. 39-49.

Goodglass, H. Understanding aphasia. San Diego: Academic Press; 1993.

Gorno-Tempini ML, Dronkers NF, Rankin KP, Ogar JM, Phengrasamy L, Rosen HJ, Johnson JK, Weiner MW, Miller BL. Cognition and anatomy in three variants of primary progressive aphasia. Annals of Neurology. 2004; 55:335–346. [PubMed: 14991811]

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]

Haruno M, Wolpert DM, Kawato M. Mosaic model for sensorimotor learning and control. Neural Computation. 2001; 13:2201–2220. [PubMed: 11570996]

Hickok G. Computational neuroanatomy of speech production. Nature Reviews Neuroscience. 2012; 13:135–145.

Hickok G. The role of mirror neurons in speech perception and action word semantics. Language and Cognitive Processes. (in press).

Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. Journal of Cognitive Neuroscience. 2003; 15:673–682. [PubMed: 12965041]

Hickok G, Costanzo M, Capasso R, Miceli G. The role of Broca's area in speech perception: Evidence from aphasia revisited. Brain and Language. 2011; 119:214–220. [PubMed: 21920592]

Hickok G, Erhard P, Kassubek J, Helms-Tillery AK, Naeve-Velguth S, Strupp JP, Strick PL, Ugurbil K. A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. Neuroscience Letters. 2000; 287:156–160. [PubMed: 10854735]

Hickok G, Houde J, Rong F. Sensorimotor integration in speech processing: computational basis and neural organization. Neuron. 2011; 69:407–422. [PubMed: 21315253]

Hickok G, Okada K, Barr W, Pa J, Rogalsky C, Donnelly K, Barde L, Grant A. Bilateral capacity for speech sound processing in auditory comprehension: evidence from Wada procedures. Brain and Language. 2008; 107:179–184. [PubMed: 18976806]

Hickok G, Okada K, Serences JT. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. Journal of Neurophysiology. 2009; 101:2725–2732. [PubMed: 19225172]

Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. Trends in Cognitive Sciences. 2000; 4:131–138. [PubMed: 10740277]

Hickok G, Poeppel D. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. Cognition. 2004; 92:67–99. [PubMed: 15037127]

Hickok G, Poeppel D. The cortical organization of speech processing. Nature Reviews Neuroscience. 2007; 8:393–402.

Houde JF, Jordan MI. Sensorimotor adaptation in speech production. Science. 1998; 279:1213–1216. [PubMed: 9469813]

Humphries C, Binder JR, Medler DA, Liebenthal E. Syntactic and semantic modulation of neural activity during auditory sentence comprehension. Journal of Cognitive Neuroscience. 2006; 18:665–679. [PubMed: 16768368]

Humphries C, Love T, Swinney D, Hickok G. Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. Human Brain Mapping. 2005; 26:128–138. [PubMed: 15895428]

Humphries C, Willard K, Buchsbaum B, Hickok G. Role of anterior temporal cortex in auditory sentence comprehension: An fMRI study. Neuroreport. 2001; 12:1749–1752. [PubMed: 11409752]

Indefrey P, Levelt WJ. The spatial and temporal signatures of word production components. Cognition. 2004; 92:101–144. [PubMed: 15037128]

Jackson A, Morton J. Facilitation of auditory word recognition. Memory and Cognition. 1984; 12:568–574.

Jacquemot C, Dupoux E, Bachoud-Levi AC. Breaking the mirror: Asymmetrical disconnection between the phonological input and output codes. Cognitive Neuropsychology. 2007; 24:3–22. [PubMed: 18416481]

Larson CR, Burnett TA, Bauer JJ, Kiran S, Hain TC. Comparison of voice F0 responses to pitch-shift onset and offset conditions. Journal of the Acoustical Society of America. 2001; 110:2845–2848. [PubMed: 11785786]

Leff AP, Schofield TM, Stephan KE, Crinion JT, Friston KJ, Price CJ. The cortical dynamics of intelligible speech. Journal of Neuroscience. 2008; 28:13209–13215. [PubMed: 19052212]

Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. Behavioral & Brain Sciences. 1999; 22:1–75. [PubMed: 11301520]

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

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. Cerebral Cortex. 2005; 15:1621–1631. [PubMed: 15703256]

Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. Ear and Hearing. 1998; 19:1–36. [PubMed: 9504270]

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

Mazoyer BM, Tzourio N, Frak V, Syrota A, Murayama N, Levrier O, Salamon G, Dehaene S, Cohen L, Mehler J. The cortical representation of speech. Journal of Cognitive Neuroscience. 1993; 5:467–479.

