Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language

Gregory Hickok[a],*, David Poeppel[b]

[a]*University of California, Irvine, CA, USA*
[b]*University of Maryland, College Park, MD, USA*

Abstract

Despite intensive work on language–brain relations, and a fairly impressive accumulation of knowledge over the last several decades, there has been little progress in developing large-scale models of the functional anatomy of language that integrate neuropsychological, neuroimaging, and psycholinguistic data. Drawing on relatively recent developments in the cortical organization of vision, and on data from a variety of sources, we propose a new framework for understanding aspects of the functional anatomy of language which moves towards remedying this situation. The framework posits that early cortical stages of speech perception involve auditory fields in the superior temporal gyrus bilaterally (although asymmetrically). This cortical processing system then diverges into two broad processing streams, a ventral stream, which is involved in mapping sound onto meaning, and a dorsal stream, which is involved in mapping sound onto articulatory-based representations. The ventral stream projects ventro-laterally toward inferior posterior temporal cortex (posterior middle temporal gyrus) which serves as an interface between sound-based representations of speech in the superior temporal gyrus (again bilaterally) and widely distributed conceptual representations. The dorsal stream projects dorso-posteriorly involving a region in the posterior Sylvian fissure at the parietal–temporal boundary (area Spt), and ultimately projecting to frontal regions. This network provides a mechanism for the development and maintenance of "parity" between auditory and motor representations of speech. Although the proposed dorsal stream represents a very tight connection between processes involved in speech perception and speech production, it does not appear to be a critical component of the speech perception process under normal (ecologically natural) listening conditions, that is, when speech input is mapped onto a conceptual representation. We also propose some degree of bi-directionality in both the dorsal and ventral pathways. We discuss some recent empirical tests of this framework that utilize a range of

---

* Corresponding author. Department of Cognitive Sciences, University of California, Irvine, CA 92612, USA.
  E-mail address: gshickok@uci.edu (G. Hickok).

methods. We also show how damage to different components of this framework can account for the major symptom clusters of the fluent aphasias, and discuss some recent evidence concerning how sentence-level processing might be integrated into the framework.

1. Introduction and preliminaries

The functional anatomic framework for language which is presented in this paper is based on a rather old insight in language research dating back at least to the 19th century (e.g. Wernicke, 1874/1969), namely that sensory speech codes must minimally interface with two systems: a conceptual system and a motor–articulatory system. The existence of an interface with the conceptual system requires no motivation; such an interface is required if we are to comprehend the meaning of the words we hear. The need for an interface with the motor system may at first seem less obvious, but in fact, many areas of language science either explicitly or implicitly posit an auditory–motor connection. The simplest demonstration of this comes from development: infants must shape their articulatory gestures in a way that matches the phonetic structure of the language they are exposed to; yet the primary input to this motor learning task is acoustic. Therefore, there must be some mechanism for using auditory input to shape motor output (Doupe & Kuhl, 1999). And there is good reason to believe that this auditory–motor interface system is functional in adults as we will see below.

Recent work in the cortical organization of vision has also emphasized that sensory input must interface both with conceptual systems (for object recognition) and with motor systems (e.g. visually guided reaching/grasping) (Milner & Goodale, 1995). It has been demonstrated empirically that these two interface systems comprise functionally and anatomically differentiated (but probably interacting) processing streams in which a ventral (occipital–temporal) stream supports object recognition/understanding, and a dorsal (occipital–parietal) stream supports visuomotor integration functions (see Milner & Goodale, 1995 for a review).

The point of the present paper is to show that by thinking of aspects of language processing in terms of these two kinds of interfaces (sensory–conceptual and sensory–motor), and by making use of what is known about the cortical organization of the visual processing streams which also make use of a similar functional distinction, we can advance our understanding of the cortical organization of language. Indeed, we will argue that sensory representations of speech in auditory-related cortices (bilaterally) interface (i) with conceptual representations via projections to portions of the temporal lobe (the ventral stream), and (ii) with motor representation via projections to temporal–parietal regions (the dorsal stream).

Before presenting the details of our framework, it is worthwhile considering two related issues. One concerns the concept of linguistic specificity in the neural organization of

language, and the other concerns the impact of task selection in brain mapping studies of language.

It is common to find brain mapping studies of language which contrast the perception of a linguistic stimulus, say auditorily presented syllables or sentences, with a non-linguistic control, say, tone sequences, rotated speech, or time reversed speech. The goal of this approach is to identify cortical fields which are speech- or language-specific in their response properties, and therefore may represent dedicated cortical processing systems. The identification of language-specific processing systems is an interesting and important empirical enterprise, but is unlikely to yield, on its own, a complete understanding of the neural organization of language. The reasoning behind this assertion is as follows. Language processing systems can be viewed as a set of transformations over representations (not necessarily in series), for example, mapping between an acoustic input and a conceptual representation (as in comprehension), or between a conceptual representation and a sequence of motor gestures (as in production). Early stages of this mapping process on the input side – for example, cochlear, brain stem and thalamic processing as well as at least early cortical auditory mapping – likely perform transformations on the acoustic data that are relevant to linguistic as well as non-linguistic auditory perception. Because these early processing stages are not *uniquely* involved in language perception, they are often dismissed as being merely "auditory" areas and not relevant to understanding language processing. But clearly each stage in this analytic process interacts with other stages in important ways: the computations performed at one level depend on the input received from other levels, and therefore each transformation plays a role in the entire process of mapping sound onto meaning (or meaning onto motor articulation). Viewed in this way, a complete understanding of the functional anatomy of language will involve an understanding of each step in the translation between conceptual representations and the sensory and motor periphery, independent of the linguistic specificity of each of the computations. Of course, it may turn out to be important to determine the degree of linguistic specificity of each of these steps, but this question need not be answered to understand the relevant computations underlying the neural basis of language (e.g. we need not understand the role of orofacial articulators in eating before we understand their role in speech production). We argue that it is premature to dismiss areas that activate during language tasks as "non-linguistic" simply because they also respond vigorously to non-linguistic stimulation. Similarly, because we do not know how linguistic categories (e.g. phonetic, phonemic, etc.) map onto these neural systems, we will use these terms sparingly, and only in a very general sense. In some cases, we will resort to using generic terms, such as "sub-lexical" or "sound-based representation of speech", which reflects our agnosticism on the issue of how the processing stages we propose map onto traditional labels.

An important consequence of this view of language processes as a set of computations or mappings between representations is that the neural systems involved in a given language operation (task) will depend to some extent on what representation is being mapped onto. For example, speech input which is mapped onto a conceptual representation (as in comprehension tasks) will clearly involve a set of computations which is non-identical to those involved in mapping that same input onto a motor–articulatory representation (as in a repetition task). Of course, the mapping stages in these

two tasks will be shared up to some point, but they must diverge in accordance with the different requirements entailed by the endpoints of the mapping process. The upshot is that *the particular task which is employed to investigate the neural organization of language (that is, the mapping operation the subject is asked to compute) determines which neural circuit is predominantly activated*. This point is obvious in extreme cases, such as the example above, but the issue may also arise in more subtle cases: to what extent does mapping an auditorily presented sentence onto a judgment of grammaticality differ from mapping that input onto a representation of meaning? Considerations such as these will play an important role in our argument below.

2. Overview of the framework

The framework we have proposed (Hickok & Poeppel, 2000) and further develop here draws heavily on what is known about the functional anatomy of vision, and more recently audition, particularly the distinction that has been made between dorsal and ventral streams. Most of the discussion of dorsal and ventral streams in the literature centers on the concept of "where" and "what" pathways (Ungerleider & Mishkin, 1982). The fundamental distinction proposed by Ungerleider and Mishkin was that visual processing could be coarsely divided into two processing streams, a ventral stream projecting to inferior temporal areas which is involved in processing object identity (the "what" pathway), and a dorsal stream projecting to parietal areas which is involved in processing object location (the "where" pathway). In the last several years, however, there has been mounting evidence suggesting that the concept "where" may be an insufficient characterization of the dorsal stream (Milner & Goodale, 1995). Instead, it has been proposed that the dorsal visual stream is particularly geared for visuo-motor integration, as required in visually guided reaching or orienting responses.[1] According to this view, dorsal stream systems appear to compute coordinate transformations – for example, transform representations in retino-centric coordinates to head-, and body-centered coordinates – that allows visual information to interface with various motor-effector systems which act on that visual input (Andersen, 1997; Rizzolatti, Fogassi, & Gallese, 1997).

In the auditory system, a dorsal–ventral partitioning has also been proposed (Rauschecker, 1998). While there is general agreement regarding the role of the ventral stream in auditory "what" processing, the functional role of the dorsal stream is debated. Some groups argue for a dorsal "where" stream in the auditory system (Kaas & Hackett, 1999; Rauschecker, 1998; Romanski et al., 1999), whereas others have proposed that the dorsal stream is more involved in tracking changes in the frequency spectra of the auditory signal over time, a capacity which would make it particularly relevant for speech perception (Belin & Zatorre, 2000). We have put forward a third hypothesis, that the dorsal auditory stream is critical for *auditory–motor integration* (Hickok & Poeppel, 2000), similar to its role in the visual domain (see Wise et al., 2001 for a similar proposal). This system, we suggest, serves both linguistic and non-linguistic

---

[1] Processing of spatial location would, of course, be a sub-process in a system geared toward visuo-motor integration, making this view consistent with previous observations of the involvement of dorsal stream structures in spatial processing.
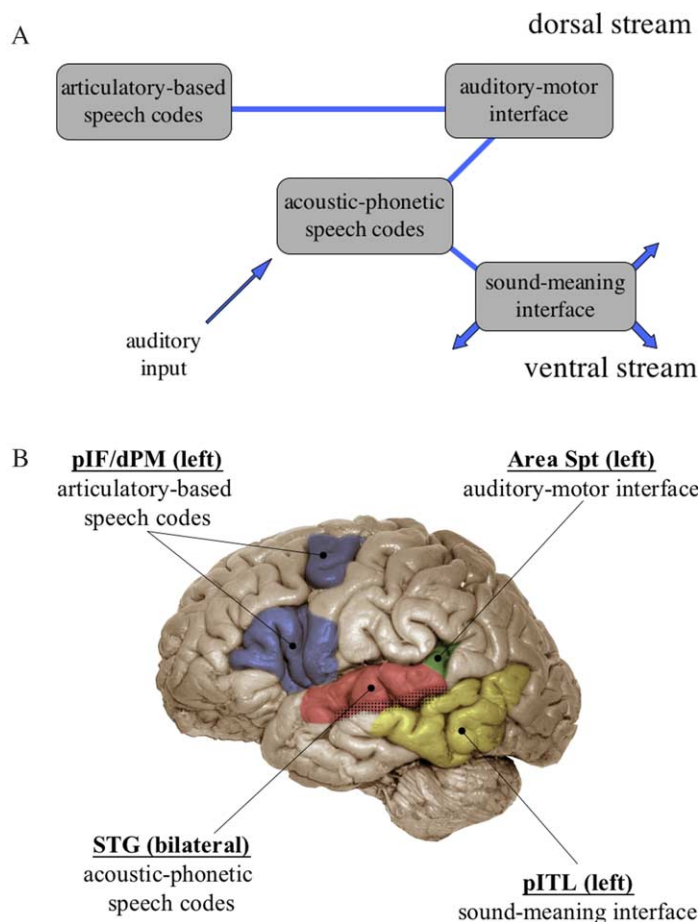
Fig. 1. (A) The proposed framework for the functional anatomy of language. See text for details. Adapted from Hickok and Poeppel (2000). (B) General locations of the model components shown on a lateral view of the brain; area boundaries are not to be interpreted as sharply as they are drawn. Note that the cortical territory associated with a given function in the model is not hypothesized to be dedicated to that function, although there may be subsystems with these fields which are functionally specialized. Delineation of frontal areas thought to support articulatory-based speech codes comes from the general distribution of activated areas in functional imaging studies of object naming and articulatory rehearsal processes (e.g. see Awh et al., 1996; Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Indefrey & Levelt, this volume). The stippled area (superior temporal sulcus) represents a region which appears to support phoneme-level representations (see text).

