Handbook of Clinical Neurology, Vol. 129 (3rd series)
The Human Auditory System
G.G. Celesia and G. Hickok, Editors

Chapter 8

Neural basis of speech perception

GREGORY HICKOK[1]* AND DAVID POEPPEL[2]
[1]*Department of Cognitive Sciences, Center for Language Science and Center for Cognitive Neuroscience, University of California, Irvine, CA, USA*
[2]*Department of Psychology, New York University, New York, NY, USA*

INTRODUCTION

The mind/brain must figure out at least two things when faced with the task of learning a natural language. One is how to transform the sound patterns of speech into a representation of the meaning of an utterance. The other is how to reproduce those sound patterns with the vocal tract (or, in the case of signed languages, with manual and facial gestures). Put differently, speech information must be processed along two different routes, an auditory-conceptual route and an auditory-motor route. These two processing streams involve partially segregated circuits in the brain and form the basis of the dual-route model of speech processing (Hickok and Poeppel, 2000, 2004, 2007), which traces its routes to the classic model of Wernicke (1874/1977), and parallels analogous proposals in the visual (Milner and Goodale, 1995) and somatosensory (Dijkerman and de Haan, 2007) systems. Thus, the division of labor proposed in dual-route models, wherein one route is sensory-conceptual and the other sensory-motor, appears to be a general organizational property of the cerebral cortex.

This chapter outlines the dual-route model as a foundation for understanding the functional anatomy of speech and language processing.

THE DUAL-ROUTE MODEL OF SPEECH PROCESSING

The dual-route model (Fig. 8.1) 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 (at the parietal-temporal junction) and the posterior frontal lobe, is involved in translating acoustic-based representations of speech signals into articulatory representations essential for speech production. In contrast to the canonical 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 compelling extent to which neuroimaging data implicate both hemispheres has recently been reviewed (Turkeltaub and Coslett, 2010; Price, 2012; Schirmer et al., 2012). The dorsal stream, on the other hand, is traditionally, and in the model outlined here, held to be strongly left-dominant.

Ventral stream: mapping from sound to meaning

Bilateral organization and parallel computation

The ventral stream is bilaterally organized, although not computationally redundant in the two hemispheres. This may not be obvious based on a cursory evaluation of the clinical data. After all, left-hemisphere damage yields language deficits of a variety of sorts, including comprehension impairment, while, in most cases, right-hemisphere damage has little effect on phonologic, lexical, or sentence-level language abilities. A closer look tells a different story. In particular, research in the 1980s showed that auditory comprehension deficits in aphasia (caused by unilateral left-hemisphere lesions) were not caused primarily by impairment in the ability to perceive speech sounds, as Wernicke and later Luria proposed (Wernicke, 1874/1969; Luria, 1970). For example, when

*Correspondence to: Gregory Hickok, Department of Cognitive Sciences, University of California, Irvine, CA 92697-5100, USA. E-mail: greg.hickok@uci.edu

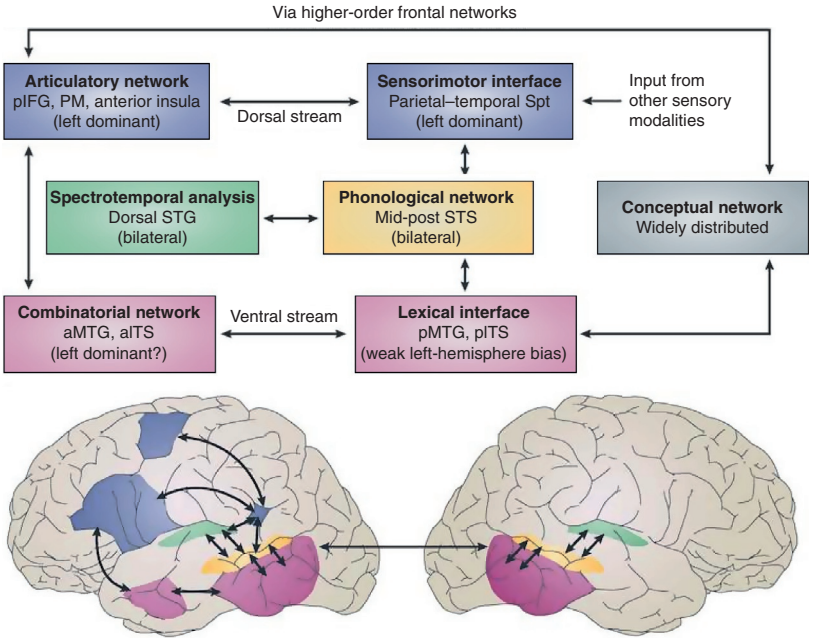

Fig. 8.1. Dual-stream model of speech processing. The dual-stream model (Hickok and Poeppel, 2000, 2004, 2007) holds that early stages of speech processing occur bilaterally in auditory regions on the dorsal superior temporal gyrus (STG) (spectrotemporal analysis: green) and superior temporal sulcus (STS) (phonologic 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 Sylvian parietal–temporal (Spt) junction 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; p, posterior; a, anterior. (Reproduced from Hickok and Poeppel, 2007, with permission.)

Wernicke's aphasics are asked to match pictures to auditorily presented words, their overall performance is well above chance, and when they do make errors they tend to confuse the correct answer with semantically similar alternatives more often than with phonemically similar foils (Miceli et al., 1980; Baker et al., 1981; Rogalsky et al., 2008, 2011). A similar pattern of performance has been observed following acute deactivation of the entire left hemisphere in Wada procedures (Fig. 8.2) (Hickok et al., 2008). "Speech perception" deficits can be identified in left-injured patients, but only on metalinguistic tasks, such as syllable discrimination, that involve some level of conscious attention to phonemic structure and working memory; the involvement of the left hemisphere in these tasks likely follows from the relation between working memory and speech articulation (Hickok and Poeppel, 2000, 2004, 2007). In contrast to the (minimal) effects of unilateral lesions on the processing of phoneme-level information during auditory comprehension, bilateral lesions involving the superior temporal lobe can have a devastating effect, as cases of word deafness attest (see Chapter 32) (Buchman et al., 1986; Poeppel, 2001).

