

Proof

PART 5

**Auditory Discrimination and
Recognition**

Taylor & Francis
Not for distribution

Proof

Proof

Taylor & Francis
Not for distribution

Proof

MODELS OF SPEECH PROCESSING

Michael Grosvald, Martha W. Burton, and Steven L. Small

Introduction

One of the fundamental questions about language is how listeners map the acoustic signal onto syllables, words, and sentences, resulting in understanding of speech. For normal listeners, this mapping is so effortless that one rarely stops to consider just how it takes place. However, studies of speech have shown that this acoustic signal contains a great deal of underlying complexity. A number of competing models seek to explain how these intricate processes work. Such models have often narrowed the problem to mapping the speech signal onto isolated words, setting aside the complexity of segmenting continuous speech. Continuous speech has presented a significant challenge for many models because of the high variability of the signal and the difficulties involved in resolving the signal into individual words.

The importance of understanding speech becomes particularly apparent when neurological disease affects this seemingly basic ability. Lesion studies have explored impairments of speech sound processing to determine whether deficits occur in perceptual analysis of acoustic-phonetic information or in stored abstract phonological representations (e.g., Basso, Casati, & Vignolo, 1977; Blumstein, Cooper, Zurif, & Caramazza, 1977). Furthermore, researchers have attempted to determine in what ways underlying phonological/phonetic impairments may contribute to auditory comprehension deficits (Blumstein, Baker, & Goodglass, 1977).

In this chapter, we discuss several psycholinguistic models of word recognition (the process of mapping the speech signal onto the lexicon), and outline how components of such models might correspond to the functional anatomy of the brain. We will also relate evidence from brain lesion and brain activation studies to components of such models. We then present some approaches that deal with speech perception more generally, and touch on a few current topics of debate.

Psycholinguistic Models of Word Recognition

The speech signal contains well-known characteristics that provide constraints on models of word recognition (Frauenfelder & Floccia, 1998; Jusczyk, 1986). First, the fine acoustic details of every word or sentence differ each time it is produced. Thus, one challenge for models of word recognition is to explain how acoustically different tokens of a word, which may vary in myriad ways, are mapped into a single word (from what may be a number of similar words). Among other factors, variation among tokens may occur as a function of who is speaking, that is, the “talker” (e.g.,

through differences in voice or pronunciation); neighboring words (e.g., through coarticulatory effects); and prosodic context. In addition, the speech signal is continuous, lacking easily discernible discrete boundaries between sounds or words. Thus, any model must account for how diverse phonetic tokens that are part of a continuous signal can be mapped onto discrete units that can be recognized as individual sounds and words.

Different models of word recognition diverge on particular details, especially with regard to the nature of intermediate representations, but there does appear to be general consensus about word recognition that accounts for a wide range of experimental results from the psycholinguistic literature (Frauenfelder & Floccia, 1998).¹ Most models consist of processes by which sublexical units, such as phonetic features or phonemes, are extracted from the acoustic signal and then matched with the appropriate lexical entry.

Sublexical units have been postulated in most models of word recognition, including TRACE (McClelland & Elman, 1986), Cohort (Marslen-Wilson, 1987; Marslen-Wilson & Warren, 1994), and Merge (Norris et al., 2000), to provide a less variable, more abstract form of the acoustic signal. This is done in order to simplify the mapping of sublexical information onto the lexical level.

A primary difference among these models relates to the nature of input representations. Some models, like the Cohort model (Marslen-Wilson, 1987; Lahiri & Marslen-Wilson, 1991; Marslen-Wilson & Warren, 1994) rely on relatively small units like phonetic features that correspond to either acoustic or articulatory characteristics. For example, the feature of “voicing” refers to the presence of vocal cord vibration during the articulation of a speech sound and distinguishes voiceless sounds like /p/ from voiced ones like /b/. Other models, such as TRACE, proposed by McClelland and Elman (1986), rely more heavily on sets or bundles of phonetic features that define more abstract units, called segments or phonemes, that are considered the minimal units of sound distinguishing the meanings of words. For example, the initial phonemes, /p/ and /b/, which share the same place and manner of articulation features, but contrast in terms of voicing, differentiate the words “pear” and “bear.”

Different realizations of a given phoneme are not acoustically identical; some (though not all) of this variation tends to be systematic. For example, in English, the same phoneme, /p/, is produced with different phonetic realizations that depend partly on where it occurs in a syllable or word (e.g., with aspiration noise in initial position, as in “pill,” and without aspiration after /s/, as in “spill”). Such variations across spoken instances of a given speech sound are said to occur at the phonetic level, while at the more general phonemic level the fine detail of the physical characteristics of the speech signal is not relevant.