(e.g. orienting responses, aspects of musical ability) processes. A more detailed discussion of the necessity for an auditory–motor integration system for speech is taken up in Section 5 below.

With this background, we turn to a brief outline of the framework (Fig. 1A,B), which has been developed primarily in the context of single word processing. (We consider an extension of this proposal into sentence-level processing in Section 4.) The model posits that early cortical stages of speech perception involve auditory-responsive fields in

the superior temporal gyrus (STG) *bilaterally*.[2] This cortical processing system then diverges into two processing streams, a *ventral stream*, which is involved in mapping sound onto meaning, and a *dorsal stream*, which is involved in mapping sound onto articulatory-based representations. Although we have nothing new to say about the cortical organization of frontal speech-related areas, we have provided some candidate locations for this system in Fig. 1 (see figure legend) based on existing data from production-related studies. These regions include a posterior inferior frontal region that, depending on the study, includes various parts of Broca's area, frontal operculum/insula, the motor face area, and a more dorsal premotor site. Nothing in the present model turns on the precise location of these frontal-production related sites. The dorsal stream is fairly strongly left-lateralized; the ventral stream also appears to be left-dominant, but perhaps to a lesser degree.

The ventral stream projects ventro-laterally and involves cortex in the superior temporal sulcus (STS) and ultimately in the posterior inferior temporal lobe (pITL, i.e. portions of the middle temporal gyrus (MTG) and inferior temporal gyrus (ITG)).[3] These pITL structures serve as an interface between sound-based representations of speech in STG and widely distributed conceptual representations (Damasio, 1989). In psycholinguistic terms, this sound–meaning interface system may correspond to the lemma level of representation (Levelt, 1989).

The dorsal stream projects dorso-posteriorly toward the parietal lobe and ultimately to frontal regions. Based on available evidence in the literature (Jonides et al., 1998), we previously hypothesized that the posterior portion of this dorsal stream was located in the posterior parietal lobe (areas 7, 40). Recent evidence, however, suggests instead that the critical region is deep within the posterior aspect of Sylvian fissure at the *boundary* between the parietal and temporal lobes, a region we have referred to as area Spt (Sylvian–parietal–temporal) (Buchsbaum, Humphries, & Hickok, 2001; Hickok et al., 2003). Area Spt, then, is a crucial part of a network which performs a coordinate transform, mapping between auditory representations of speech and motor representations of speech. As far as the nature of the computations performed by this auditory–motor interface, we have in mind something like the neural network model proposed by Guenther and colleagues (Guenther, Hampson, & Johnson, 1998), in which articulatory gestures are planned in auditory space and then mapped onto motor representations. We also hypothesize that this network provides a mechanism for the development and maintenance of "parity" between auditory and motor representations of speech (Liberman & Mattingly, 1985). Although the proposed dorsal stream represents a very

---

[2] We had previously postulated the posterior half of the STG as the critical site, but recent evidence suggests that auditory representations of speech are organized hierarchically, and roughly concentrically around Heschl's gyrus. This places the critical region more anteriorly than we had previously claimed. See below for further discussion.

[3] In Hickok and Poeppel (2000) we proposed the temporal–parietal–occipital junction as a "good candidate" for the anatomical localization of systems involved in the sound–meaning interface. This was based on the distribution of lesions associated with transcortical sensory aphasia (see Section 4.3). The present claim involving pITL reflects recent observations from both lesion and imaging work suggesting a critical role of this area in lexical processing. This issue is discussed at length in Section 4.3.

tight connection between processes involved in speech perception and speech production, it does not appear to be a critical component of the speech perception process under normal (ecologically natural) listening conditions, that is when speech input is mapped onto a conceptual representation.

We also propose bi-directionality in both the dorsal and ventral pathways. Thus, in the ventral stream, pITL networks mediate the relation between sound and meaning both for perception and production (the involvement need not be symmetrical in perception and production). Similarly, we hypothesize that sectors of the *left* STG participate not only in sub-lexical aspects of the perception of speech, but also in sub-lexical aspects of the production of speech (again, perhaps non-symmetrically).[4] In the dorsal stream, we suggest that temporal–parietal systems can map auditory speech representations onto motor representations (as in verbatim repetition tasks, in which access to a motor-based representation is necessary), as well as map motor speech representations onto auditory speech representations. This sensory–motor loop in the dorsal stream provides the functional anatomic basis for verbal working memory (Baddeley, 1992), that is, the ability to use articulatory-based processes (rehearsal) to keep auditory-based representations (storage) active.

The extent to which the dorsal or ventral stream is utilized in a language task depends on the extent to which that task involves mapping between auditory and motor systems on the one hand, or between auditory and conceptual systems on the other. The involvement of these systems in a given task will also depend strongly on the strategies employed by individual subjects. A task which ostensibly involves only comprehension (say, passive sentence listening in a functional activation experiment) will primarily drive bilateral auditory, ventral stream areas, but may additionally recruit dorsal stream mechanisms if the subject uses articulatory re-mapping as an aid in task performance.

In the following sections we outline the evidence relevant to our proposal, focusing on data that have not received extensive treatment in our initial description of the framework (Hickok & Poeppel, 2000). We will also discuss how the proposal can be extended to encompass language processing beyond the word level, and how it relates to the classical symptom complex of aphasia.

3. Task dissociations in "speech perception"

One central thesis of our approach is that the execution of different linguistic tasks (functions) involves non-identical neural networks, even with stimulus conditions held constant. In this section we review evidence that supports this assumption in the domain of speech perception. In particular, the evidence shows that the ability to perform sub-lexical speech tasks (phoneme identification, rhyming tasks, and so on) double-dissociates from the ability to comprehend words (which presumably involves processing sub-lexical

---

[4] The left-lateralized involvement of the STG in production probably arises from the fact that this system is interfacing with motor planning systems which tend to be fairly strongly left-lateralized for sub-lexical aspects of speech.

information). This is a paradoxical result, on standard assumptions. Suppose word comprehension involves several stages of processing, as it typically assumed: acoustic–phonetic analysis → sub-lexical processing/representation (sequences of phonemes or syllabic representations) → lexical–semantic access → comprehension. Sub-lexical tasks (syllable discrimination/identification) presumably represent an attempt to isolate and study the early stages in this normal comprehension process, that is, the acoustic–phonetic analysis and/or the sub-lexical processing stage. The paradox, of course, stems from the fact that patients exist who cannot accurately perform syllable discrimination/identification tasks, yet have normal word comprehension: if sub-lexical tasks isolate and measure early stages of the word comprehension process, deficits on sub-lexical tasks should be highly predictive of auditory comprehension deficits, yet they are not. What we suggest is that performance on sub-lexical tasks involves neural circuits beyond (i.e. a superset of) those involved in the normal comprehension process. This is an important observation because there are many studies of the functional anatomy of "speech perception" that utilize sub-lexical tasks. Because sub-lexical tasks recruit neural circuits beyond those involved in word comprehension, the outcome of such studies may paint a misleading picture of the neural organization of speech perception, as it is used under more normal listening conditions.

3.1. Evidence from aphasia

The crucial observation from the aphasia literature is that there is not a strong correlation between performance on sub-lexical speech perception tasks and performance on auditory comprehension tasks (Basso, Casati, & Vignolo, 1977; Blumstein, Cooper, Zurif, & Caramazza, 1977; Miceli, Gainotti, Caltagirone, & Masullo, 1980). In fact, performance on the two classes of tasks doubly dissociates: patients have been reported who fail syllable discrimination and identification tasks yet have very good word-level auditory comprehension, and vice versa. Consider, for example, Table 1, which reproduces data reported in Miceli et al. (1980), showing a double-dissociation between performance on a phoneme discrimination task and a single-word auditory comprehension task in a series of 69 unselected right-handed aphasic patients.

Basso et al. (1977) report a similar dissociation in a series of 84 unilateral brain lesioned patients (22 right lesioned, 62 left lesioned of which 50 were aphasic). They used a phoneme identification task in the context of a voice-onset-time continuum (da–ta) and found that non-aphasics and all but one of the right lesioned patients performed within normal limits; but 74% of the aphasics were impaired on this task to some degree. Of interest was that the phoneme identification deficit was more common among non-fluent aphasics with good comprehension (10 of 11, 91%) than among fluent aphasics with poor comprehension (13 of 18, 72%). Furthermore, it was not simply the case that the non-fluent aphasics were only mildly impaired on the phoneme identification task: 36% (4 of 11) were classified as "very severely" impaired on this task showing "no trend towards correct identification" (p. 91). So 36% of the non-fluent

Table 1
Relation between a phoneme discrimination task (CCVC pairs) and a word-to-picture matching auditory comprehension task (four alternative forced choice with phonemic, semantic, and unrelated foils)

Word comprehension	Phoneme discrimination	
	Normal	Pathological
Normal	23	19
Pathological	9	15

Normal vs. pathological classification based on data from >60 normal age and education-level matched controls. For word comprehension, "Normal" subjects all scored 100% correct. Note the double-dissociation: 19 patients had normal word comprehension but pathological scores on the phoneme discrimination task, and nine patients had pathological scores on word comprehension yet were in the normal range on phoneme discrimination. Data from Table 3 in Miceli et al. (1980).

patients had good auditory comprehension despite being severely impaired in phoneme identification, whereas 28% (5 of 18) of the fluent patients had poor auditory comprehension yet performed normally on the phoneme task. These findings again demonstrate a double-dissociation between auditory comprehension and performance on a sub-lexical task.

Finally, more recent work on phonemic perception as studied using sub-lexical speech tasks provides additional relevant evidence. Caplan, Gow, and Makris (1995) studied a group of ten unselected aphasics with "acoustic–phonetic" deficits as defined by their performance on an extensive battery involving a range of phonemic contrasts, with both natural and synthetic stimuli, and in both discrimination and identification paradigms. Three of the ten patients had a clinical diagnosis of Broca's aphasia and one had a diagnosis of conduction aphasia; both of these syndromes, by definition, are characterized by good auditory comprehension, showing again that impaired performance on phoneme discrimination and identification is not predictive of poor auditory comprehension. As with the Basso study, it is not the case that the Broca's aphasics are only mildly impaired on sub-lexical tasks: one of the Broca's patients, R.Wi., had the worst score on the phoneme identification task, and the second worse composite score (discrimination + identification) in the sample. R.Wi.'s lesion involved left frontal cortex (Broca's area plus surrounding regions), and spared the temporal lobe, showing that poor sub-lexical task performance can occur with lesions which fully spare auditory cortex (case A.P. from that sample is another clear example of this effect).

It might be argued that patients with poor syllable discrimination and identification abilities can nonetheless comprehend words because contextual information constrains the construction of phonemic representations. This view would predict that if contextual cues were removed, as for example in single word-to-picture comprehension tasks with phonemic foils, auditory comprehension performance would fall off dramatically. This prediction has not been borne out however (see the experiment by Miceli et al., 1980 described above, and Section 4.1 below).