Data from neuroimaging have 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 connected, intelligible speech is contrasted with various acoustic baselines, some studies have reported left-dominant activation patterns (Scott et al., 2000; Narain et al., 2003), leading some authors to argue for a fully left-lateralized network for speech perception (Scott et al., 2000; Rauschecker and Scott, 2009). Other studies report bilateral activation even when acoustic controls are subtracted out of the activation pattern (Okada et al., 2010; for a review, see Hickok and Poeppel, 2007). The issue is still being actively debated within the functional imaging literature, although recent reviews and meta-analyses support the conjecture of bilateral STG/STS involvement (Turkeltaub and Coslett, 2010; Price, 2012; Schirmer et al., 2012).

Computational asymmetries

The hypothesis that sublexical-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

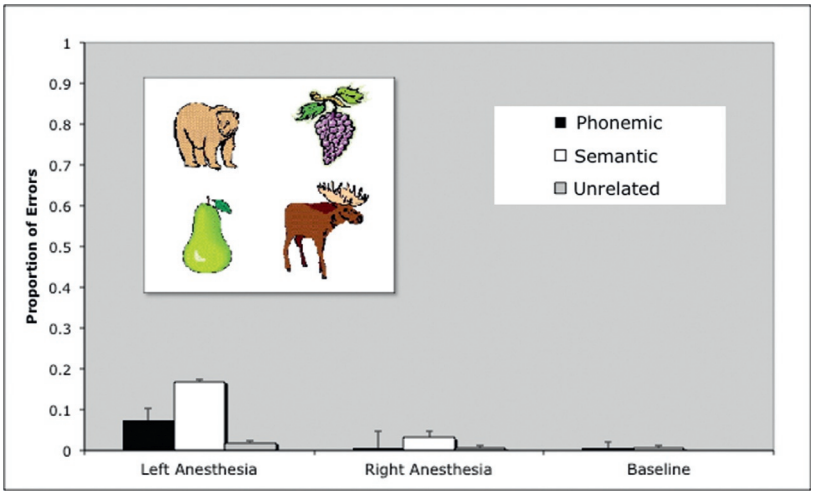

Fig. 8.2. Auditory comprehension performance during Wada procedure. Data show mean error rate on a four-alternative forced-choice auditory comprehension task with phonemic, semantic, and unrelated foils (inset shows sample stimulus card for the spoken target word, bear) in 20 patients undergoing clinically indicated Wada procedures. Error rate is shown as a function error type and amytal condition: left-hemisphere injection, right-hemisphere injection, and baseline. Note overall low error rate, even with left-hemisphere injection, and the dominance of semantic misselections when errors occur. (Modified from Hickok et al., 2008.)

in the processing of acoustic/speech information (Zatorre et al., 2002; Boemio et al., 2005; Giraud et al., 2007; Hickok and Poeppel, 2007; Abrams et al., 2008). The basis of these differences is currently being debated. One view, arguing for a domain-general perspective for all sounds, is that the difference turns on the selectivity for temporal (left-hemisphere) versus spectral (right-hemisphere) resolution (Zatorre et al., 2002; Obleser et al., 2008). That is, the left hemisphere may be particularly well suited for resolving rapid acoustic change (such as a formant transition), while the right hemisphere may have an advantage in resolving spectral frequency information. A closely related proposal is that the two hemispheres differ in terms of their preferred "sampling rate," with some left auditory cortical regions incorporating at bias for faster-rate (25–50 Hz) sampling and the right hemisphere for slower-rate sampling (4–8 Hz) (Poeppel, 2003). These two proposals are not incompatible as there is a relation between sampling rate and spectral vs temporal resolution: rapid sampling allows the system to detect changes that occur over short timescales, but sacrifices spectral resolution, and vice versa (Zatorre et al., 2002).

Further research is needed to address these hypotheses. 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 lexical meaning (Hickok and Poeppel, 2007), similar to well-accepted dual-route models of reading (phoneme-to-grapheme conversion and whole-word routes) (Coltheart et al., 1993). Although the parallel-pathway view differs from standard models of speech recognition (McClelland and Elman, 1986; Marslen-Wilson, 1987; Luce and Pisoni, 1998), wherein the processor proceeds from small to larger units in serial stages, it is consistent with the fact that speech contains redundant cues to phonemic information (e.g., in the speech envelope and fine spectral structure cues) and with behavioral evidence suggesting that the speech system can take advantage of these different cues (Remez et al., 1981; Shannon et al., 1995). It is worth bearing in mind that such computational asymmetries apply to all sounds that the auditory system analyzes. They reflect properties of neuronal ensembles that are like filters acting on any incoming signal. Specialization is likely to occur at the next stage at which signals are translated into a format suitable for lexical access.

Phonologic processing and the superior temporal sulcus

Beyond the earliest stages of speech recognition there is accumulating evidence that portions of the STS are important for representing and/or processing phonologic information (Price et al., 1996; Binder et al., 2000; Hickok and Poeppel, 2004, 2007; Indefrey and Levelt, 2004; Liebenthal et al., 2005). The STS is activated by language tasks that require access to phonologic information, including both the perception and production of speech (Indefrey and Levelt, 2004), and during active maintenance of phonemic information (Buchsbaum et al., 2001; Hickok et al., 2003). Portions of the STS seem to be relatively selective for acoustic signals that contain

phonemic information when compared to complex non-speech signals (yellow shaded portion of Fig. 8.1) (Narain et al., 2003; Liebenthal et al., 2005; Hickok and Poeppel, 2007; Okada et al., 2010). STS activation can be modulated by the manipulation of psycholinguistic variables that tap phonologic networks (Okada and Hickok, 2006), such as phonologic neighborhood density (the number of words that sound similar to a target word), and this region shows neural adaptation effects to phonologic-level information (Vaden et al., 2010).