The nature of these abstract representations remains controversial. Early evidence indicated that listeners showed poor discrimination between acoustically different members of the same phonetic category (e.g., acoustically different exemplars of /d/, as in the classic categorical perception experiments of Liberman, Harris, Hoffman, & Griffith, 1957), suggesting that listeners perceive speech sounds using abstract phoneme category information and discard much of the fine acoustic detail.

Subsequent studies have indicated that listeners retain more acoustic detail in stored representations of speech than many researchers had initially believed (Kuhl, 1991; Pisoni, 1993). Kuhl and colleagues have argued that listeners maintain within-category distinctions for speech sounds such as vowels. In other words, some sounds within a given speech category (e.g., spoken instances of the vowel /i/) are considered by listeners to be better exemplars of the category than others, and in this sense can be considered more prototypical. Furthermore, these prototypical sounds serve as perceptual magnets, affecting the perception of other similar speech stimuli, whereas non-prototypical sounds do not show such effects. Along these lines, Kuhl (1991) provided evidence that listeners are sensitive to some within-category acoustic detail, and that speech prototypes may serve to organize categories of sounds during perception.

Pisoni (1993) further challenged traditional views of speech perception, arguing that even more acoustic information (including variability from utterances of individual talkers and context effects, such as rate) is preserved from initial perceptual analysis of the speech signal, and in addition provides important information for later recognition of specific talkers (Nygaard, Sommers, & Pisoni, 1995). According to Pisoni (1993), information about specific acoustic characteristics of particular talkers is stored in long-term memory.

Despite these findings, models of speech perception still tend to assume that during perceptual analysis of the signal listeners are able to “normalize” variability in the signal and map acoustic input onto abstract representations such as features (Marslen-Wilson & Warren, 1994) and phonemes (Norris et al., 2000). Some models of speech perception, such as the connectionist TRACE model of McClelland and Elman (1986), incorporate abstract segmental units as well as phonetic features, whereas other models map spectral properties of the acoustic signal directly onto syllables (Klatt, 1979). The motor theory of speech perception posits that these intermediate units correspond to articulatory gestures (Lieberman et al., 1967; Liberman & Mattingly, 1985; Galantucci et al., 2006).

The mapping of these input representations onto the word level characteristically involves activation of lexical entries, for which the level of activation is determined by the goodness of fit between the input representation and the lexical representation. Typically, once a threshold of activation has been reached a word is recognized. An example of such a model implemented using the connectionist framework is TRACE (McClelland & Elman, 1986), which has been used to account for a number of effects of lexical influence on phonetic categorization (e.g., Pitt & Samuel, 1993).

Despite the common use of an activation mechanism to represent how words are recognized, models differ in the mechanistic details of the activation process, such as the amount of top-down processing that is permitted from higher levels to lower levels, or the degree to which inhibition between competing representations plays a role. For example, the Merge model maintains that top-down (lexical) information does not influence categorization of lower-level phonetic information (Norris et al., 2000), whereas models such as TRACE allow such feedback (McClelland & Elman, 1986). Other models, such as the neighborhood activation model, emphasize the role of competition among lexical entries in facilitating word recognition (Luce & Pisoni, 1998). In that model, the number of similar words, their degree of phonetic similarity, and frequency of occurrence in a language (similarity-neighborhood) affect the speed and accuracy of word recognition.

Despite such differences among models, the generally accepted view—that the speech signal is analyzed into sublexical units that activate word entries—provides a framework for how the most basic processes of word recognition proceed (Frauenfelder & Floccia, 1998). Although lesion studies have often seemed inconsistent regarding the physical loci of these processes in the brain, more recent results offer some revealing insights. Evidence from a variety of experimental tasks in neuropsychological studies, and from functional neuroimaging of normal adults, reveals a network of regions of the brain that appear to correspond to specific functions postulated in models of speech processing. Below we will consider what specific areas of the brain may be involved in the analysis and mapping of the speech signal onto lexical information. We will argue that although listeners have access to different kinds of sublexical representations, their use depends critically on task demands, and the effects of such task demands are reflected in activation patterns seen in neuroimaging studies.

Evidence from Aphasia

Experiments exploring speech perception in stroke patients with aphasia have focused attention on segmental contrasts within words (e.g., /p/ vs. /b/ in *pear/bear*) or in nonsense syllables (e.g., *pa/ba*) (Basso et al., 1977; Benson, 1988; Blumstein, Cooper, et al., 1977; Carpenter &

Rutherford, 1973; Miceli, Caltagirone, Gainotti, & Payer-Rigo, 1978; Oscar-Berman, Zurif, & Blumstein, 1975; Riedel & Studdert-Kennedy, 1985). Nearly all patients show some impairment in discrimination (“same-different” judgments) and/or labeling or identification (e.g., “Is the first sound a *p* or a *b*?”) (Blumstein, 1998). Patterns of errors on these tasks follow those seen in speech production errors. For example, more errors occur in medial or final position in the syllable than in initial position (Blumstein, 1998). Individuals with lesions in the temporal lobe might be expected to show speech discrimination deficits because primary auditory areas are located in the temporal lobe and have direct connections to the auditory association areas. However, individuals with anterior lesions also manifest such impairments. These data have challenged the traditional view that speech perception impairments should be associated solely with posterior lesions in people with aphasia.