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

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264:746–748. [PubMed: 1012311]

Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M. The essential role of premotor cortex in speech perception. Current Biology. 2007; 17:1692–1696. [PubMed: 17900904]

Miglioretti DL, Boatman D. Modeling variability in cortical representations of human complex sound perception. Experimental Brain Research. 2003; 153:382–387.

Miller GA, Heise GA, Lichten W. The intelligibility of speech as a function of the context of the test materials. Journal of Experimental Psychology. 1951; 41:329–335. [PubMed: 14861384]

Miller LM, D'Esposito M. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. Journal of Neuroscience. 2005; 25:5884–5893. [PubMed: 15976077]

Milner, AD.; Goodale, MA. The visual brain in action. Oxford: Oxford University Press; 1995.

Mottonen R, Watkins KE. Motor representations of articulators contribute to categorical perception of speech sounds. Journal of Neuroscience. 2009; 29:9819–9825. [PubMed: 19657034]

Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD, Matthews PM. Defining a left-lateralized response specific to intelligible speech using fMRI. Cerebral Cortex. 2003; 13:1362–1368. [PubMed: 14615301]

Nath AR, Beauchamp MS. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. Journal of Neuroscience. 2011; 31:1704–1714. [PubMed: 21289179]

Nath AR, Beauchamp MS. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. Neuroimage. 2012; 59:781–787. [PubMed: 21787869]

Nath AR, Fava EE, Beauchamp MS. Neural correlates of interindividual differences in children's audiovisual speech perception. Journal of Neuroscience. 2011; 31:13963–13971. [PubMed: 21957257]

Okada K, Hickok G. Identification of lexical-phonological networks in the superior temporal sulcus using fMRI. Neuroreport. 2006; 17:1293–1296. [PubMed: 16951572]

Okada K, Rong F, Venezia J, Matchin W, Hsieh IH, Saberi K, Serences JT, Hickok G. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. Cerebral Cortex. 2010; 20:2486–2495. [PubMed: 20100898]

Patterson K, Nestor PJ, Rogers TT. Where do you know what you know? The representation of semantic knowledge in the human brain. Nature Reviews Neuroscience. 2007; 8:976–987.

Poeppel D. Pure word deafness and the bilateral processing of the speech code. Cognitive Science. 2001; 25:679–693.

Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as "asymmetric sampling in time". Speech Communication. 2003; 41:245–255.

Price CJ, Wise RJS, Warburton EA, Moore CJ, Howard D, Patterson K, Frackowiak RSJ, Friston KJ. Hearing and saying: The functional neuro-anatomy of auditory word processing. Brain. 1996; 119:919–931. [PubMed: 8673502]

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]

Rauschecker JP. Cortical processing of complex sounds. Current Opinion in Neurobiology. 1998; 8:516–521. [PubMed: 9751652]

Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nature Neuroscience. 2009; 12:718–724.

Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. Science. 1981; 212:947–950. [PubMed: 7233191]

Rissman J, Eliassen JC, Blumstein SE. An event-related FMRI investigation of implicit semantic priming. Journal of Cognitive Neuroscience. 2003; 15:1160–1175. [PubMed: 14709234]

Rodd JM, Davis MH, Johnsrude IS. The neural mechanisms of speech comprehension: fMRI studeis of semantic ambiguity. Cerebral Cortex. 2005; 15:1261–1269. [PubMed: 15635062]

Rogalsky C, Love T, Driscoll D, Anderson SW, Hickok G. Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system. Neurocase. 2011; 17:178–187. [PubMed: 21207313]

Rogalsky C, Pitz E, Hillis AE, Hickok G. Auditory word comprehension impairment in acute stroke: relative contribution of phonemic versus semantic factors. Brain and Language. 2008; 107:167–169. [PubMed: 18823655]

Rosenblum LD, Schmuckler MA, Johnson JA. The McGurk effect in infants. Perception and Psychophysics. 1997; 59:347–357. [PubMed: 9136265]