One currently unresolved question concerns the relative contribution of anterior versus posterior STS regions in phonologic processing. Lesion evidence indicates that damage to posterior temporal-lobe areas is most predictive of auditory comprehension deficits (Bates et al., 2003) and a majority of functional imaging studies targeting phonologic processing in perception have identified regions in the posterior half of the STS (Hickok and Poeppel, 2007). Other studies, however, have reported anterior STS activation in perceptual speech tasks (Mazoyer et al., 1993; Scott et al., 2000; Narain et al., 2003; Spitsyna et al., 2006). These studies typically involved sentence-level stimuli, raising the possibility that anterior STS regions may be responding to some other aspect of the stimulus, such as its syntactic or prosodic organization (Friederici et al., 2000; Humphries et al., 2001, 2005, 2006; Vandenberghe et al., 2002). Recent electrophysiologic work supports the hypothesis that the left anterior temporal lobe (ATL) is critical to elementary structure building (Bemis and Pylkkanen, 2011) in line with the view that intelligibility tasks tap into additional operations beyond speech recognition. It will, in any case, be important in future work to understand the role of various portions of the STS in auditory speech perception and language processing.

Lexical-semantic access

During auditory comprehension, the goal of speech processing is to use phonologic information to access words and conceptual-semantic representations that are critical to comprehension. The dual-stream model holds that conceptual-semantic representations are widely distributed throughout the cortex. However, a more focal system serves as a computational interface that maps between phonologic-level representations of words or morphologic roots and distributed conceptual representations (Hickok and Poeppel, 2000, 2004, 2007; Lau et al., 2008). This interface is not the site for storage of conceptual information. Instead, it is hypothesized to store information regarding the relation (or correspondences) between phonologic information on the one hand and conceptual information on the other. Most authors agree that the temporal lobes play a critical role in this process, but there is disagreement regarding the role of anterior versus posterior regions. The evidence for both of these viewpoints is briefly presented below.

Damage to posterior temporal-lobe regions, particularly along the middle temporal gyrus, has long been associated with auditory comprehension deficits (Damasio, 1991; Dronkers et al., 2000; Bates et al., 2003), an effect confirmed in a large-scale study involving 101 patients (Bates et al., 2003). We infer that these deficits are primarily postphonemic in nature, as phonemic deficits following unilateral lesions to this area are mild (Hickok and 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 and Boatman, 2003). Functional imaging studies have also implicated posterior middle temporal regions in lexical-semantic processing (Binder et al., 1997; Rissman et al., 2003; Rodd et al., 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. Electrophysiologic studies have successfully used paradigms building on the N400 response to study lexical-semantic processing. This response is very sensitive to a range of variables known to implicate lexical-level properties. A review of that literature (including source localization studies of the N400) also suggests that the posterior middle temporal gyrus plays a key role, although embedded in a network of anterior temporal, parietal, and inferior frontal regions (Lau et al., 2008).

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 (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 phonologic and conceptual representations and indeed appear to involve more general semantic integration (Patterson et al., 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 (Mazoyer et al., 1993; Friederici et al., 2000; Humphries et al., 2001, 2005; 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). Recent electrophysiologic data (e.g. Brennan and Pylkkanen, 2012; Bemis and Pylkkanen, 2013) also implicate the left ATL in elementary structure building. Damage to the ATL has also been linked to deficits in comprehending complex syntactic structures (Dronkers et al., 2004). However, data from semantic dementia are 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 (Vandenberghe et al., 2002; Humphries et al., 2005, 2006); however, the neuropsychologic evidence remains equivocal.

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 and Mishkin, 1982). More recently, there has been some convergence on the idea that the dorsal stream supports auditory-motor integration (Hickok and Poeppel, 2000, 2004, 2007; Wise et al., 2001; Scott and Wise, 2004; Rauschecker and Scott, 2009; Rauschecker, 2011). 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 (Milner and Goodale, 1995; Andersen, 1997).

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 direct link between sensory and motor representations of words 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 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 (phonologic form) of a word. We know that the targets are auditory in nature because manipulating a speaker's auditory feedback during speech production results in compensatory changes in motor speech acts (Houde and Jordan, 1998; Larson et al., 2001; Purcell and Munhall, 2006). For example, if a subject is asked to produce one vowel and the feedback that she hears is manipulated so that it sounds like another vowel, 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 et al., 1998). The role of auditory input is nowhere more apparent than in development, where the child must use acoustic information in the linguistic environment to shape vocal tract movements that must reproduce those sounds.

A great deal of progress has been made in mapping the neural organization of sensorimotor integration for speech. Early functional imaging studies identified an auditory-related area in the left planum temporale region as involved in speech production (Hickok et al., 2000; Wise et al., 2001). Subsequent studies showed that this left-dominant region, dubbed Spt for its location in the Sylvian fissure at the parietal–temporal boundary (Fig. 8.3A) (Hickok et al., 2003), exhibited a number of properties characteristic of sensorimotor integration areas such as those found in macaque parietal cortex (Andersen, 1997; Colby and Goldberg, 1999). Most fundamentally, Spt exhibits sensorimotor response properties, activating both during the passive perception of speech and during covert (subvocal) speech articulation (Buchsbaum et al., 2001, 2005; Hickok et al., 2003), and further that different subregional patterns of activity are apparent during the sensory and motor phases of the task (Hickok et al., 2009), likely reflecting the activation of different neuronal subpopulations (Dahl et al., 2009), some sensory- and others motor-weighted. Figure 8.3B–D shows examples of the sensory-motor response properties of Spt and the patchy organization of this region for sensory- versus motor-weighted voxels (Fig. 8.3C, inset).