3.2. Evidence from functional neuroimaging

The neuroimaging literature also contains examples of dissociations between tasks involving conscious attention to speech segments vs. those that do not. Passive listening to speech of various sorts (syllables, words, sentences) reliably activates portions of the STG bilaterally (Norris & Wise, 2000). Activation of other sites such as Broca's area, while found in many studies involving passive listening, tends to be less robust both in spatial distribution and amplitude of the response (e.g. Binder et al., 2000; Schlosser, Aoyagi, Fulbright, Gore, & McCarthy, 1998), appears in fewer individual subjects (e.g. Nakai et al., 1999), or is not reported at all (e.g. Wise et al., 1991). Broca's area is strongly and reliably activated, however, when subjects are asked to perform various sub-lexical tasks involving auditorily presented speech (Burton, Small, & Blumstein, 2000; Zatorre, Evans, Meyer, & Gjedde, 1992; Zatorre, Meyer, Gjedde, & Evans, 1996). Recent work has suggested that this anterior activation is driven primarily by processes involved in explicit segmentation (Burton et al., 2000). These studies are considered in more detail in Section 5.2.

3.3. Section summary

In hindsight, it should come as no surprise that performance on sub-lexical speech perception tasks dissociates from performance on auditory comprehension tasks: in normal conversational speech, listeners have no explicit awareness of the phonemic structure of the input, only the semantic content. Of course, sub-lexical information contained in heard speech can be accessed explicitly, but only with prior instruction to attend to such information in an utterance, and only apparently if the listener is literate (Morais, Bertelson, Cary, & Alegria, 1986). Clearly, then, explicit access to sub-lexical structure entails cognitive mechanisms that are not typically engaged in listening to speech for comprehension. We hypothesize that there is overlap in the neural systems supporting these two classes of tasks up to the level of the STG – all speech tasks activate this region – with subsequent processing relying predominantly on non-overlapping systems (Fig. 2). In other words, accurate performance on sub-lexical speech tasks requires the integrity of speech processing systems in the STG as well as frontal circuits. The fact that deficits on sub-lexical tasks can occur in a variety of clinical aphasic subtypes (e.g. Broca's, conduction and Wernicke's aphasia; see cases in Caplan et al., 1995), with their associated lesion distributions involving frontal and posterior regions, is consistent with this claim. Because speech processing systems in the STG are also part of the system supporting auditory comprehension, this hypothesis predicts that deficits on sub-lexical speech tasks should be partially predictive of auditory comprehension deficits *for patients whose sub-lexical deficits are attributed to STG lesions*.

4. The ventral stream

The ventral stream, which one can broadly conceptualize as an auditory 'what' system, deals with the conversion of sensory information into a format suitable for linguistic
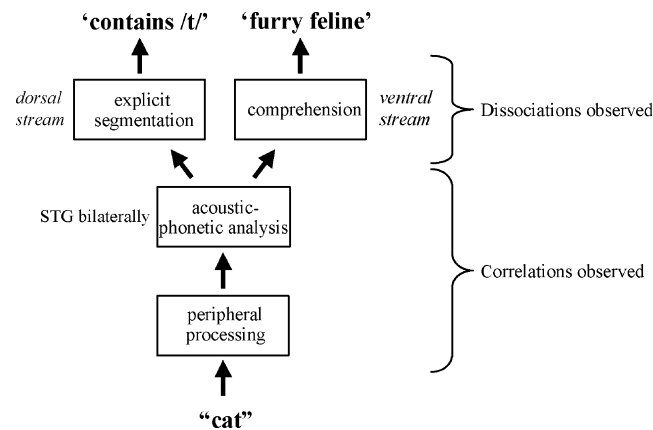
Fig. 2. Schematic of the relation between systems supporting sub-lexical segmentation ability and auditory comprehension ability. Observed dissociations between these abilities arise when damage or functional imaging studies affect portions of the system after the divergence point, whereas we would predict some degree of correlation between these abilities if damage or functional imaging targets earlier shared components of the system.

computation (in the case of speech input). As such, this pathway deals with (probably multiple levels of) acoustic–phonetic processing, the interface of acoustic–phonetic representations with lexical representations, and the interface of the lexical items or roots with the computational system responsible for syntactic and morphological operations. In summary, this pathway mediates comprehension broadly construed, i.e. from sound to meaning. (Note that some stages of this process, e.g. "acoustic–phonetic" processing, are likely common to both the ventral and dorsal streams.)

When the neural organization of speech perception (or acoustic–phonetic processing, we will use these terms interchangeably) is examined from the perspective of auditory comprehension tasks, the picture that emerges is one in which acoustic–phonetic processing is carried out in the STG bilaterally (although asymmetrically) and then interfaces with conceptual systems via a left-dominant network in posterior inferior temporal regions (e.g. MTG, ITG, and perhaps extending to regions around the temporal–parietal–occipital boundary). The arguments supporting this claim follow in Sections 4.1 and 4.2. Sentence-level processes may additionally involve anterior temporal regions (see Section 4.4).

4.1. Bilateral organization at early stages of the cortical processing hierarchy: the construction of sound-based representations

The idea of bilateral organization of speech perception is not new. An interesting historical note is that while Wernicke's area is classically associated with the left STG, Wernicke himself indicates that both hemispheres can represent "sound images" of speech. According to Wernicke (1874/1969), the left STG becomes dominant

for language processes by virtue of its connection with the left-lateralized motor speech area:

> All the available information supports the view that the sensory nerves, most of which are bilateral in function, deliver memory images to identical points in both hemispheres. The locus of sound images must thus be present on the right as well as the left… But only the left sound center is effectively connected with the motor speech center, and thus probably only the left sound center has established well-worn connections with the conceptual regions. But the right sound center can completely replace the left one very quickly… (f.n. #15, p. 97)

Data collected over the last century strongly support Wernicke's contention that sound images (what we have termed acoustic–phonetic speech codes) are represented bilaterally in auditory cortical fields, and that these representations in either hemisphere are sufficient to support access to the mental lexicon. The data come from several sources (for a more comprehensive treatment see Hickok, 2000; Hickok & Poeppel, 2000; Norris & Wise, 2000; Poeppel, 2001). First, while lesions in left posterior temporal cortex frequently produce auditory comprehension impairments, an examination of the nature of the comprehension errors reveals (i) relatively few phoneme-based errors (patients select a picture of a pear when presented with the word "bear" about 12–20% of the time in the most severe cases), and (ii) relatively more semantic errors than phonemic errors (e.g. presented with the word "bear", patients are more likely to select a picture of a moose than a picture of a pear) (Barde, Baynes, Gage, & Hickok, 2000; Blumstein, Baker, & Goodglass, 1977; Gainotti, Micelli, Silveri, & Villa, 1982). These findings show that unilateral lesions do not regularly cause profound speech perception deficits (in auditory comprehension tasks) as one would predict if acoustic–phonetic processing systems were exclusively left lateralized.[5] Good auditory comprehension abilities (at least at the lexical level) in the right hemisphere in split brain subjects (Barde et al., 2000; Zaidel, 1985) and patients undergoing Wada procedures (McGlone, 1984) is consistent with these findings.

The bilateral hypothesis predicts that profound speech perception deficits (functional deafness for speech) will be found in association with bilateral lesions. The second source of data, therefore, comes from cases of word deafness, a syndrome characterized by the profound impairment in speech perception not found among the classical aphasias. The most common lesion pattern is bilateral damage to the superior temporal lobe, consistent with the bilateral hypothesis (Buchman, Garron, Trost-Cardamone, Wichter, & Schwartz, 1986; Poeppel, 2001). A number of cases have been

---

[5] Again, it is important not to confuse "acoustic–phonetic" deficits in the context of auditory comprehension tasks with "acoustic–phonetic" deficits in the context of sub-lexical tasks. What we are claiming in this section is that unilateral lesions that produce auditory comprehension deficits, as in Wernicke's aphasia, do not produce profound disruptions of the ability to construct acoustic–phonetic representations of speech, but rather only impair this ability mildly. This is shown by the fact that relatively few phonemic errors are made by such patients. Whether or not these representations, once constructed, can be explicitly segmented in a sub-lexical task is a separate issue that could and perhaps should be studied explicitly. For example, we know that at least some Wernicke's patients have trouble with sub-lexical tasks. Within the model, this could arise either because partial damage to acoustic–phonetic processing systems in the STG causes deficits on sub-lexical tasks, and/or because there may also be damage involving the dorsal stream auditory–motor interface system. See Fig. 2.

reported with unilateral lesions in the left hemisphere, but these cases have to be considered in proportion to the number of cases of unilateral left hemisphere lesions that do *not* cause word deafness. The fact that only 17 clear unilateral cases of word deafness have appeared in the literature (Poeppel, 2001) in spite of the very common occurrence of left unilateral lesions suggests that this pattern is exceedingly rare and may, as Goodglass has suggested (Goodglass, 1993, p. 125), represent anomalous cases.[6]

A third line of evidence supporting bilateral organization of speech perception comes from neuroimaging. Physiological recordings of normal subjects listening to speech stimuli uniformly indicate bilateral activation in the STG (see review by Norris & Wise, 2000). In many discussions of this pattern, authors interpret the left activation as "phonemic" and the right as "acoustic", but as we have pointed out above, the neuropsychological data do not support this interpretation.

Recent work has provided some clues concerning the internal organization of cortical auditory systems supporting speech perception (e.g. Binder et al., 2000; Scott, Blank, Rosen, & Wise, 2000). The earliest levels in the cortical auditory processing hierarchy correspond anatomically to portions of Heschl's gyrus, which responds well even to relatively "simple" auditory stimuli such as unmodulated broad spectrum noise. The next hierarchical level involves the supratemporal plane, both anterior and posterior to Heschl's gyrus. These regions respond more vigorously to time-structured signals than to unstructured stimuli (noise). Although speech activates these regions robustly, less complex spectro-temporal signals, such as sequences of pure tones, also produce strong responses. Finally, ventro-lateral portions of the STG, extending into both anterior and posterior portions of the STS, appear to respond best to complex spectro-temporal signals such as speech. Lexical or semantic manipulations do not modulate the response in these later regions, nor does the intelligibility of phonemic segments (words, pseudowords, and time-reversed speech activate these regions equally well) (Binder et al., 2000; Scott et al., 2000). These STS regions also show sustained activation during the active maintenance (silent rehearsal) of phonological information (Buchsbaum, Hickok, & Humphries, 2001; Hickok et al., 2003). All of these observations suggest that cortex in ventro-lateral portions of the STG, including the STS (stippled portion in Fig. 1B), comprises advanced stages in the auditory processing hierarchy which are critical to phoneme-level processing (although not necessarily exclusive to it). The point in this processing hierarchy at which our proposed dorsal- and ventral-streams split off remains to be investigated intensively, but it now seems

---

[6] Although it is clear that bilateral STG damage is the more common etiology of word deafness which supports the bilateral hypothesis, it may also be relevant that 16 of the 17 unilateral cases involved a relatively small white matter lesion in the left temporal lobe. Perhaps an exquisitely placed lesion within the left hemisphere (thus explaining its rarity) can produce word deafness by some mechanism possibly involving an interruption of auditory radiations in the left hemisphere and callosal projections from the right hemisphere. This would not explain, however, why larger subcortical lesions in the left hemisphere do not produce word deafness, nor why the isolated right hemisphere has good auditory comprehension, unless we assume some inhibitory influence of intact left cortical areas on the speech perception abilities of the right hemisphere, which would again implicate both hemispheres. This is an important issue that deserves more attention, but which does not detract from the main line of argumentation, namely that the majority of word deaf cases involve bilateral STG lesions as predicted.

likely that they share auditory processing resources up to the most advanced level in the auditory cortical hierarchy described above.