Sams M, Mottonen R, Sihvonen T. Seeing and hearing others and oneself talk. Brain Research: Cognitive Brain Research. 2005; 23:429–435. [PubMed: 15820649]

Sato M, Grabski K, Glenberg AM, Brisebois A, Basirat A, Menard L, Cattaneo L. Articulatory bias in speech categorization: Evidence from use-induced motor plasticity. Cortex. 2011; 47:1001–1003. [PubMed: 21501836]

Scott SK, Blank CC, Rosen S, Wise RJS. Identification of a pathway for intelligible speech in the left temporal lobe. Brain. 2000; 123:2400–2406. [PubMed: 11099443]

Scott SK, Wise RJ. The functional neuroanatomy of prelexical processing in speech perception. Cognition. 2004; 92:13–45. [PubMed: 15037125]

Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 1995; 270:303–304. [PubMed: 7569981]

Siva N, Stevens EB, Kuhl PK, Meltzoff AN. A comparison between cerebral-palsied and normal adults in the perception of auditory-visual illusions. Journal of the Acoustical Society of America. 1995; 98:2983.

Skipper JI, Nusbaum HC, Small SL. Listening to talking faces: motor cortical activation during speech perception. Neuroimage. 2005; 25:76–89. [PubMed: 15734345]

Spitsyna G, Warren JE, Scott SK, Turkheimer FE, Wise RJ. Converging language streams in the human temporal lobe. Journal of Neuroscience. 2006; 26:7328–7336. [PubMed: 16837579]

Sumby WH, Pollack I. Visual contributions to speech intelligibility in noise. Journal of the Acoustical Society of America. 1954; 26

Ungerleider, LG.; Mishkin, M. Two cortical visual systems. In: Ingle, DJ.; Goodale, MA.; Mansfield, RJW., editors. Analysis of visual behavior. Cambridge, MA: MIT Press; 1982. p. 549-586.

Vaden KI Jr, Muftuler LT, Hickok G. Phonological repetition-suppression in bilateral superior temporal sulci. Neuroimage. 2009

van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. Procdings of the National Academy of Sciences. 2005; 102:1181–1186.

Vandenberghe R, Nobre AC, Price CJ. The response of left temporal cortex to sentences. Journal of Cognitive Neuroscience. 2002; 14:550–560. [PubMed: 12126497]

Ventura MI, Nagarajan SS, Houde JF. Speech target modulates speaking induced suppression in auditory cortex. BMC Neuroscience. 2009; 10:58. [PubMed: 19523234]

Wernicke, C. The symptom complex of aphasia: A psychological study on an anatomical basis. In: Cohen, RS.; Wartofsky, MW., editors. Boston studies in the philosophy of science. Dordrecht: D. Reidel Publishing Company; 1874/1969. p. 34-97.

Wernicke, C. Der aphasische symptomencomplex: Eine psychologische studie auf anatomischer basis. In: Eggert, GH., editor. Wernicke's works on aphasia: A sourcebook and review. The Hague: Mouton; 1874/1977. p. 91-145.

Wilson SM, Iacoboni M. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. Neuroimage. 2006; 33:316–325. [PubMed: 16919478]

Wise RJS, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA. Separate neural sub-systems within "Wernicke's area". Brain. 2001; 124:83–95. [PubMed: 11133789]