Spt is not speech-specific; its sensorimotor responses are equally robust when the sensory stimulus consists of

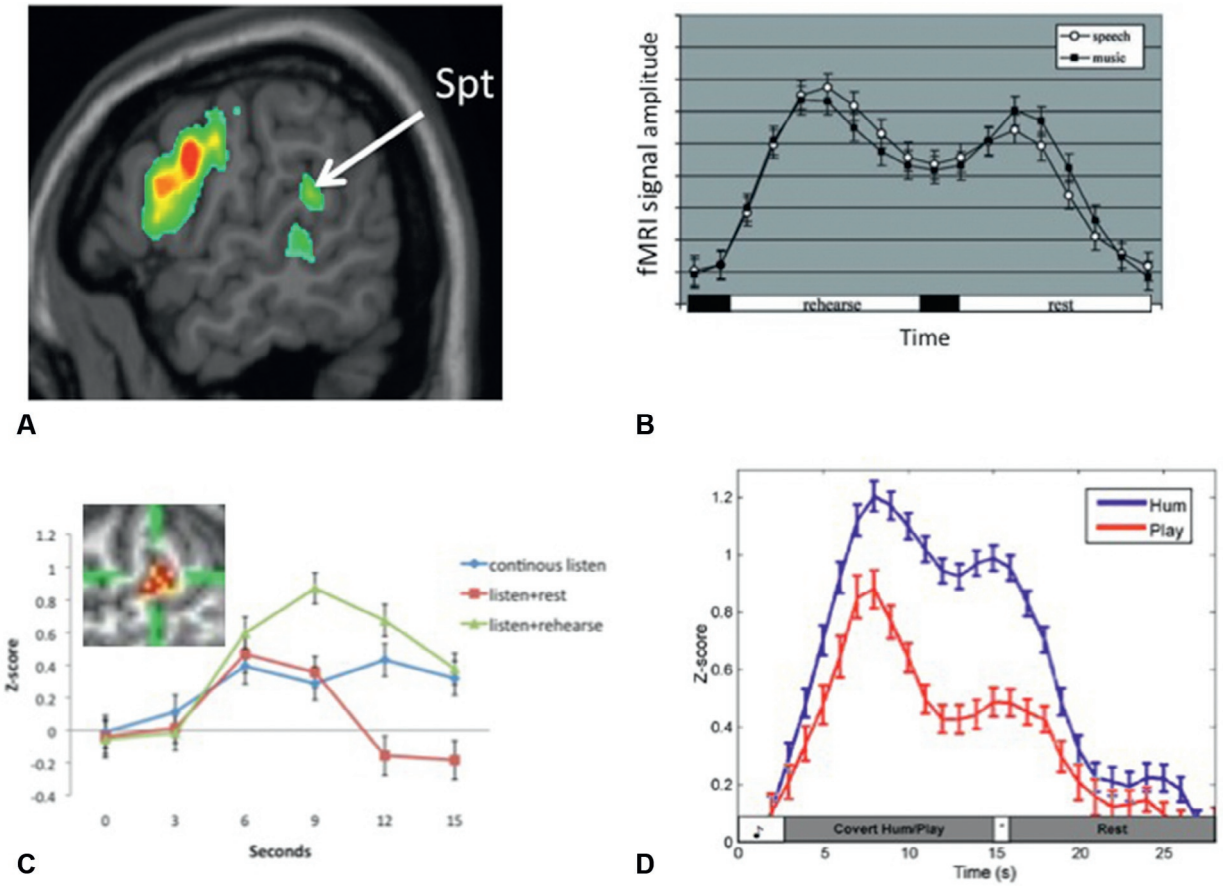

Fig. 8.3. Location and functional properties of area Sylvian parietal–temporal (Spt) junction. (**A**) Activation map for covert speech articulation (rehearsal of a set of non-words). (**B**) Activation timecourse (functional magnetic resonance imaging (fMRI) signal amplitude) in Spt during a sensorimotor task for speech and music. A trial is composed of 3 seconds of auditory stimulation followed by 15 seconds of covert rehearsal/humming of the heard stimulus, followed by 3 seconds of auditory stimulation, followed by 15 seconds of rest. The two humps represent the sensory responses, the valley between the humps is the motor (covert rehearsal) response, and the baseline values at the onset and offset of the trial reflect resting activity levels. Note similar response to both speech and music. (Adapted from Hickok et al., 2003.) (**C**) Activation timecourse in Spt in three conditions: continuous speech (15 seconds, blue curve), listen + rest (3 seconds speech, 12 seconds rest, red curve), and listen + covert rehearse (3 seconds speech, 12 seconds rehearse, green curve). The pattern of activity within Spt (inset) was found to be different for listening to speech compared to rehearsing speech assessed at the end of the continuous listen versus listen + rehearse conditions despite the lack of a significant signal amplitude difference at that time point. (Adapted from Hickok et al., 2009.) (**D**) Activation timecourse in Spt in skilled pianists performing a sensorimotor task involving listening to novel melodies and then covertly humming them (blue curve) vs listening to novel melodies and imagining playing them on a keyboard (red curve). This indicates that Spt is relatively selective for vocal tract actions. (Reproduced with permission from Hickok, 2009.) (Reproduced from Hickok and Buchsbaum, 2003.)

tonal melodies and (covert) humming is the motor task (see the two curves in Fig. 8.3B) (Hickok et al., 2003). Activity in Spt is highly correlated with activity in the pars opercularis (Buchsbaum et al., 2001, 2005), which is the posterior sector of Broca's region. White-matter tracts identified via diffusion tensor imaging suggest that Spt and the pars opercularis are densely connected anatomically (for review, see Friederici, 2009; Rogalsky and Hickok, 2011). Finally, consistent with some sensorimotor integration areas in the monkey parietal lobe (Andersen, 1997; Colby and Goldberg, 1999), Spt appears to be motor-effector-selective, responding more robustly when the motor task involves the vocal tract than the manual effectors (Fig. 8.2D) (Pa and Hickok, 2008). More broadly, Spt is situated in the middle of a network of auditory (STS) and motor (pars opercularis, premotor cortex) regions (Buchsbaum et al., 2001, 2005; Hickok et al., 2003), perfectly positioned both functionally and anatomically to support sensorimotor integration for speech and related vocal tract functions.