4.2. Asymmetries embedded in bilateral organization

The observation documented in Section 4.1, that the speech code is mediated bilaterally in the STG, i.e. that the STG in both hemispheres is capable of extracting speech-relevant information from the auditory stream sufficiently well to access lexical representations, does not imply that the left and right STGs perform exactly the same computation on incoming speech information. Neuropsychological (Robin, Tranel, & Damasio, 1990; Zaidel, 1985), electrophysiological (Gage, Poeppel, Roberts, & Hickok, 1998; Shtyrov et al., 1998), and imaging data (Belin et al., 1998) show that the input signal is analyzed bilaterally – but not identically.

There are several proposals in the literature concerning how aspects of auditory computation are lateralized. First, a recent model by Ivry and colleagues (Ivry & Robertson, 1998), *the double-filtering-by-frequency theory*, posits that (auditory or visual) input signals are initially characterized by their spectral content. In the context of a perceptual task, an attentional filter identifies the relevant spectral range for the given perceptual requirements (filter 1). Subsequent to the identification of a (task-specific) spectral center point, the representations are asymmetrically elaborated in the two hemispheres; the spectral point defined by the attentional system acts as a frequency center point around which the information is filtered. The high-pass data are passed to left hemisphere areas, low-pass data to the right (filter 2). One advantage of this model is that it does not posit absolute frequency ranges that are distributed to the two hemispheres but rather predicts that relatively higher vs. lower frequency portions of a stimulus are distributed across the hemispheres.

A second model, argued for by Zatorre and colleagues (Zatorre, 1997; Zatorre, Belin, & Penhune, 2002), suggests that left superior temporal cortical areas are specialized for the *analysis of temporal changes* of signals and right mechanisms are for *spectral analysis*. This proposal has the virtue that it accounts well for many neuropsychological and imaging data that have shown, for example, that right temporal cortex lesions are associated with poor performance on pitch change tasks, prosodic phenomena, and other tasks that require fine spectral analysis (Zatorre et al., 2002). In contrast, left lesions show poor performance on tasks that require the analysis of rapidly changing information (see Nicholls, 1996 for a review).

A third proposal, closely related to the previous two, is the asymmetric sampling in time (AST) model (Poeppel, 2001, 2003). That model proposes that the speech signal is asymmetrically analyzed in the time domain, with left-hemisphere mechanisms preferentially extracting information over shorter (25–50 ms) temporal integration windows and right mechanisms over longer (150–250 ms) windows. The initial spectro-temporal representation of an auditory input signal is bilaterally symmetric; subsequently the same signal is analyzed on two different time scales. These time scales are commensurate with the requirements to process small frequency changes (right) and rapid temporal changes (left) at the same time. The AST proposal has the disadvantage that it

commits to specific temporal integration windows and is thus committed to (relatively broad) absolute hemispheric asymmetries.

What all models must account for is (a) that speech signals are processed bilaterally, (b) that the output of these processes are representations that must be able to interface with (at least) lexical representations, and (c) that the left and right hemispheres are differentially sensitive to aspects of speech signals, for example, rapid spectral changes such as in formants and slow changes such as in small pitch shifts in prosody. It is, of course, an empirical question as to which of these models, or some other model, captures the relevant phenomena more accurately.

4.3. The sound–meaning interface

The existence of some mechanism for interfacing sound and meaning is uncontroversial. This process is likely a multistage operation: for example, Stevens (2002) postulates multiple steps just in the derivation from the acoustic speech signal to a representation of a sequence of segments. These computational steps (or steps similar to them) are subsumed in our component labeled "acoustic–phonetic speech codes" and have a functional neuroanatomical organization that is yet to be determined. The mapping from a sequence of segments onto a conceptual–semantic representation is also often believed to involve at least one additional step, such as some form of word-level representation (e.g. Forster, 1976; Marslen-Wilson, 1987; McClelland & Elman, 1986; Shelton & Caramazza, 1999).[7] There is debate (at least on the production side) over the question of whether these word-level representations are simply phonological word-forms, "lexemes" (Caramazza, 1997), or whether in addition a more abstract lexical representation, "lemma", is maintained (Levelt, Roelofs, & Meyer, 1999). For the present purposes, these distinctions, should they exist, would constitute subdivisions of our "auditory–conceptual interface" system. Thus, we are taking an agnostic stand on the computational details of this interface system. Our claim is simply that there exists a cortical network which performs a mapping between (or binds) acoustic–phonetic representations on the one hand, and conceptual–semantic representations on the other.

Can this system be localized in the brain? We have proposed that this sound–meaning interface system can be coarsely localized to cortex in the temporal–parietal–occipital junction, predominantly on the left (Hickok & Poeppel, 2000). The evidence for this claim was based on transcortical sensory aphasia, a syndrome in which auditory comprehension is impaired, but syntactic and phonological abilities appear to be spared, as indicated by preserved repetition ability (Kertesz, Sheppar, & MacKenzie, 1982). Semantic paraphasias dominate the production errors in this syndrome suggesting a deficit not only in comprehension but also in production (Damasio, 1992; Kertesz et al., 1982). The lesions associated with transcortical sensory aphasia are varied, but generally occur in regions posterior and/or inferior to the Sylvian fissure in the left hemisphere, that is, various sites in and around the T-P-O junction, particularly pITL structures (Damasio, 1991; Kertesz et al., 1982). This clinical–anatomic pattern is consistent with the proposal

---

[7] Models of speech production also posit one (or more) word-level representations (Caramazza, 1997; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Levelt, 1989).

that left posterior extra-Sylvian regions, particularly inferior aspects (e.g. pITL), comprise a network involved in mapping sound onto meaning and vice versa.[8]

Is there any evidence beyond transcortical sensory aphasia implicating left pITL structures in sound–meaning mapping? Several lines of evidence are supportive of the hypothesis. One line of evidence comes from work on the "basal temporal language area" which was discovered in electrical stimulation studies of speech/language abilities (Lüders et al., 1991). Stimulation of the basal portion of the dominant temporal lobe (corresponding to the fusiform/inferior temporal gyri) produced speech interruption during reading in eight of 22 cases studied. (Stimulation of classical language areas produced higher rates of speech interruption: Broca's area = 15/22, Wernicke's area = 14/22.) Three of the patients who showed deficits in the reading task were studied further, revealing that basal temporal stimulation produced deficits on both production tasks (e.g. object naming) and comprehension tasks (Token Test). The naming deficit was not due to visual agnosia as all three patients could subsequently recall and name the object presented during stimulation. Stimulation of the non-dominant hemisphere did not cause speech/language deficits. Repetition was also affected in the two patients tested for this ability. The existence of a ventral temporal language area region which participates both in comprehension and production is certainly consistent with our proposed ventral processing stream. However, recent findings suggest that the language disruptions elicited by stimulation of the basal temporal language area may be caused by remote after-discharges in the left posterior STG, rather than by stimulation of the basal region itself (Ishitobi et al., 2000), so it is unclear what role this region plays in language processing. Nonetheless, this line of research in general highlights possible links between superior temporal regions and more ventral temporal structures which may play some role in language processing.

Another line of evidence comes from imaging studies of "semantic processing" (typically lexical semantics) which generally implicate inferior posterior temporal regions and posterior parietal cortex (Binder et al., 1997).[9] This distribution of activation corresponds quite well to the distribution of lesions associated with transcortical sensory aphasia (TSA) which lends support to the claim of meaning-based integration networks in posterior ITL (again perhaps extending to regions around the temporal–parietal–occipital junction). But it is difficult to know exactly how the tasks used in such imaging experiments relate to the proposed sound–meaning interface system, so caution is warranted in interpreting these findings. See Binder et al. (2000) for further discussion and for limited evidence suggesting that pITL regions may activate more during listening to words compared with non-words, which is consistent with a functional role in mapping sound onto meaning.

A third source of evidence for the localization of a sound–meaning interface system comes from Wernicke's aphasia. It has been shown that single-word comprehension

---

[8] Auditory comprehension deficits in TSA often resolve over weeks or months post-stroke suggesting that lesions to left pITL structures do not permanently impair sound–meaning mappings. There are two explanations of this observation which can be put forth within the framework developed here. One is that the wide distribution of sound–meaning mapping systems within left pITL makes it unlikely that a lesion will compromise large enough portions of this system to induce lasting deficits. A second explanation is that the sound–meaning mapping system is bilaterally organized to some degree.

[9] Frontal regions are also typically activated. We will concentrate on the posterior distribution in this paper.

deficits in Wernicke's aphasia have a prominent semantic component suggesting a breakdown in the mapping between sound and meaning (Baker, Blumstein, & Goodglass, 1981). Since lesions restricted to the posterior STG do not lead to lasting comprehension deficits (Dronkers, Redfern, & Knight, 2000), it is likely that it is the extra-Sylvian inferior extension of the lesion in Wernicke's aphasia that is largely responsible for comprehension problems (Dronkers et al., 2000). This extra-Sylvian involvement in Wernicke's aphasia overlaps partially with the distribution of lesions in TSA (Kertesz et al., 1982), thus lending support to the view that these regions play an important role in mapping between sound and meaning.

A fourth line of evidence comes from neuropsychological studies which specifically target word-level semantic deficits. For example, Hart and Gordon (1990) correlated lesion location in a series of aphasic patients with single-word semantic deficits as measured by several tasks and argued for an association between left posterior–temporal lesions and single word semantic deficits. Similar evidence comes from semantic dementia (SD) in which patients suffer a progressive decline in the ability to comprehend or name items from common conceptual categories. Phonological and syntactic aspects of language ability are claimed to be relatively preserved, as is performance on episodic memory tasks, and other standard non-verbal neuropsychological measures (Garrard & Hodges, 2000). A recent review of 45 cases of SD (Garrard & Hodges, 2000) revealed atrophy predominantly in the left temporal lobe. Two regions appear to be particularly affected: the temporal pole, and infero-lateral regions. Although some authors have emphasized the role of the temporal pole in producing semantic deficits in SD (Mummery et al., 2000), we argue that it is the posterior temporal involvement which is the primary contributor to the symptom complex in this syndrome. The logic of the argument is first that data from SD alone are ambiguous with respect to whether it is the temporal pole or more posterior structures (or both) which are responsible for the language deficit, and second that data from other sources suggest a minimal role for the anterior temporal lobe in lexical–semantic processing, yet an important role for posterior inferior temporal structures in lexical–semantic processing (see immediately below). Therefore, the weight of the data implicates posterior structures.

Let's consider this argument in more detail. While the temporal pole does appear to be the site of maximal atrophy in SD, it does not follow that this region is necessarily the site which causes the behavioral deficit: it could be that the extensive temporal pole atrophy has no behavioral consequences, and that the deficits arise from relatively milder atrophic changes which occur in areas critical for language processing. Since atrophic changes have been observed, not only in the temporal pole but also in other inferior temporal regions (as well as frontal regions), we cannot conclude that the temporal pole is the critical site.[10] Evidence from physiological measurements further implicates posterior structures in SD. For example, hypo-perfusion of left posterior temporal areas as measured by SPECT or

---

[10] One study (Mummery et al., 2000) correlated semantic deficits with the degree of temporal pole atrophy and found a significant correlation. But it is difficult to interpret this study because posterior areas were not studied, leaving open the possibility that atrophy in posterior temporal cortex would also show a strong relation to semantic deficits. In other words, it is possible that semantic deficits were correlated with atrophy in anterior temporal regions only because atrophy in anterior temporal regions is correlated with atrophy in language-critical posterior temporal regions.