Zatorre RJ, Belin P, Penhune VB. Structure and function of auditory cortex: music and speech. Trends in Cognitive Sciences. 2002; 6:37–46. [PubMed: 11849614]
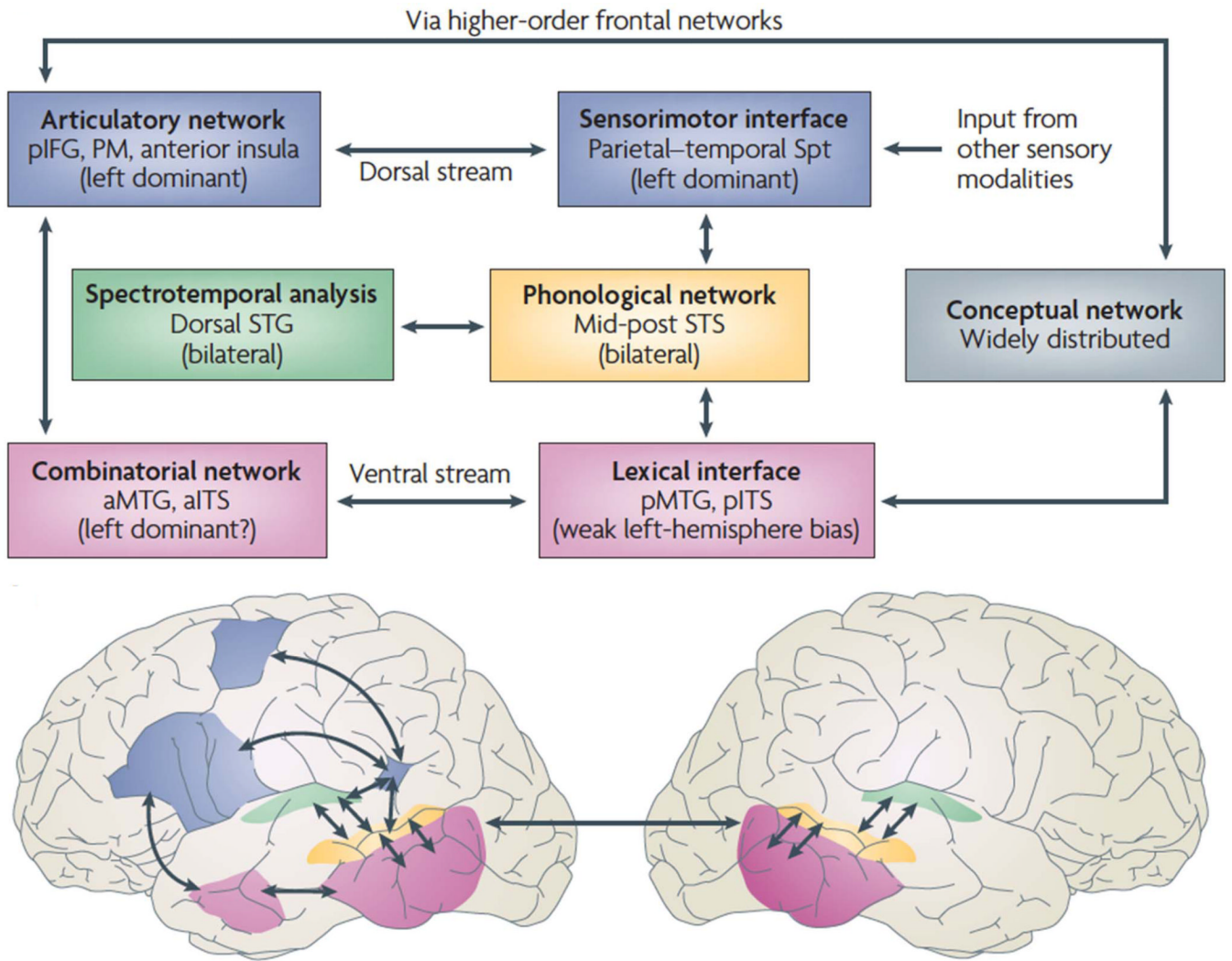
Fig. 1. Dual stream model of speech processing

The dual stream model (Hickok & Poeppel, 2000, 2004, 2007) holds that early stages of speech processing occurs bilaterally in auditory regions on the dorsal STG (spectrotemporal analysis; green) and STS (phonological access/representation; yellow), and then diverges into two broad streams: a temporal lobe ventral stream supports speech comprehension (lexical access and combinatorial processes; pink) whereas a strongly left dominant dorsal stream supports sensory-motor integration and involves structures at the parietal-temporal junction (Spt) and frontal lobe. The conceptual network (gray box) is assumed to be widely distributed throughout cortex. IFG, inferior frontal gyrus; ITS, inferior temporal sulcus; MTG, middle temporal gyrus; PM, premotor; Spt, Sylvian parietal-temporal; STG, superior temporal gyrus; STS, superior temporal sulcus. Figure reprinted with permission from (Hickok & Poeppel, 2007).