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 (H. Damasio, 1991; A. R. Damasio, 1992), 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 (Damasio and Damasio, 1980; Goodglass, 1992; Baldo et al., 2008), and the lesion distribution overlaps with the location of functional area Spt (Fig. 8.4) (Buchsbaum et al., 2011). 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 verbatim repetition of strings of speech with little semantic constraint (Goodglass, 1992, 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 and Poeppel, 2000, 2004, 2007).

Recent theoretic work has clarified the computational details underlying auditory-motor integration in the dorsal stream. Drawing on advances in understanding motor control generally, speech researchers have emphasized the role of internal forward models in speech motor control (Golfinopoulos et al., 2010; Hickok et al., 2011; Houde and Nagarajan, 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 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 (Guenther et al., 1998; Golfinopoulos et al., 2010; Hickok et al., 2011; Houde and Nagarajan, 2011). One such model is shown in Figure 8.5 (Hickok et al., 2011). Input to the system comes from a lexical-conceptual network as assumed by psycholinguistic models of speech production (Dell et al., 1997; Levelt et al., 1999). In between the input/output system is a phonologic system that is split into two components, corresponding to sensory input and motor output subsystems and mediated by a sensorimotor translation system, which corresponds to area Spt (Buchsbaum et al., 2001; Hickok et al., 2003, 2009). Parallel inputs to sensory and motor systems are needed to explain neuropsychologic observations (Jacquemot et al., 2007), such as conduction aphasia, as we will see below. Inputs to the auditory-phonologic 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 conduction aphasia. A lesion to Spt would disrupt the ability to generate forward predictions in auditory cortex and

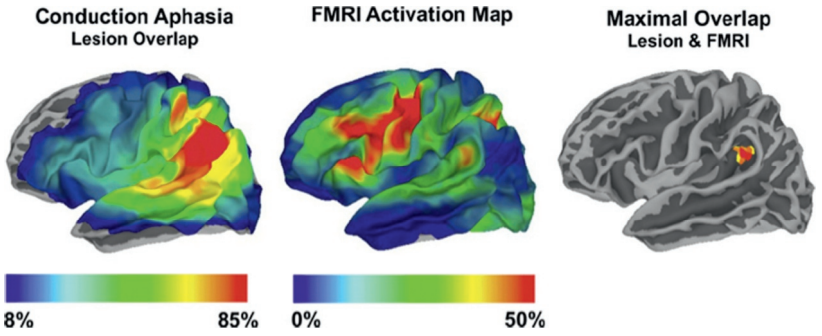

Fig. 8.4. Relation between lesions associated with conduction aphasia and the cortical auditory-motor network. A comparison of conduction aphasia, an auditory-motor task (listening to and then repeating back speech) in functional magnetic resonance imaging (fMRI), and their overlap. The uninflated surface in the left panel shows the regional distribution lesion overlap in patients with conduction aphasia (maximum is 12/14 or 85% overlap). The middle panel shows the auditory-motor network in the fMRI analysis. The right panel shows the area of maximal overlap between the lesion and fMRI surfaces (lesion >85% overlap and significant fMRI activity). (Modified from Buchsbaum et al., 2011.)

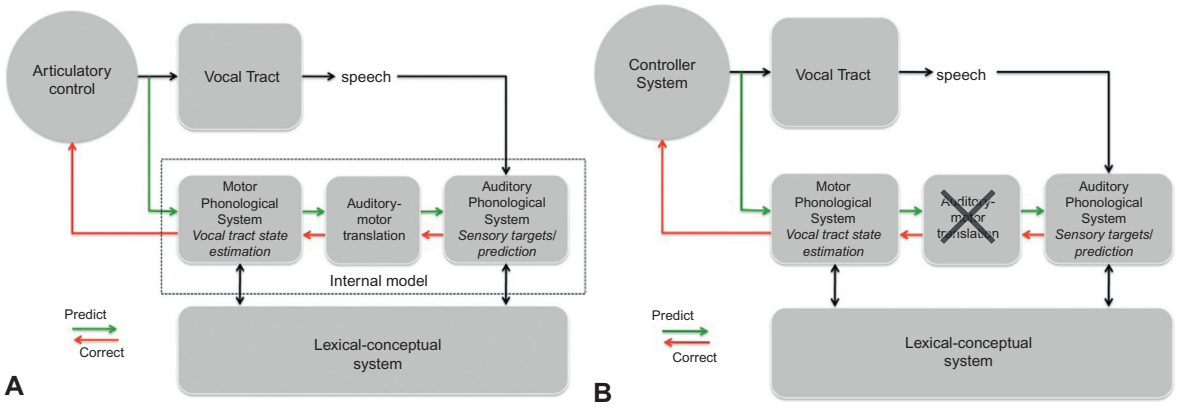

Fig. 8.5. An integrated state feedback control (SFC) model of speech production. (**A**) Speech models derived from the feedback control, psycholinguistic, and neurolinguistic literatures are integrated into one framework, presented here. The architecture is fundamentally that of an SFC system with a controller, or set of controllers (Haruno et al., 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 generate 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 phonologic system," which corresponds to the neurolinguistically elucidated phonologic 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 phonologic system," which corresponds to the neurolinguistically elucidated phonologic input lexicon, and is localized to the superior temporal gyrus/superior temporal sulcus. Motor and auditory phonologic systems are linked via an auditory-motor translation system, localized to the Sylvian parietal–temporal junction. The system is activated via parallel inputs from the lexical-conceptual system to the motor and auditory phonologic 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 phonologic systems is 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. (Reproduced with permission from Hickok and Poeppel, 2004; Hickok et al., 2011.)

thereby the ability to perform internal feedback monitoring, making errors more frequent than in an unimpaired system (Fig. 8.5B). However, this would not disrupt the activation of auditory targets via the lexical-semantic system, thus leaving patients 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.