PET is not uncommon in SD (Garrard & Hodges, 2000), and posterior inferior temporal areas (e.g. Brodmann area 37), but not anterior temporal areas, show less activation in SD patients performing a semantic judgment task than do normal controls (Mummery et al., 1999). Given that posterior as well as anterior areas are affected in SD, the data are ambiguous with respect to the anatomical source of the deficit. If we look to evidence from other sources concerning the relative contributions of anterior vs. posterior temporal areas in lexical–semantic abilities, a very clear picture emerges, one in which posterior structures play a dominant role. Left anterior temporal lobectomies have minimal effects on language abilities (Saykin et al., 1995), and certainly do not produce symptoms of SD. Likewise, language disorders that have been associated with left anterior temporal lobe strokes are relatively mild including naming deficits for unique entities (Damasio et al., this volume) and the comprehension of syntactically complex sentences (Dronkers et al., this volume) – certainly not the substantial lexical–semantic processing deficits seen in SD. However, lexical–semantic processing deficits similar to those reported in SD are associated with posterior temporal lobe strokes, as reviewed above. Taken together, these observations (including those from SD) are consistent with our claim of a sound–meaning mapping system in posterior ITL.

Finally, Indefrey and Levelt (this volume), in their meta-analysis of a large number of functional imaging studies, have identified the middle portion of the MTG as a site which plays a role in "conceptually-driven lexical retrieval" during speech production; this region was also shown to be consistently active during speech perception in their analysis. This stage in processing is compatible with our sound–meaning interface. While our localization of this network extends more posteriorly due to our consideration of neuropsychological data (e.g. TSA), Indefrey and Levelt's localization does overlap ours in the lateral MTG (~middle third).

The above discussion makes a reasonably strong case for the involvement of left posterior inferior temporal regions in non-phonemic aspects of language comprehension and production. We have interpreted this set of results as indicating that this region is important in relating sound to meaning and vice versa, and therefore may correspond to something like a lemma level of representation. Of course, it is quite difficult to determine whether word-level deficits, or activations produced by lexical–semantic tasks, actually reflect the computational operations involved in mapping between sound and meaning. The ability to perform lexical–semantic tasks surely relies on a complicated neural network, components of which may or may not be speech-specific, and which may or may not be task-dependent. By using data from lexical–semantic tasks generally to localize systems which map between sound and meaning, we are very likely oversimplifying the picture. It is an empirical question whether this oversimplification will prove to be a helpful generalization or not. What we do know is that (i) several lines of evidence point to lexical–semantic processes of a variety of sorts being linked fairly consistently to pITL (unlike, for example, anterior temporal regions), and (ii) pITL is implicated not only in meta-semantic tasks (e.g. category membership judgments), but also in more simple comprehension and naming tasks (indeed both comprehension and naming; see Indefrey and Levelt, this volume). For this reason, we take the position that pITL plays a central role in lexical–semantic processing, and although we don't claim to understand precisely what it is doing, we hypothesize generally that it maps sound onto meaning and vice versa

(again we are not necessarily claiming exclusivity of function in this region). Clearly, much work remains to be done.

4.4. The interface with grammatical processing

We have discussed the interface between sound-based representations and meaning-based representations associated with single words, and have suggested that this sound–meaning interface can be coarsely localized to left inferior temporal cortex. How does grammatical processing fit into this network? The answer to this question remains unclear, but some clues are beginning to emerge. For example, several recent studies have implicated the left anterior temporal regions (e.g. anterior STS) in sentence-level processing. Anterior temporal regions activate in response to listening to meaningful or jabberwocky sentences, but not to (or less well to) word lists, foreign sentences, backwards sentences, or meaningful environmental sound sequences (Friederici, Meyer, & von Cramon, 2000; Humphries, Willard, Buchsbaum, & Hickok, 2001; Mazoyer et al., 1993; Schlosser et al., 1998). Morphosyntactic comprehension deficits have also been linked to pathology in left anterior temporal regions (Dronkers, 1994; Grossman et al., 1998; and see Dronkers, this volume).

Together, these data suggest a possible role for anterior temporal cortex in aspects of grammatical processing, but it is also clear that this is not the only region involved. Left anterior temporal lobectomies do not substantially interfere with language comprehension (Saykin et al., 1995), patients with left aTL lesions and morphosyntactic comprehension deficits typically have lesions that involve additional regions (see Dronkers et al., this volume), and many of the imaging studies cited above report sentence > non-sentence activation in other areas such as posterior STS, MTG, and Broca's area.

A recent MEG study by Friederici et al. (Friederici, Wang, Hermann, Maess, & Oertel, 2000) exemplifies how anterior temporal and inferior frontal cortex might be coordinated when a grammatical representation is constructed. These authors recorded electrophysiological signals from subjects engaged in a psycholinguistic paradigm in which the component associated with supposed early structure building was elicited (ELAN, early left anterior negativity). This early component is argued to reflect very rapid, automatic structure building. When the sources underlying the component were modeled, it was observed that two sources best accounted for the electrophysiological data, one source being in left inferior frontal gyrus (Broca's area) and one source in the superior anterior temporal lobe.

Although speculative, one possibility is that aTL serves as an interface between posterior lexical–semantic systems and frontal systems involved in structuring information over time periods that extend beyond the duration of a sensory trace (Fuster, 1995). On this view, grammatical processing is not localized to any one place (which would be surprising given that grammatical processing is not monolithic but a complex amalgam of computations), but rather is instantiated in the collective activity of a large-scale distributed network involving frontal, anterior temporal, and posterior temporal systems (Caplan, Hildebrandt, & Makris, 1996), each of which nonetheless play different roles in the process.

5. The dorsal stream

Using the organization of the visual system as a guide, we have hypothesized the existence of a dorsal auditory stream which is critical for auditory–motor integration (Hickok & Poeppel, 2000). In this section we first outline current views on dorsal-stream sensory–motor integration networks in vision, and then specify the role that an *auditory–motor* integration system might play in speech/language. Finally, we turn to neural evidence relevant to mapping the spatial distribution of this network.

5.1. Sensory–motor integration in the parietal lobe

While dorsal stream processing in vision has traditionally been aligned with spatial "where" functions (Ungerleider & Mishkin, 1982), there is a growing literature which demonstrates the existence of visuo-motor integration systems in the dorsal stream (Andersen, 1997; Milner & Goodale, 1995; Rizzolatti et al., 1997). Milner and Goodale (1995) provide an extensive review of this literature, so we will only summarize some of the arguments. Single-unit recordings in the parietal lobe of primates have shown that many cells are sensitive not only to visual stimulation, but also to the monkey's action towards that visual stimulation. For example, a unit may respond not only when an object is presented, but also when the monkey reaches for that object in an appropriate way, even if the object is no longer in view (Murata, Gallese, Kaseda, & Sakata, 1996; Rizzolatti et al., 1997). Not surprisingly, these visuo-motor fields are densely connected with frontal lobe regions involved in controlling motor behavior (Rizzolatti et al., 1997). Additional evidence comes from the documentation of visual shape-sensitive units in parietal cortex (Murata et al., 1996; Sereno & Maunsell, 1998). Shape coding has traditionally been associated with ventral stream function, and presumably is irrelevant to a purely spatial "where" code. But shape information *is* critical in guiding appropriate grasping behavior, and so this finding supports a visuo-motor integration model of parietal lobe function. Finally, in humans, dissociations have been observed between the conscious perception of visual information (the ventral stream) and the ability to act on that information appropriately (the dorsal stream). For example, in optic ataxia, patients can judge the location and orientation of visual stimuli, but have substantial deficits in reaching for those same stimuli; parietal lesions are associated with optic ataxia (Perenin & Vighetto, 1988). Conversely, Goodale and colleagues (Goodale, Milner, Jakobson, & Carey, 1991) report a case of visual agnosia in which perceptual judgments of object orientation and shape were severely impaired, yet reaching behavior towards those same stimuli showed normal shape- and orientation-dependent anticipatory movements. Similar dissociations have been reported in the context of visual illusions in normal subjects (Aglioti, DeSouza, & Goodale, 1995), although not without controversy (Carey, 2001).

5.2. Sensory–motor integration in speech

Does the concept of a sensory–motor integration network make sense in the context of speech? The answer is 'yes', such a network must exist (Doupe & Kuhl, 1999). The primary argument is developmental: for the child to learn to articulate the speech sounds in

his or her linguistic environment, there must be a mechanism by which (i) sensory representations of speech uttered by others can be stored, (ii) the child's articulatory attempts can be compared against these stored representations, and (iii) the degree of mismatch revealed by this comparison can be used to shape future articulatory attempts. Although such a network obviously assumes less importance in adult speakers, there is evidence – from articulatory decline following late-onset deafness (Waldstein, 1989), from the effects of delayed auditory feedback on speech articulation (Yates, 1963), and from altered speech feedback experiments (Houde & Jordan, 1998) – that it continues to operate throughout life. Further, the fact that it is possible to repeat pseudowords accurately demonstrates that this network involves interactions between auditory and motor speech systems at a fairly low level, that is, without extensive mediation via conceptual systems.

We have suggested that verbal working memory relies on this auditory–motor integration network (Aboitiz & García, 1997; Hickok & Poeppel, 2000). Indeed, verbal working memory (and perhaps working memory in general) can be viewed as a form of sensory–motor integration (Wilson, 2001). For example, in Baddeley's model (Baddeley, 1992), the phonological loop is essentially a mechanism for using motor systems (articulatory rehearsal) to keep sensory-based representations (the phonological store) active. In our model, the phonological store overlaps with the STG systems supporting sound-based representations of speech, and the articulatory rehearsal component maps onto frontal systems supporting articulatory-based representations of speech. Verbal working memory in our model differs from Baddeley's in that our model postulates an explicit network which translates between the articulatory and storage components (Hickok & Poeppel, 2000).

On the assumption that verbal working memory is a special case of auditory–motor integration, we can use data concerning the localization of working memory as a means to identify the neural systems supporting auditory–motor integration. There is a body of work, primarily from functional imaging, which indicates that the articulatory rehearsal component involves left frontal cortices, notably portions of Broca's area and a more dorsal pre-motor region (Awh et al., 1996) (depicted in Fig. 1). The neural basis of the phonological store has proved to be more difficult to work out, but several reports have implicated left parietal cortex (Jonides et al., 1998; Paulesu, Frith, & Frackowiak, 1993). Our interpretation of the posterior activation is that it reflects the operation, not of the storage component per se, but rather of the auditory–motor integration network (Hickok & Poeppel, 2000). The storage component, in our framework, should be located in auditory-responsive fields in the STG – a location which had not been associated with verbal working memory tasks. A recent fMRI study (Buchsbaum, Hickok, & Humphries, 2001), however, has provided strong support for our prediction. Subjects listened to three multi-syllabic pseudowords and then covertly rehearsed them for 27 seconds followed by a period of rest, and then another set of pseudowords, and so on. In each participant (analyzed separately), robust activation during the rehearsal phase was found in two posterior sites: a site at the boundary of the parietal and temporal lobes deep inside the Sylvian fissure (area Spt), and a more lateral and inferior site on the STG/STS (Fig. 3). This latter finding, of STG/STS activation, confirms our prediction of STG involvement in verbal working memory.
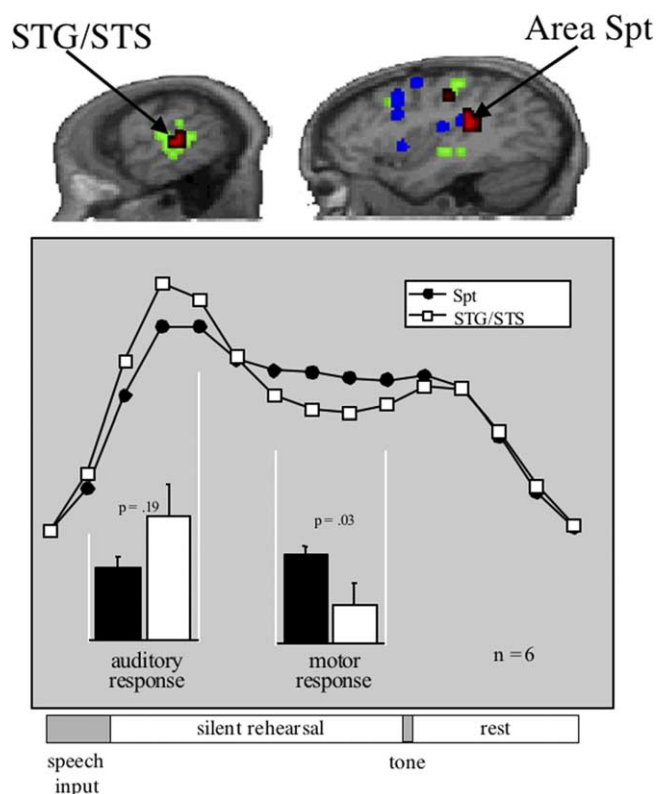
Fig. 3. (Top) Left sagittal views of a representative subject from Buchsbaum, Hickok, and Humphries (2001) illustrating two sites with auditory–motor response properties (red). Green: pixels that responded only to the auditory stimulus; blue: pixels that responded only during rehearsal. See text for details. (Bottom) Time course of activation in the two temporal lobe auditory–motor sites (Spt and STG/STS) averaged across six subjects.