Several lines of evidence suggest that speech perception impairments in aphasic individuals do not occur at the stage of extracting sublexical information from the acoustic signal. Patients’ performance is generally better on discrimination tasks than on identification tasks, indicating that they are sensitive to acoustic-phonetic differences in the stimuli, but may have difficulty providing accurate responses based on segmental information (Gow & Caplan, 1996). Furthermore, patterns of results in discrimination tasks are similar to those of healthy controls in the location of boundaries between phonetic categories and in the overall shape of the discrimination functions, even in cases where patients cannot label the stimuli (Blumstein, 1998).

Blumstein and colleagues (2000), in an investigation of the effects of phonological priming, offered further evidence that participants with aphasia are able to extract useful acoustic properties from the signal. In a set of priming tasks, aphasic participants presented with pairs of words or nonwords performed a lexical decision on the second member of the pair. The phonological properties of the first member of the pair were systematically varied (e.g., producing either rhyming or unrelated pairs like “pear-bear” vs. “pen-bear”). Both Broca’s and Wernicke’s aphasic participants showed effects of rhyming primes on lexical decision times, suggesting that both types of aphasic participants are sensitive to phonological similarity (see also Gordon & Baum, 1994). In a second experiment, the researchers tested repetition priming with these same aphasic participants using lexical decision on repeated words occurring at different intervals (i.e., presentation of the same word immediately following the prime, two words following the prime, eight words following the prime, and so on). The results suggested the presence of a second impairment in these participants, namely, maintaining an acoustic form in short-term memory. Unlike normals, the aphasic participants showed neither increased repetition effects at shorter intervals compared to longer intervals nor any repetition effects for nonwords. As long as the lexicon was activated by a real word, aphasic participants showed priming effects. However, the lack of increased repetition effects at shorter intervals suggests that the aphasic participants were “matching” the meaning of the stimulus, not its phonetic form, which may have dissipated in working memory faster for the aphasic participants than for controls. The lack of increased repetition effects at longer intervals and nonword effects were consistent across all aphasic participants.

These findings are also in agreement with other patterns of deficits described in single case studies. For example, Martin, Bredin, and Damian (1999) report a patient able to perform phoneme discrimination at a level similar to that of normals, yet speech discrimination declined with increased interstimulus interval, suggesting a short-term memory deficit. Such short-term memory impairments of maintaining information over time certainly contribute to auditory language comprehension deficits, such as those in which listeners must actively retrieve specific pieces of information over the longer periods of time required by sentence processing (Caplan & Waters, 1995; Martin & Romani, 1994; Miyake, Carpenter, & Just, 1994). In contrast, low-level speech

perception impairments may have some limited role in higher-level language comprehension, but do not appear to account for severe auditory language comprehension difficulties (Blumstein, Baker, et al., 1977). Blumstein and colleagues found that the level of performance on identification and discrimination of consonant-vowel syllables that do not require maintenance over long intervals was a poor predictor of auditory comprehension as measured by standard clinical diagnostic tests.

In summary, a number of studies have investigated speech perception in aphasic individuals using different experimental tasks. The results have generally demonstrated that impairments are not due to initial acoustic analysis of the speech signal. Because the patterns of impairment have not clearly corresponded to particular levels of representation in speech perception models and have had unclear localization, they have not played a significant role in models of normal speech perception. Although one might hypothesize that components of sublexical processing could be selectively impaired (e.g., perceptual analysis of the signal vs. segmentation), neuropsychological studies have shown that nearly all aphasic individuals display some impairments in speech perception. These difficulties performing a range of tasks do not generally appear to be due to early stages of sublexical speech processing in which the acoustic information, such as temporal or spectral properties, are extracted from the signal. Thus, even if aphasic participants have left temporal damage, they may be able to perform some speech discrimination tasks because of spared right hemisphere structures.

In an effort to better understand the relationship between location of brain lesions and deficits in language comprehension, Dronkers et al. (2004) conducted a study of 64 participants with aphasia due to left hemisphere stroke, who were asked to choose drawings that best matched the meanings of auditorily presented sentences. The results of a voxel-based lesion-symptom mapping analysis (Bates et al., 2003) identified several relevant left hemisphere regions: posterior middle temporal gyrus, anterior superior temporal gyrus, superior temporal sulcus and angular gyrus, and two frontal regions—Brodmann areas 46 and 47. Of these regions, middle temporal gyrus seemed to be the one most involved in word-level speech comprehension, with the others apparently more involved in comprehension at the sentence level. Interestingly, the analysis found that significant deficits in language comprehension were neither associated with lesions of Broca's area nor Wernicke's area.