CLINICAL CORRELATES OF THE DUAL-STREAM MODEL

The dual-stream model, like the classic Wernicke–Lichtheim model, provides an account of the major clinical aphasia syndromes (Hickok and Poeppel, 2004). Within the dual-stream model, Broca's aphasia and conduction aphasia are considered to be dorsal stream-related syndromes, while Wernicke's aphasia, word deafness, and transcortical sensory aphasia are considered ventral-stream syndromes. We have already noted that conduction aphasia can be conceptualized as a disruption of auditory-motor integration resulting from damage to area Spt. Broca's aphasia can be viewed as a disruption to representations that code for speech-related actions at multiple levels, from coding low-level phonetic features, to sequences of syllables, to sequences of words in structured sentences. Although Broca's area and Broca's aphasia are widely considered to be associated with deficits in receptive syntactic processing (Caramazza and Zurif, 1976; Grodzinsky, 2000), this issue is now being seriously questioned and remains debatable (Rogalsky and Hickok, 2011).

Word deafness is the "lowest-level" ventral-stream syndrome, according to the dual-stream model, affecting the processing of phonemic information during speech recognition. This differs from classic interpretations of word deafness as a disconnection syndrome (Geschwind, 1965). Due to the key role that auditory systems play in speech production, as discussed above, we should expect that disruption to auditory speech systems, as in word deafness, will impact production as well. Although the canonic description of word deafness

is a syndrome in which speech production is preserved, the majority of case descriptions that provide information on the speech output of word-deaf patients report the presence of paraphasic errors (Buchman et al., 1986).

Wernicke's aphasia is explained in terms of damage to multiple ventral-stream processing levels in the dual-stream model. Given the rather extensive posterior lesions that are typically required to yield a chronic Wernicke's aphasia (Dronkers and Baldo, 2009), it is likely that this syndrome results from damage to auditory-motor area Spt, left-hemisphere auditory areas, and posterior middle temporal lexical-semantic interface systems. Such damage can explain the symptom complex: relatively good phonologic-level speech recognition (due to the bilateral organization, as described above), poor comprehension at the higher semantic level (due to damage to lexical-semantic interface systems), fluent speech (due to preserved motor-speech systems), poor repetition (due to disruption of auditory-motor interface network), and paraphasic errors (due to disruption of auditory motor interface network).

Transcortical sensory aphasia, which is similar to Wernicke's aphasia but with preserved repetition, is conceptualized as a functionally more focal deficit involving the lexical-semantic interface network but sparing the auditory-motor network. Damage to the lexical-semantic interface explains the poor comprehension, while sparing of the auditory-motor interface explains the preserved repetition.

SEX DIFFERENCES IN LANGUAGE ORGANIZATION

Substantial evidence exists for sexual dimorphism in the brain (Cahill, 2006), which raises the question of whether there are sex differences in organization within the dorsal and/or ventral speech streams. This issue has not been thoroughly investigated, in part because existing evidence for sex differences in language-related brain function has not yielded consistent results (Wallentin, 2009). More work is needed to address this question.

SUMMARY

Dual-stream models of cortical organization have proven useful in understanding both language and visual-related systems and indeed have been a recurrent theme in neural models stretching back more than a century (Wernicke, 1874/1977). Thus, the general concept underlying the model – that the brain must interface sensory information with two different systems, conceptual and motor – is not only intuitively appealing but has a proven track record across domains. In the language domain, the dual-stream model provides an explanation of classic language disorders (Hickok and Poeppel, 2004; Hickok et al., 2011) and provides a framework for integrating and unifying research across psycholinguistic, neurolinguistic, and neurophysiologic traditions. Recent work has shown that still further integration with motor control models is possible (Hickok et al., 2011). All of this suggests that the dual-stream framework is on the right track as a model of language organization and provides a rich context for guiding future research.

References

Abrams DA, Nicol T, Zecker S et al. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. J Neurosci 28 (15): 3958–3965.

Andersen R (1997). Multimodal integration for the representation of space in the posterior parietal cortex. Phil Trans Roy Soc Lond B Bio Sci 352: 1421–1428.

Anderson JM, Gilmore R, Roper S et al. (1999). Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model. Brain Lang 70: 1–12.

Baker E, Blumsteim SE, Goodglass H (1981). Interaction between phonological and semantic factors in auditory comprehension. Neuropsychologia 19: 1–15.

Baldo JV, Klostermann EC, Dronkers NF (2008). It's either a cook or a baker: patients with conduction aphasia get the gist but lose the trace. Brain Lang 105 (2): 134–140.

Bates E, Wilson SM, Saygin AP et al. (2003). Voxel-based lesion-symptom mapping. Nat Neurosci 6 (5): 448–450.

Bemis DK, Pylkkanen L (2011). Simple composition: a magnetoencephalography investigation into the comprehension of minimal linguistic phrases. J Neurosci 31 (8): 2801–2814.

Bemis DK, Pylkkanen L (2013). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. Cereb Cortex 23 (8): 1859–1873.

Binder JR, Frost JA, Hammeke TA et al. (1997). Human brain language areas identified by functional magnetic resonance imaging. J Neurosci 17: 353–362.

Binder JR, Frost JA, Hammeke TA et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10: 512–528.

Boemio A, Fromm S, Braun A et al. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci 8 (3): 389–395.

Brennan J, Pylkkanen L (2012). The time-course and spatial distribution of brain activity associated with sentence processing. Neuroimage 60 (2): 1139–1148.

Buchman AS, Garron DC, Trost-Cardamone JE et al. (1986). Word deafness: One hundred years later. J Neurol Neurosurg Psychiatr 49: 489–499.

Buchsbaum B, Hickok G, Humphries C (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. Cognit Sci 25: 663–678.

Buchsbaum BR, Olsen RK, Koch P et al. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. Neuron 48 (4): 687–697.

Buchsbaum BR, Baldo J, Okada K et al. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory – an aggregate analysis of lesion and fMRI data. Brain Lang 119 (3): 119–128.