Several additional, relevant observations emerged from this study. First, the precise location of both of these activations varied substantially from subject to subject, not so much in terms of their location relative to prominent anatomical landmarks, but in terms of their absolute locations in stereotaxic space. In fact, when the data were group averaged, both activation loci diminished in strength and spatial distribution. This is likely due to the high degree of individual anatomical variability in posterior Sylvian regions (Ide, Rodriguez, Zaidel, & Aboitiz, 1996), and probably explains why previous studies had not made similar observations. Second, both of these activation sites responded both during the auditory and motor (rehearsal) phases of the trial (Fig. 3). This sensory–motor response pattern is consistent with similar sensory–motor responses in single units located both within parietal lobe visuo-motor integration networks and within sensory cortex (e.g. so called "memory cells" in visual cortex) (Fuster, 1995). Third, although both sites responded to both the auditory and motor phases of the trial, the subject-averaged response pattern of area Spt differed significantly from the more lateral STG/STS site: Spt responded slightly less well to the auditory phase, but more vigorously to the motor phase, than did the STG/STS site.

This is consistent with the lateral site being more sensory-weighted than Spt. Finally, two frontal sites responded both to the auditory and motor phase of the trial, area 44 and a more superior pre-motor site (consistent with previous work, see above). Of particular interest was that the activation pattern in Spt and area 44 were the most tightly correlated of all four sites. This is particularly relevant given that the cytoarchitectonics of area Tpt (a region that subsumes the functionally identified Spt) and area 44 are virtually indistinguishable, suggesting a tight functional relation (Galaburda, 1982).[11]

Area Spt appears not to be speech-specific: in another fMRI experiment we replicated the Buchsbaum et al. result – sensory–motor response in area Spt – using jabberwocky speech stimuli, and then showed that this region responded equally well when participants listened to, and then silently "rehearsed" (hummed) short novel piano melodies (Hickok et al., 2003). The fact that Spt responds well to melodic stimuli argues against the view that activation in this region can be attributed to a specifically phonological store.[12]

Based on all of these findings, our hypothesis is that area Spt is a critical interface between sensory- and motor-based representations of speech (as well as for other functions).[13] This system can operate bi-directionally. Auditory-to-motor mappings would be used, for example, in verbatim repetition of heard speech (Buchsbaum, Humphries, & Hickok, 2001; Hickok et al., 2003; Wise et al., 2001) and in spontaneous speech production,[14] and motor-to-auditory mappings would be used in working memory tasks in which frontal circuits controlling articulatory rehearsal can be used to activate auditory representations of speech via area Spt (Hickok et al., 2003). Again, this notion is consistent with the idea from other domains that, in addition to frontal regions, both sensory representations (Fuster, 1995) and sensory–motor interface systems (Gallese, Fadiga, Fogassi, Luppino, & Murata, 1997) take part in working memory circuits.

The ability to perform sub-lexical speech tasks seems to rely on a neural network that is virtually identical to that involved in verbal working memory tasks. For example, data from both lesion (Caplan et al., 1995) and imaging studies (Burton et al., 2000; Zatorre et al., 1992) have implicated Broca's area, left inferior parietal, and left STG in "phonological"

---

[11] Area Tpt "…exhibits a degree of specialization like that of Area 44 in Broca's region. It contains prominent pyramids in layer IIIc and a broad lamina IV… Thus 44 and Tpt are equivalent transitional areas between the paramotor and the generalized cortices of the prefrontal area, and between parakoniocortex and the temporoparietal occipital junction areas respectively… the intimate relationship and similar evolutionary status of Areas 44 and Tpt allows for a certain functional overlap. In fact, architectonic similarities between anterior and posterior language areas, and the overlap in their connectional organization makes it a somewhat surprising finding that lesions in either region produce such different aphasic syndromes." (Galaburda, 1982)

[12] It has been suggested that the "phonological store" is specialized for phonological information, and does not code auditory information more generally (Salamé & Baddeley, 1982). However, sequences of pitch varying tonal stimuli have also been shown to interfere with short-term memory for speech stimuli arguing against this view (Jones & Macken, 1996), and consistent with the results of Hickok et al. (2003).

[13] Critical involvement in one function (sensory–motor interface for speech) does not preclude involvement in another function (e.g. sensory–motor interface for music-related abilities). For example, the vocal tract is critical, even specifically adapted for speech, yet it is also critically involved in and even specifically adapted for a completely unrelated function: digestion.

[14] Its role in speech production appears to be dependent on phonological load. For example, conduction aphasics, who typically have lesions in this region, have more difficulty repeating or naming multi-syllabic words, and a fMRI study (Okada, Smith, Humphries & Hickok, 2003) suggests that Spt activates to a greater extent when subjects covertly name objects with multi- vs. mono-syllabic names.

ability as revealed by a variety of sub-lexical tasks. This spatial overlap in the systems involved in verbal working memory and sub-lexical phonological tasks raises the possibility that they both rely on computations carried out by our proposed temporal–parietal–frontal sensory–motor integration network. If true, it would explain the spatially disparate neuroanatomy associated with sub-lexical speech tasks in terms of the spatially distributed location of the different functional components of the network (sensory, motor, or sensory–motor transformation) which together support the ability to perform such tasks.

But while there is a conceptual (and now empirical) argument which links working memory and sensory–motor integration, the relation between sensory–motor integration and sub-lexical speech tasks is less obvious. Why, for example, would one need to recode acoustic input, say a pair of syllables, in an articulatory format in order to decide whether they are the same or different? A possible answer comes from a recent study by Burton and colleagues (Burton et al., 2000). Using fMRI, these investigators studied onset phoneme discrimination embedded in two types of syllables, those that had identical rhymes (e.g. *dip–tip*) and those that had differing rhymes (e.g. *dip–ten*). The relevant contrast is that same–different judgment on the onset phoneme in a rhyming pair can be carried out on a global sensory representation (any perceived difference is indicative of a "different" response), whereas for non-rhyming pairs one presumably needs to segment out the onset phoneme and make a same–different judgment on that segmented representation (i.e. the pairs always differ in the global sensory representation). They found activation only in the STG, bilaterally, with rhyming pairs, but found additional activation in inferior parietal and inferior frontal cortex for non-rhyming pairs. It seems that frontal systems are only required for sub-lexical tasks that require explicit segmentation (Burton et al., 2000; Zatorre et al., 1996). We hypothesize that the frontal and parietal involvement in phonemic segmentation is correlated with processing load (explicit segmentation is harder than sensory discrimination): as load increases, a purely sensory representation is no longer sufficient (or efficient) to carry out the task, and so articulatory recoding (i.e. working memory) is used strategically to complete the task successfully. It will be informative to compare directly in the same subjects activations associated with standard verbal working memory tasks and activations associated with phonemic segmentation tasks.

6. Perception–production overlap in posterior "sensory" cortex

A critical component of Wernicke's 1874 model was that auditory representations of speech played an important role in speech production; this is how speech production errors (paraphasias) were explained in aphasia caused by lesions to left auditory areas. Available evidence suggests he was correct (Buchsbaum, Humphries, & Hickok, 2001; Hickok, 2001; Hickok et al., 2000, 2003).

Some of the best evidence comes from conduction aphasia (Hickok, 2000). Such patients have two primary deficits, phonemic paraphasias in their output, and naming difficulty.[15]

---

[15] Repetition impairment has gained prominence in the diagnosis of conduction aphasia. But Wernicke initially identified the syndrome on the basis of the existence of paraphasic errors in spontaneous speech but with good auditory comprehension. Repetition deficit can be thought of as an instance of a more general phonologically-based production deficit in conduction aphasia.

The difficulty at the phonemic level in production has been observed in a variety of conditions including spontaneous speech, naming, reading aloud, and repetition, but is often most evident under conditions of high phonological load and low semantic constraint such as repeating multi-syllabic words (Goodglass, 1992). Indeed, some authors consider conduction aphasia to be "an aphasic disorder in which the ability to encode phonological information is selectively impaired" (Wilshire & McCarthy, 1996). Relevant to our purposes is the fact that conduction aphasia has been associated with damage to – or direct electrical stimulation of – the left posterior STG (Anderson et al., 1999; Damasio & Damasio, 1980), a region which has also been strongly implicated in speech *perception*, based on neuroimaging data. This suggests some degree of overlap in the systems supporting speech perception and speech production, particularly in the left hemisphere (Hickok, 2000).

Neuroimaging data provide further evidence of left STG involvement in speech production (and therefore overlap in perception production systems). Portions of the left STG activate in a variety of speech production tasks including repetition (Wise et al., 2001), object naming (Hickok et al., 2000; Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998), and word generation tasks (Wise et al., 1991) – this is true even when production is covert so there is no auditory input (Hickok et al., 2000). And the studies described above (Buchsbaum, Hickok, & Humphries, 2001; Hickok et al., 2003) provided direct evidence for overlap between speech perception and production, showing that there are subfields in the STG (including Spt) which respond both during the perception and production of speech.

The evidence cited above argues strongly for overlap in the neural systems supporting speech perception and speech production. The site of overlap appears to be in the posterior STG. The perception/production overlap is partial, however: there appear to be perceptual systems which are not critically involved in production (e.g. early cortical auditory areas, and perhaps the right STG), and there are production systems which are not critically involved in perception (e.g. frontal areas, the anterior insula) (Dronkers, 1996; Wise, Greene, Büchel, & Scott, 1999). This arrangement, with partially overlapping perception/production systems is consistent with previous claims of both overlap (Allport, 1984; Coleman, 1998; MacKay, 1987) and independence (Dell et al., 1997; Levelt et al., 1999) of speech input and output functions (Hickok, 2001).

The idea that there is a tight relation between speech perception and speech production is not new. Analysis-by-synthesis proposals, in general, and the motor theory of speech perception, in particular, have long argued for this position (Liberman & Mattingly, 1985). Our proposal builds on this work. The present model differs in that the mapping of sensory representations of speech onto motor representations may not be an automatic consequence of speech perception, and indeed are not necessary for auditory comprehension (Hickok & Poeppel, 2000); this view has recently been advanced also by researchers in the Haskins group (Benson et al., 2001). On our view, motor representations of speech *can* be activated and utilized strategically to assist in task performance.