A.
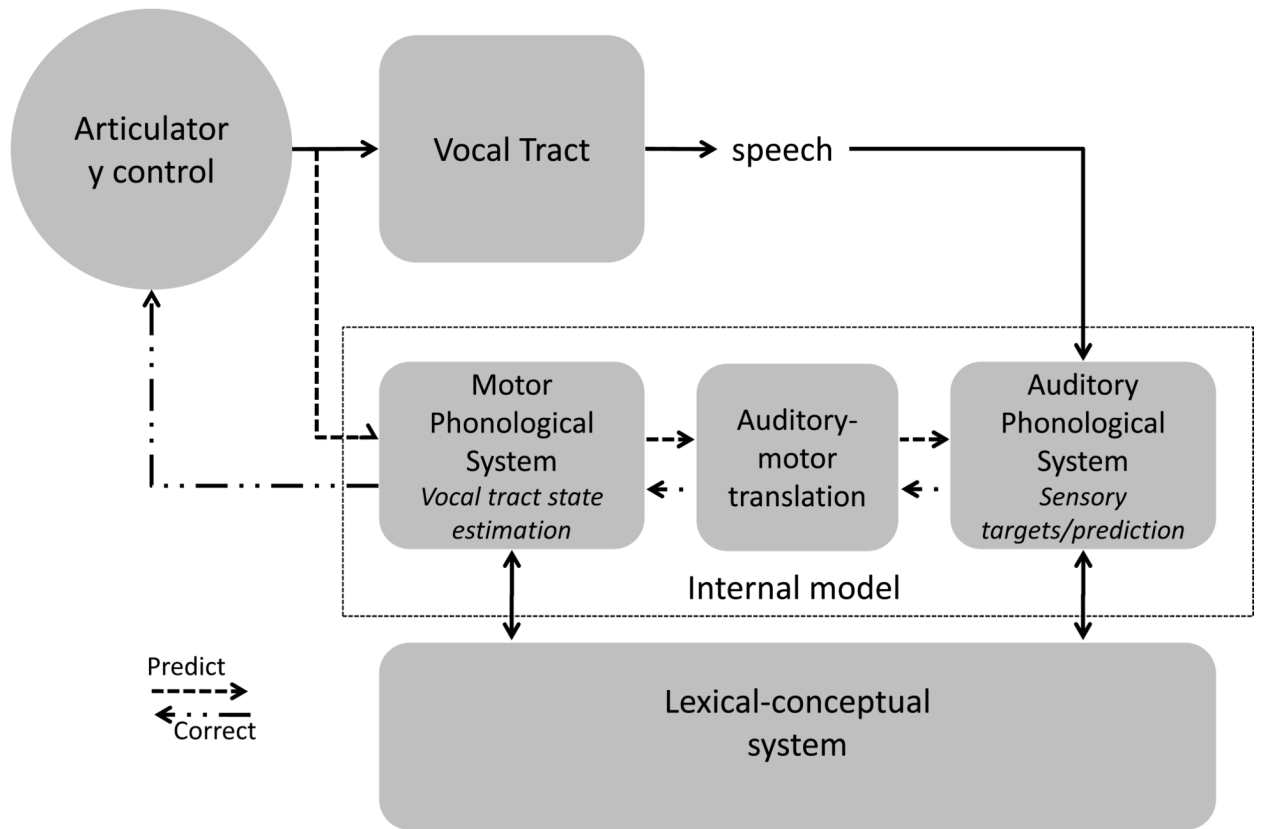
B.
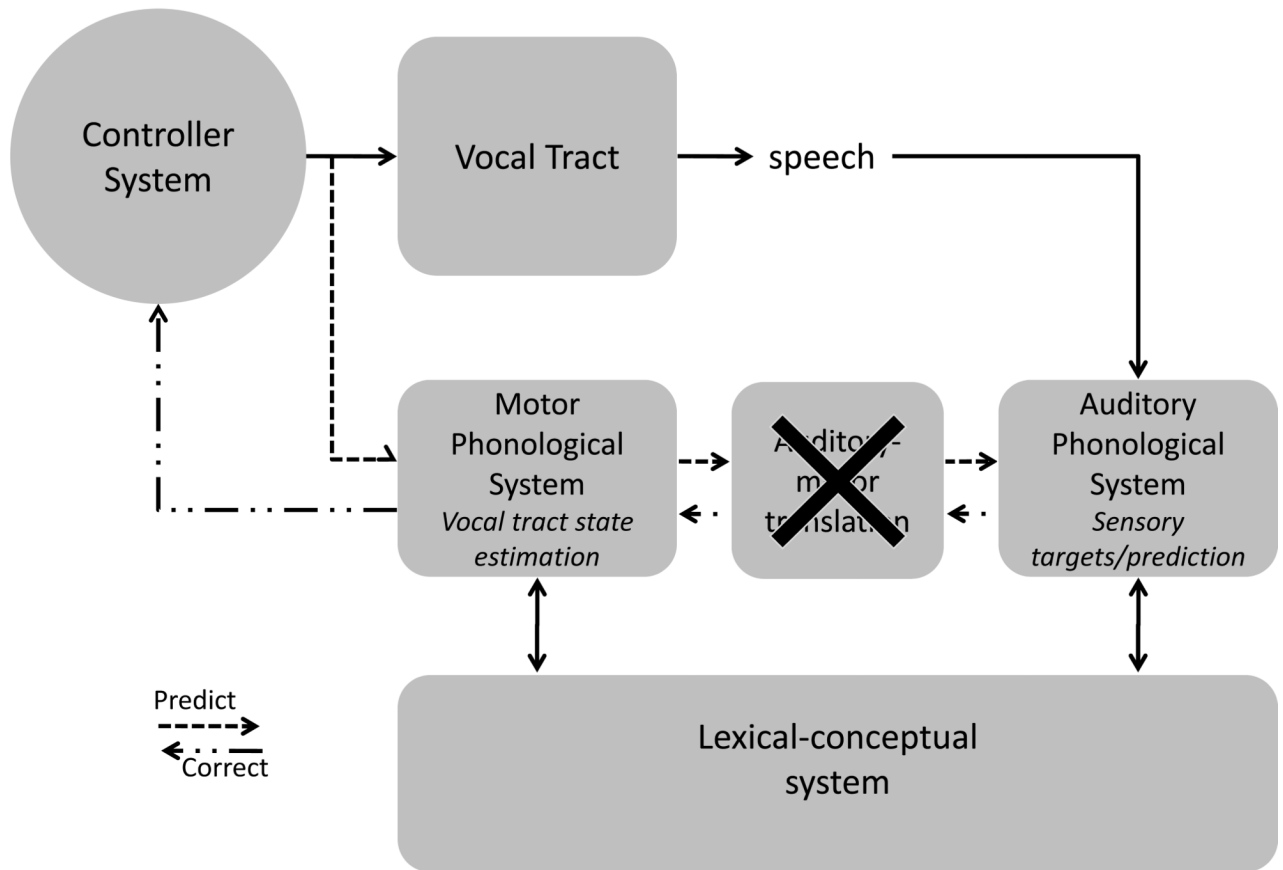
Fig. 2. An integrated state feedback control (SFC) model of speech production
A. Speech models derived from the feedback control, psycholinguistic, and neurolinguistic literatures are integrated into one framework, presented here. The architecture is fundamentally that of a SFC system with a controller, or set of controllers (Haruno, Wolpert, & Kawato, 2001), localized to primary motor cortex, which generates motor commands to the vocal tract and sends a corollary discharge to an internal model which makes forward predictions about both the dynamic state of the vocal tract and about the sensory consequences of those states. Deviations between predicted auditory states and the intended targets or actual sensory feedback generates an error signal that is used to correct and update the internal model of the vocal tract. The internal model of the vocal tract is instantiated as a "motor phonological system", which corresponds to the neurolinguistically elucidated *phonological output lexicon*, and is localized to premotor cortex. Auditory targets and forward predictions of sensory consequences are encoded in the same network, namely the "auditory phonological system", which corresponds to the neurolinguistically elucidated *phonological input lexicon*, and is localized to the STG/STS. Motor and auditory phonological systems are linked via an auditory-motor translation system, localized to area Spt. The system is activated via parallel inputs from the lexical-conceptual system to the motor and auditory phonological systems. **B**. Proposed source of the deficit in conduction aphasia: damage to the auditory-motor translation system. Input from the lexical conceptual system to motor and auditory phonological systems are unaffected allowing for fluent output and accurate activation of sensory targets. However, internal forward sensory predictions are not possible leading to an increase in error rate. Further, errors detected as a consequence of mismatches between sensory targets and actual sensory feedback cannot be used to correct

motor commands. Reprinted with permission from (Hickok, Houde, et al., 2011; Hickok & Poeppel, 2004)