Cahill L (2006). Why sex matters for neuroscience. Nat Rev Neurosci 7 (6): 477–484.

Caramazza A, Zurif EB (1976). Dissociation of algorithmic and heuristic processes in sentence comprehension: Evidence from aphasia. Brain Lang 3: 572–582.

Colby CL, Goldberg ME (1999). Space and attention in parietal cortex. Annu Rev Neurosci 22: 319–349.

Coltheart M, Curtis B, Atkins P et al. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. Psychol Rev 100: 589–608.

Dahl CD, Logothetis NK, Kayser C (2009). Spatial organization of multisensory responses in temporal association cortex. J Neurosci 29 (38): 11924–11932.

Damasio H (1991). Neuroanatomical correlates of the aphasias. In: M Sarno (Ed.), Acquired aphasia, 2nd edn. Academic Press, San Diego, pp. 45–71.

Damasio AR (1992). Aphasia. New England Journal of Medicine 326: 531–539.

Damasio H, Damasio AR (1980). The anatomical basis of conduction aphasia. Brain 103: 337–350.

Dell GS, Schwartz MF, Martin N et al. (1997). Lexical access in aphasic and nonaphasic speakers. Psychol Rev 104: 801–838.

Dijkerman HC, de Haan EH (2007). Somatosensory processes subserving perception and action. Behav Brain Sci 30 (2): 189–201, discussion 201–139.

Dronkers N, Baldo J (2009). Language: Aphasia. In: LR Squire (Ed.), Encyclopedia of Neuroscience, Vol. 5. Academic Press, Oxford, pp. 343–348.

Dronkers NF, Redfern BB, Knight RT (2000). The neural architecture of language disorders. In: MS Gazzaniga (Ed.), The new cognitive neurosciences, MIT Press, Cambridge, MA, pp. 949–958.

Dronkers NF, Wilkins DP, Van Valin Jr RD et al. (2004). Lesion analysis of brain regions involved in language comprehension. Cognition 92: 145–177.

Friederici AD (2009). Pathways to language: fiber tracts in the human brain. Trends Cogn Sci 13 (4): 175–181.

Friederici AD, Meyer M, von Cramon DY (2000). Auditory languge comprehension: An event-related fMRI study on the processing of syntactic and lexical information. Brain Lang 74: 289–300.

Geschwind N (1965). Disconnexion syndromes in animals and man. Brain 88 (237–294): 585–644.

Giraud AL, Kleinschmidt A, Poeppel D et al. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron 56 (6): 1127–1134.

Golfinopoulos E, Tourville JA, Guenther FH (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. Neuroimage 52 (3): 862–874.

Goodglass H (1992). Diagnosis of conduction aphasia. In: SE Kohn (Ed.), Conduction aphasia, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 39–49.

Goodglass H (1993). Understanding aphasia, Academic Press, San Diego.

Gorno-Tempini ML, Dronkers NF, Rankin KP et al. (2004). Cognition and anatomy in three variants of primary progressive aphasia. Ann Neurol 55 (3): 335–346.

Grodzinsky Y (2000). The neurology of syntax: Language use without Broca's area. Behav Brain Sci 23: 1–21.

Guenther FH, Hampson M, Johnson D (1998). A theoretical investigation of reference frames for the planning of speech movements. Psychol Rev 105: 611–633.

Haruno M, Wolpert DM, Kawato M (2001). Mosaic model for sensorimotor learning and control. Neural Comput 13 (10): 2201–2220.

Hickok G (2009). The functional neuroanatomy of language. Phys Life Rev 6: 121–143.

Hickok G, Buchsbaum B (2003). Temporal lobe speech perception systems are part of the verbal working memory circuit: evidence from two recent fMRI studies. Behav Brain Sci 26: 740–741.

Hickok G, Poeppel D (2000). Towards a functional neuroanatomy of speech perception. Trends Cogn Sci 4: 131–138.

Hickok G, Poeppel D (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. Cognition 92: 67–99.

Hickok G, Poeppel D (2007). The cortical organization of speech processing. Nat Rev Neurosci 8 (5): 393–402.

Hickok G, Erhard P, Kassubek J et al. (2000). A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. Neurosci Lett 287: 156–160.

Hickok G, Buchsbaum B, Humphries C et al. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. J Cogn Neurosci 15: 673–682.

Hickok G, Okada K, Barr W et al. (2008). Bilateral capacity for speech sound processing in auditory comprehension: evidence from Wada procedures. Brain Lang 107 (3): 179–184.

Hickok G, Okada K, Serences JT (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. J Neurophysiol 101 (5): 2725–2732.

Hickok G, Houde J, Rong F (2011). Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69 (3): 407–422.

Houde JF, Jordan MI (1998). Sensorimotor adaptation in speech production. Science 279: 1213–1216.

Houde JF, Nagarajan SS (2011). Speech production as state feedback control. Front Hum Neurosci 5: 82.

Humphries C, Willard K, Buchsbaum B et al. (2001). Role of anterior temporal cortex in auditory sentence comprehension: An fMRI study. Neuroreport 12: 1749–1752.

Humphries C, Love T, Swinney D et al. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. Hum Brain Mapp 26: 128–138.

Humphries C, Binder JR, Medler DA et al. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. J Cogn Neurosci 18 (4): 665–679.

Indefrey P, Levelt WJ (2004). The spatial and temporal signatures of word production components. Cognition 92 (1–2): 101–144.

Jacquemot C, Dupoux E, Bachoud-Levi AC (2007). Breaking the mirror: asymmetrical disconnection between the phonological input and output codes. Cogn Neuropsychol 24 (1): 3–22.

Larson CR, Burnett TA, Bauer JJ et al. (2001). Comparison of voice F0 responses to pitch-shift onset and offset conditions. J Acoust Soc Am 110 (6): 2845–2848.

Lau EF, Phillips C, Poeppel D (2008). A cortical network for semantics: (de)constructing the N400. Nat Rev Neurosci 9 (12): 920–933.