7. Understanding aphasia

The framework outlined in this article and schematized in Fig. 1 is in part motivated by findings from the deficit-lesion literature and should account in natural ways for relevant aphasic syndromes. Here we summarize how the proposal provides a framework to discuss deficit-lesion data using four types of clinical deficits.

7.1. Word deafness

Word deafness is an auditory dysfunction in which the integrity of the speech perception system is severely compromised. Early stages of hearing (e.g. frequency discrimination) are intact, but patients are functionally deaf for speech. As noted in Section 4.1, the lesions associated with word deafness typically involve the STG bilaterally.

In our framework, word deafness would be caused by complete (i.e. bilateral) disruption of the system supporting acoustic–phonetic representations of speech. Because we assume that the left portion of this system also participates in phonological aspects of production, we predict that word deaf patients with lesions involving these left regions should have phonemic paraphasias in their speech output. Of the cases where production data are reported, greater than 70% of word deaf patients indeed have such production deficits (Buchman et al., 1986).[16]

7.2. Conduction aphasia

Conduction aphasia is characterized by good auditory comprehension, fluent speech production, relatively poor speech repetition, frequent phonemic errors in production, and naming difficulties (Damasio, 1992; Goodglass, 1992). Although the repetition disorder has gained clinical prominence in the diagnosis of conduction aphasia, phonemic errors are frequently observed not only in repetition tasks, but also in spontaneous speech, oral reading, and naming tasks (Goodglass, 1992). Lesions associated with conduction aphasia are distributed around the left posterior temporal–parietal boundary involving either the STG or the supramarginal gyrus, or both (Damasio & Damasio, 1980; Dronkers et al., 2000).

The classical analysis of conduction aphasia (Geschwind, 1965; Lichtheim, 1885; Wernicke, 1874/1969) holds that the disorder results from a disconnection of posterior and anterior language centers caused by a lesion involving the white matter pathways (typically assumed to be the arcuate fasciculus; Geschwind, 1965) which normally connect these two classic language regions. Anderson et al. (1999) recently reviewed

---

[16] Are the 30% of word deaf cases without production deficits problematic for our claim? Not necessarily. First, it seems likely that speech production is not something that is investigated intensively in word deaf patients. Perhaps production ability was noted only in a cursory fashion without investigating it thoroughly, e.g. with phonological complex items. Second, under our bilateral hypothesis, not all lesion patterns which might produce word deafness would be expected to also produce phonemic errors in production. Word deafness associated with bilateral STG lesions would produce production errors on our view. However, theoretically, it would also be possible to produce word deafness with a lesion to right STG, plus a subcortical lesion which interrupts the left auditory radiations. This would leave left STG intact, therefore preserving speech production abilities.

a range of data which call into question the validity of this claim. In particular, they point out that focal lesions to the arcuate fasciculus do not cause conduction aphasia (Tanabe et al., 1987), that conduction aphasia is typically associated with *cortical* lesions including cases which spare the arcuate fasciculus (Benson et al., 1973; Damasio, 1992; Damasio & Damasio, 1980; Green & Howes, 1978), and that *cortical* stimulation can produce conduction aphasia-like symptoms (Anderson et al., 1999; Shuren et al., 1995). Conduction aphasia, therefore, is better viewed as a syndrome which results from cortical dysfunction affecting the computational systems underlying aspects of phonemic processing, rather than as a white matter disconnection syndrome (Anderson et al., 1999; Hickok, 2000).

In the context of the present framework, conduction aphasia can be explained in terms of damage to the sound-based speech processing systems in left STG and/or to the temporal–parietal system (Spt) which interfaces these systems with motor–articulatory networks. Comprehension is relatively preserved because right hemisphere speech perception systems are intact, as are sound–meaning interface systems. Production (including repetition) contains frequent phonemic paraphasias because of the left lateralized involvement of the STG region in phonemic aspects of production (see Section 6). It is an open question whether damage to STG systems vs. the Spt interface will yield subtle differences in the deficit pattern. Perhaps there is a connection between (i) these two proposed processing systems, (ii) the two different forms of conduction aphasia which have been identified, reproduction conduction aphasia and repetition conduction aphasia (Shallice & Warrington, 1977; Wilshire & McCarthy, 1996), and (iii) the two lesion patterns (STG vs. supramarginal gyrus). Clearly further work is needed.

7.3. Transcortical sensory aphasia

Transcortical sensory aphasics have impaired comprehension, fluent but paraphasic speech production (mostly semantic errors), but with relatively preserved repetition (Damasio, 1992; Goodglass, 1993). The ability to repeat speech, and sometimes spontaneously correct grammatical errors, suggests that phonological and syntactic systems are largely preserved in TSA. Lesions associated with TSA are typically found in various locations around the posterior temporal lobe, particularly inferior areas (Kertesz et al., 1982); classical Wernicke's area is usually preserved (Damasio, 1991).

TSA can be understood as a deficit involving the mapping between sound and meaning – the sound–meaning interface system in the present framework. Comprehension is impaired because the mapping of auditory speech representations onto meaning is disrupted. Production contains semantic paraphasias because the reverse mapping is impaired. And repetition is spared because the auditory–motor circuit is intact. It is worth highlighting the complementary distribution of symptoms and lesions in conduction aphasia and TSA: conduction aphasia involves good comprehension, phonemic errors, poor repetition, and damage to left posterior Sylvian cortex; TSA involves poor comprehension, semantic errors, good repetition, and damage to left pITL.

7.4. Wernicke's aphasia

Wernicke's aphasia is classically considered to be a "primary" aphasia in that it is thought to result from damage to a single cortical system (e.g. auditory word representations in Wernicke's model). In our model, Wernicke's aphasia is a composite disorder involving damage to both auditory speech systems in left STG and sound–meaning mapping systems in left pITL. A close look at the symptom complex of the syndrome will reveal its composite nature: comprehension is impaired as in TSA, repetition is impaired as in conduction aphasia, and production contains both semantic and phonemic paraphasias as in TSA and conduction aphasia, respectively. Thus, the deficits in Wernicke's aphasia are essentially those of the negative symptoms of conduction aphasia and the negative symptoms of TSA combined (Hickok, 2000). The lesion pattern typical of Wernicke's aphasia also shows overlap between those associated with conduction and transcortical sensory aphasias, with damage usually involving large sectors of left posterior cortex including STG, MTG, and supramarginal and angular gyri (Damasio, 1991). One symptom associated with Wernicke's aphasia, paragrammatic production (Goodglass, 1993), has not been described in connection with either conduction aphasia or TSA. Perhaps this symptom can be related to the frequent involvement of MTG in Wernicke's aphasia (Dronkers et al., 2000).

8. Summary and conclusions

The framework for the functional anatomy of language which we have outlined here has strengths and weaknesses. The limitations are straightforward. It is very broad in scope, and therefore glosses over many important details: what exactly is an "acoustic–phonetic representation of speech"? What are the computations involved in mapping sound onto meaning, or auditory onto motor representations? (However, some existing models may fit well into the current framework, as suggested above.) It does not attempt to deal with the functional organization of frontal language systems, and consequently does not advance our understanding of the non-fluent aphasias. And, it does not address the role of subcortical systems in language processing, such as the basal ganglia, thalamus, and cerebellum.

On the other hand, this model integrates language processing networks into the broader scheme of cortical functional anatomy. It provides a context to interpret the neural basis not only of "traditional" language functions (broadly defined), such as speech perception, auditory comprehension, and speech production, but also provides a natural account of verbal working memory. And it provides a coherent framework for interpreting the symptom complex of several of the classical aphasias.

Acknowledgements

References

Aboitiz, F., & García, V. R. (1997). The evolutionary origin of language areas in the human brain. A neuroanatomical perspective. *Brain Research Reviews*, *25*, 381–396.

Aglioti, S., DeSouza, J. F. X., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, *5*, 679–685.

Allport, D. A. (1984). Speech production and comprehension: one lexicon or two? In W. Prinz, & A. F. Sanders (Eds.), *Cognition and motor processes* (pp. 209–228). Berlin: Springer-Verlag.

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

Anderson, J. M., Gilmore, R., Roper, S., Crosson, B., Bauer, R. M., Nadeau, S., Beversdorf, D. Q., Cibula, J., Rogish, M., III, Kortencamp, S., Hughes, J. D., Gonzalez Rothi, L. J., & Heilman, K. M. (1999). Conduction aphasia and the arcuate fasciculus: a reexamination of the Wernicke-Geschwind model. *Brain and Language*, *70*, 1–12.

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in working memory: PET evidence. *Psychological Science*, *7*, 25–31.

Baddeley, A. D. (1992). Working memory. *Science*, *255*, 556–559.

Baker, E., Blumstein, S. E., & Goodglass, H. (1981). Interaction between phonological and semantic factors in auditory comprehension. *Neuropsychologia*, *19*, 1–15.

Barde, L. F., Baynes, K., Gage, N., & Hickok, G. (2000). "Phonemic" perception in aphasia and in the isolated right hemisphere. *Cognitive Neuroscience Society Annual Meeting Program*, 43.

Basso, A., Casati, G., & Vignolo, L. A. (1977). Phonemic identification defects in aphasia. *Cortex*, *13*, 84–95.

Belin, P., & Zatorre, R. J. (2000). "What", "where" and "how" in auditory cortex. *Nature Neuroscience*, *3*, 965–966.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M. C., & Samson, Y. (1998). Lateralization of speech and auditory temporal processing. *Journal of Cognitive Neuroscience*, *10*, 536–540.

Benson, D. F., Sheremata, W. A., Bouchard, R., Segarra, J. M., Price, D., & Geschwind, N. (1973). Conduction aphasia: a clinicopathological study. *Archives of Neurology*, *28*, 339–346.

Benson, R. R., Whalen, D. H., Richardson, M., Swainson, B., Clark, V. P., Lai, S., & Liberman, A. M. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language*, *78*, 364–396.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, *10*, 512–528.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience*, *17*, 353–362.

Blumstein, S. E., Baker, E., & Goodglass, H. (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, *15*, 19–30.

Blumstein, S. E., Cooper, W. E., Zurif, E. B., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia*, *15*, 371–383.

Buchman, A. S., Garron, D. C., Trost-Cardamone, J. E., Wichter, M. D., & Schwartz, M. (1986). Word deafness: one hundred years later. *Journal of Neurology, Neurosurgery, and Psychiatry*, *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. *Cognitive Science*, *25*, 663–678.

Buchsbaum, B., Humphries, C., & Hickok, G. (2001). A new perspective on the functional anatomy of phonological working memory: fMRI investigations. *Cognitive Neuroscience Society Eighth Annual Meeting Program*, 87.

Burton, M. W., Small, S., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: an fMRI investigation. *Journal of Cognitive Neuroscience*, *12*, 679–690.

Caplan, D., Gow, D., & Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology*, *45*, 293–298.

Caplan, D., Hildebrandt, N., & Makris, N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain*, *119*, 933–949.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.

Carey, D. P. (2001). Do action systems resist visual illusions? *Trends in Cognitive Sciences*, *5*, 109–113.

Coleman, J. (1998). Cognitive reality and the phonological lexicon: a review. *Journal of Neurolinguistics*, *11*(3), 295–320.

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, *1*, 123–132.

Damasio, A. R. (1992). Aphasia. *New England Journal of Medicine*, *326*, 531–539.

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

Damasio, H., & Damasio, A. R. (1980). The anatomical basis of conduction aphasia. *Brain*, *103*, 337–350.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*, 801–838.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, *22*, 567–631.

Dronkers, N. (1994). Neural mechanisms of morphosyntactic comprehension deficits. *Cognitive Neuroscience Society Abstracts*, *1*, 63.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, *384*, 159–161.

Dronkers, N. F., Redfern, B. B., & Knight, R. T. (2000). The neural architecture of language disorders. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 949–958). Cambridge, MA: MIT Press.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales, & E. Walker (Eds.), *New approaches to language mechanisms*. Amsterdam: North-Holland.

Friederici, A. D., Meyer, M., & von Cramon, D. Y. (2000). Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language*, *74*, 289–300.

Friederici, A. D., Wang, Y., Hermann, C. S., Maess, B., & Oertel, U. (2000). Localization of early syntactic processes in frontal and temporal cortical areas: a magnetoencephalographic study. *Human Brain Mapping*, *11*, 1–11.

Fuster, J. M. (1995). Memory in the cerebral cortex. Cambridge, MA: MIT Press.

Gage, N., Poeppel, D., Roberts, T. P. L., & Hickok, G. (1998). Auditory evoked M100 reflects onset acoustics of speech sounds. *Brain Research*, *814*, 236–239.

Gainotti, G., Micelli, G., Silveri, M. C., & Villa, G. (1982). Some anatomo-clinical aspects of phonemic and semantic comprehension disorders in aphasia. *Acta Neurologica Scandinavica*, *66*, 652–665.

Galaburda, A. M. (1982). Histology, architectonics, and asymmetry of language areas. In M. A. Arbib, D. Caplan, & J. C. Marshall (Eds.), *Neural models of language processes* (pp. 435–445). San Diego, CA: Academic Press.

Gallese, V., Fadiga, L., Fogassi, L., Luppino, G., & Murata, A. (1997). A parietal-frontal circuit for hand and grasping movements in the monkey: evidence from reversible inactivation experiments. In P. Thier, & H.-O. Karnath (Eds.), *Parietal lobe contributions to orientation in 3D space* (pp. 255–270). Heidelberg: Springer-Verlag.

Garrard, P., & Hodges, J. R. (2000). Semantic dementia: clinical radiological, and pathological perspectives. *Journal of Neurology*, *247*, 409–422.

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

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, *349*, 154–156.

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

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

Green, E., & Howes, D. H. (1978). The nature of conduction aphasia: a study of anatomic and clinical features and of underlying mechanisms. In H. Whitaker, & H. A. Whitaker (Eds.), (*3*). *Studies in neurolinguistics*, New York: Academic Press.

Grossman, M., Payer, F., Onishi, K., D'Esposito, M., Morrison, D., Sadek, A., & Alavi, A. (1998). Language comprehension and regional cerebral defects in frontotemporal degeneration and Alzheimer's disease. *Neurology*, *50*, 157–163.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, *105*, 611–633.

Hart, J. J., & Gordon, B. (1990). Delineation of single-word semantic comprehension deficits in aphasia, with anatomical correlation. *Annals of Neurology*, *27*, 226–231.

Hickok, G. (2000). Speech perception, conduction aphasia, and the functional neuroanatomy of language. In Y. Grodzinsky, L. Shapiro, & D. Swinney (Eds.), *Language and the brain* (pp. 87–104). San Diego, CA: Academic Press.

Hickok, G. (2001). Functional anatomy of speech perception and speech production: psycholinguistic implications. *Journal of Psycholinguistic Research*, *30*, 225–234.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, *15*, 673–682.

Hickok, G., Erhard, P., Kassubek, J., Helms-Tillery, A. K., Naeve-Velguth, S., Strupp, J. P., Strick, P. L., & Ugurbil, K. (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. *Neuroscience Letters*, *287*, 156–160.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*, 131–138.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*, 1213–1216.

Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *NeuroReport*, *12*, 1749–1752.

Ide, A., Rodriguez, E., Zaidel, E., & Aboitiz, F. (1996). Bifurcation patterns in the human sylvian fissure: hemispheric and sex differences. *Cerebral Cortex*, *6*, 717–725.

Ishitobi, M., Nakasato, N., Suzuki, K., Nagamatsu, K., Shamoto, H., & Yoshimoto, T. (2000). Remote discharges in the posterior language area during basal temporal stimulation. *NeuroReport*, *13*, 2997–3000.

Ivry, R., & Robertson, L. (1998). The two sides of perception. Cambridge, MA: MIT Press.

Jones, D. M., & Macken, W. J. (1996). Irrelevant tones produce an irrelevant speech effect: implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 369–381.

Jonides, J., Schumacher, E. H., Smith, E. E., Koeppe, R. A., Awh, E., Reuter-Lorenz, P. A., Marshuetz, C., & Willis, C. R. (1998). The role of parietal cortex in verbal working memory. *Journal of Neuroscience*, *18*, 5026–5034.

Kaas, J. H., & Hackett, T. A. (1999). "What" and "where" processing in auditory cortex. *Nature Neuroscience*, *2*, 1045–1047.

Kertesz, A., Sheppar, A., & MacKenzie, R. (1982). Localization in transcortical sensory aphasia. *Archives of Neurology*, *39*, 475–478.

Levelt, W. J. M. (1989). Speaking: from intention to articulation. Cambridge, MA: MIT Press.

Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, *10*, 553–567.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–75.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Lichtheim, L. (1885). On aphasia. *Brain*, *7*, 433–484.

Lüders, J., Lesser, R. P., Hahn, J., Dinner, D. S., Morris, H. H., Wyllie, E., & Godoy, J. (1991). Basal temporal language area. *Brain*, *114*, 743–754.

MacKay, D. G. (1987). The organization of perception and action: a theory for language and other cognitive skills. New York: Springer-Verlag.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., & Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, *5*, 467–479.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McGlone, J. (1984). Speech comprehension after unilateral injection of sodium amytal. *Brain and Language*, *22*, 150–157.

Miceli, G., Gainotti, G., Caltagirone, C., & Masullo, C. (1980). Some aspects of phonological impairment in aphasia. *Brain and Language*, *11*, 159–169.

Milner, A. D., & Goodale, M. A. (1995). The visual brain in action. Oxford: Oxford University Press.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, *24*, 45–64.

Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S. J., & Hodges, J. R. (2000). A voxel-based morphometry study of the semantic dementia: relationship between temporal lobe atrophy and semantic memory. *Annals of Neurology*, *47*, 36–45.

Mummery, C. J., Patterson, K., Wise, R. J. S., Vandenbergh, R., Price, C. J., & Hodges, J. R. (1999). Disrupted temporal lobe connections in semantic dementia. *Brain*, *122*, 61–73.

Murata, A., Gallese, V., Kaseda, M., & Sakata, H. (1996). Parietal neurons related to memory-guided hand manipulation. *Journal of Neurophysiology*, *75*, 2180–2186.

Nakai, T., Matsuo, K., Kato, C., Matsuzawa, M., Okada, T., Glover, G. H., Moriya, T., & Inui, T. (1999). A functional magnetic resonance imaging study of listening comprehension of languages in human at 3 tesla-comprehension level and activation of the language areas. *Neuroscience Letters*, *263*, 33–36.

Nicholls, M. (1996). Temporal processing asymmetries between the cerebral hemispheres: evidence and implications. *Laterality*, *1*, 97–137.

Norris, D., & Wise, R. (2000). The study of prelexical and lexical processes in comprehension: psycholinguistics and functional neuroimaging. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 867–880). Cambridge, MA: MIT Press.

Okada, K., Smith, K.R., Humphries, C. & Hickok, G. (2003). Word length modulates neural activity in auditory cortex during covert object naming. *Neuroreport*, *14*, 2323–2326.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, *362*, 342–345.

Perenin, M.-T., & Vighetto, A. (1988). Optic ataxia: a specific disruption in visuomotor mechanisms. I. Different aspects of the deficit in reaching for objects. *Brain*, *111*, 643–674.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, *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.

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, *8*(4), 516–521.

Rizzolatti, G., Fogassi, L., & Gallese, V. (1997). Parietal cortex: from sight to action. *Current Opinion in Neurobiology*, *7*, 562–567.

Robin, D. A., Tranel, D., & Damasio, H. (1990). Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain and Language*, *39*, 539–555.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*, *2*, 1131–1136.

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, *21*, 150–164.

Saykin, A. J., Stafiniak, P., Robinson, L. J., Flannery, K. A., Gur, R. C., O'Connor, M. J., & Sperling, M. R. (1995). Language before and after temporal lobectomy: specificity of acute changes and relation to early risk factors. *Epilepsia*, *36*, 1071–1077.

Schlosser, M. J., Aoyagi, N., Fulbright, R. K., Gore, J. C., & McCarthy, G. (1998). Functional MRI studies of auditory comprehension. *Human Brain Mapping*, *6*, 1–13.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*, 2400–2406.

Sereno, A. B., & Maunsell, J. H. (1998). Shape selectivity in primate lateral intraparietal cortex. *Nature*, *395*, 500–503.

Shallice, T., & Warrington, E. (1977). Auditory-verbal short-term memory impairment and conduction aphasia. *Brain and Language*, *4*, 479–491.

Shelton, J. R., & Caramazza, A. (1999). Deficits in lexical and semantic processing: implications for models of normal language. *Psychonomic Bulletin and Review*, *6*, 5–27.

Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R. J., & Näätanen, R. (1998). Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: magnetic mismatch negativity study. *Neuroscience Letters*, *251*, 141–144.

Shuren, J. E., Schefft, B. K., Yeh, H. S., Privitera, M. D., Cahill, W. T., & Houston, W. (1995). Repetition and the arcuate fasciculus. *Journal of Neurology*, *242*(9), 596–598.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, *111*, 1872–1891.

Tanabe, H., Sawada, T., Inoue, N., Ogawa, M., Kuriyama, Y., & Shiraishi, J. (1987). Conduction aphasia and arcuate fasciculus. *Acta Neurologica Scandinavica*, *76*(6), 422–427.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Waldstein, R. S. (1989). Effects of postlingual deafness on speech production: implications for the role of auditory feedback. *Journal of the Acoustical Society of America*, *88*, 2099–2144.

Wernicke, C. (1874/1969). The symptom complex of aphasia: a psychological study on an anatomical basis. In R. S. Cohen, & M. W. Wartofsky (Eds.), *Boston studies in the philosophy of science* (pp. 34–97). Dordrecht: D. Reidel.

Wilshire, C. E., & McCarthy, R. A. (1996). Experimental investigations of an impairment in phonological encoding. *Cognitive Neuropsychology*, *13*, 1059–1098.

Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin and Review*, *8*, 44–57.

Wise, R., Chollet, F., Hadar, U., Friston, K., Hoffner, E., & Frackowiak, R. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain*, *114*, 1803–1817.

Wise, R. J. S., Greene, J., Büchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *Lancet*, *353*, 1057–1061.

Wise, R. J. S., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural sub-systems within "Wernicke's area". *Brain*, *124*, 83–95.

Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin*, *60*, 213–251.

Zaidel, E. (1985). Language in the right hemisphere. In D. F. Benson, & E. Zaidel (Eds.), *The dual brain: hemispheric specialization in humans* (pp. 205–231). New York: Guilford Press.

Zatorre, R. J. (1997). Cerebral correlates of human auditory processing: perception of speech and musical sounds. In J. Syka (Ed.), *Acoustical signal processing in the central auditory system* (pp. 453–468). New York: Plenum Press.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, *6*, 37–46.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*, 846–849.

Zatorre, R. J., Meyer, E., Gjedde, A., & Evans, A. C. (1996). PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cerebral Cortex*, *6*, 21–30.