Levelt WJM, Roelofs A, Meyer AS (1999). A theory of lexical access in speech production. Behavioral and Brain Sciences 22 (1): 1–75.

Liebenthal E, Binder JR, Spitzer SM et al. (2005). Neural substrates of phonemic perception. Cereb Cortex 15 (10): 1621–1631.

Luce PA, Pisoni DB (1998). Recognizing spoken words: the neighborhood activation model. Ear Hear 19: 1–36.

Luria AR (1970). Traumatic aphasia, Mouton, The Hague.

Marslen-Wilson WD (1987). Functional parallelism in spoken word-recognition. Cognition 25: 71–102.

Mazoyer BM, Tzourio N, Frak V et al. (1993). The cortical representation of speech. J Cogn Neurosci 5: 467–479.

McClelland JL, Elman JL (1986). The TRACE model of speech perception. Cogn Psychol 18: 1–86.

Miceli G, Gainotti G, Caltagirone C et al. (1980). Some aspects of phonological impairment in aphasia. Brain Lang 11: 159–169.

Miglioretti DL, Boatman D (2003). Modeling variability in cortical representations of human complex sound perception. Exp Brain Res 153 (3): 382–387.

Milner AD, Goodale MA (1995). The visual brain in action, Oxford University Press, Oxford.

Narain C, Scott SK, Wise RJ et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex 13 (12): 1362–1368.

Obleser J, Eisner F, Kotz SA (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J Neurosci 28 (32): 8116–8123.

Okada K, Hickok G (2006). Identification of lexical-phonological networks in the superior temporal sulcus using fMRI. Neuroreport 17: 1293–1296.

Okada K, Rong F, Venezia J et al. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. Cereb Cortex 20 (10): 2486–2495.

Pa J, Hickok G (2008). A parietal-temporal sensory-motor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians. Neuropsychologia 46: 362–368.

Patterson K, Nestor PJ, Rogers TT (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. Nat Rev Neurosci 8 (12): 976–987.

Poeppel D (2001). Pure word deafness and the bilateral processing of the speech code. Cognit Sci 25: 679–693.

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

Price CJ (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage 62 (2): 816–847.

Price CJ, Wise RJS, Warburton EA et al. (1996). Hearing and saying: the functional neuro-anatomy of auditory word processing. Brain 119: 919–931.

Purcell DW, Munhall KG (2006). Compensation following real-time manipulation of formants in isolated vowels. J Acoust Soc Am 119 (4): 2288–2297.

Rauschecker JP (1998). Cortical processing of complex sounds. Curr Opin Neurobiol 8 (4): 516–521.

Rauschecker JP (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hear Res 271 (1–2): 16–25.

Rauschecker JP, Scott SK (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12 (6): 718–724.

Remez RE, Rubin PE, Pisoni DB et al. (1981). Speech perception without traditional speech cues. Science 212: 947–950.

Rissman J, Eliassen JC, Blumstein SE (2003). An event-related FMRI investigation of implicit semantic priming. J Cogn Neurosci 15 (8): 1160–1175.

Rodd JM, Davis MH, Johnsrude IS (2005). The neural mechanisms of speech comprehension: fMRI studeis of semantic ambiguity. Cereb Cortex 15: 1261–1269.

Rogalsky C, Hickok G (2011). The role of Broca's area in sentence comprehension. J Cogn Neurosci 23: 1664–1680.

Rogalsky C, Pitz E, Hillis AE et al. (2008). Auditory word comprehension impairment in acute stroke: relative contribution of phonemic versus semantic factors. Brain Lang 107 (2): 167–169.

Rogalsky C, Love T, Driscoll D et al. (2011). Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system. Neurocase 17 (2): 178–187.

Schirmer A, Fox PM, Grandjean D (2012). On the spatial organization of sound processing in the human temporal lobe: a meta-analysis. Neuroimage 63 (1): 137–147.

Scott SK, Wise RJ (2004). The functional neuroanatomy of prelexical processing in speech perception. Cognition 92 (1–2): 13–45.

Scott SK, Blank CC, Rosen S et al. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123: 2400–2406.

Shannon RV, Zeng F-G, Kamath V et al. (1995). Speech recognition with primarily temporal cues. Science 270: 303–304.

Spitsyna G, Warren JE, Scott SK et al. (2006). Converging language streams in the human temporal lobe. J Neurosci 26 (28): 7328–7336.

Turkeltaub PE, Coslett HB (2010). Localization of sublexical speech perception components. Brain Lang 114 (1): 1–15.

Ungerleider LG, Mishkin M (1982). Two cortical visual systems. In: DJ Ingle, MA Goodale, RJW Mansfield (Eds.), Analysis of visual behavior, MIT Press, Cambridge, MA, pp. 549–586.

Vaden Jr KI, Muftuler LT, Hickok G (2010). Phonological repetition-suppression in bilateral superior temporal sulci. Neuroimage 49: 1018–1023.

Vandenberghe R, Nobre AC, Price CJ (2002). The response of left temporal cortex to sentences. J Cogn Neurosci 14 (4): 550–560.

Wallentin M (2009). Putative sex differences in verbal abilities and language cortex: a critical review. Brain Lang 108 (3): 175–183.

Wernicke C (1874/1969). The symptom complex of aphasia: A psychological study on an anatomical basis. In: RS Cohen, MW Wartofsky (Eds.), Boston studies in the philosophy of science. D. Reidel, Dordrecht, pp. 34–97.

Wernicke C (1874/1977). Der aphasische Symptomencomplex: Eine psychologische Studie auf anatomischer Basis. In: GH Eggert (Ed.), Wernicke's works on aphasia: A sourcebook and review. Mouton, The Hague, pp. 91–145.

Wise RJS, Scott SK, Blank SC et al. (2001). Separate neural sub-systems within "Wernicke's area". Brain 124: 83–95.

Zatorre RJ, Belin P, Penhune VB (2002). Structure and function of auditory cortex: music and speech. Trends Cogn Sci 6: 37–46.