Anatomically, the effects of hypometabolism in temporal and temporoparietal regions have been found for patients with various types of aphasia and thus may provide a potential common neural substrate (Metter et al., 1989, 1990). Behaviorally, many people with aphasia have problems with auditory working memory that may play a greater role in speech identification (comparison of stimulus to a stored representation) than speech discrimination (comparison between two exemplars).

People with aphasia may show some impairment in discrimination tasks as well as identification tasks. It remains unclear to what extent hemispheric differences in the posterior temporal lobe play a role in speech perception impairments. However, theories that postulate hemispheric asymmetries predict difficulties with those speech contrasts that rely on rapidly changing frequency information (discussed further in the section on superior temporal regions, below). This could explain why some adults with focal brain lesions in the left hemisphere have difficulties with rapid formant transitions in stop consonants (Ivry & Robertson, 1998).

In broad terms, however, basing our understanding of the underlying neuroanatomy of speech perception on lesion studies alone leaves a number of unanswered questions. The common finding that inferior frontal cortex is activated in phonological tasks and in tasks that are designed to require verbal working memory suggests that the role of working memory requires more attention in explaining impaired performance on certain speech perception tasks.

Functional Neuroanatomy of Speech Perception

In contrast to the lesion data, functional neuroimaging studies of healthy participants have been converging on a set of regions involved in specific functional components of speech perception. From these studies, it is possible to gain insight into why some aphasic participants may have difficulty with particular tasks, such as phoneme identification, and to determine whether processing of different types of sublexical information (e.g., acoustic-phonetic features and phonemes) produces distinct patterns of brain behavior (e.g., patterns of activation, electromagnetic responses). Evidence that listeners are sensitive to sublexical information (acoustic-phonetic features and/or phonemic category information)—as demonstrated through neurobiological responses—will help to clarify what role sublexical information may play in models of speech perception.

Functional neuroimaging studies of speech perception typically rely on one of two task-dependent imaging methods, PET (positron emission tomography) or fMRI (functional magnetic resonance imaging), to reveal areas of the brain that participate in a task, yet when damaged may not necessarily impair performance of language functions. Although PET and fMRI differ in a number of aspects of experimental design and analysis and have different spatial and temporal properties, both methods involve imaging some (presumed) correlate of neural activity during performance of a cognitive task. These methods allow observation of the particular areas of the brain that participate in a cognitive task (for reviews of neuroimaging methods see Burton & Small, 1999; Bandettini, 2009). Magnetoencephalography (MEG) provides highly accurate information about the time-course of processing for relatively small areas of cortex by measuring neuromagnetic responses to stimulation. For that reason, MEG has been used to study speech perception, which takes place within milliseconds in the temporal lobe. Other methods, such as event-related potentials (ERP), that provide highly accurate time-course information but not spatial localization, have also been used to identify stages of speech processing (e.g., Rinne et al., 1999).

Long-standing evidence indicates that three main areas of activity—superior temporal gyrus and sulcus, inferior parietal lobule, and inferior frontal gyrus—perform critical functions during the processing of speech. Figure 13.1 shows the approximate location of these areas on a structural image of the left lateral cortex of a normal subject. More recent data indicate that other brain regions are crucially involved as well, and together with the three regions just mentioned, form networks or processing streams whose configuration can depend at least partly on task demands (Hein & Knight, 2008; Londei et al., 2010; Price, 2012).

Superior Temporal Regions

Some of the most compelling evidence for the role of superior temporal cortex in early speech analysis is the consistent pattern of activation when passive speech is compared to rest or to nonspeech sounds. Although activation in primary auditory cortex and auditory association areas is seen for nonlinguistic as well as for language sounds (for an overview, see Price, 2012), much evidence suggests that during speech perception initial acoustic/phonetic analysis also takes place in these regions. During this process, it appears that phonetic/phonemic information is extracted from the acoustic signal (Binder et al., 1994; Dhankhar et al., 1997; Gage, Poeppel, Roberts, & Hickok, 1998; Kuriki, Okita, & Hirata, 1995; Petersen, Fox, Posner, Mintun, & Raichle, 1988; Poeppel et al., 1996; Price et al., 1992; Wise et al., 1991; Zatorre, Meyer, Gjedde, & Evans, 1996).

Initial research comparing speech and nonspeech perception often contrasted speech to tones, and found greater temporal activity for the speech stimuli, particularly in the left hemisphere. This

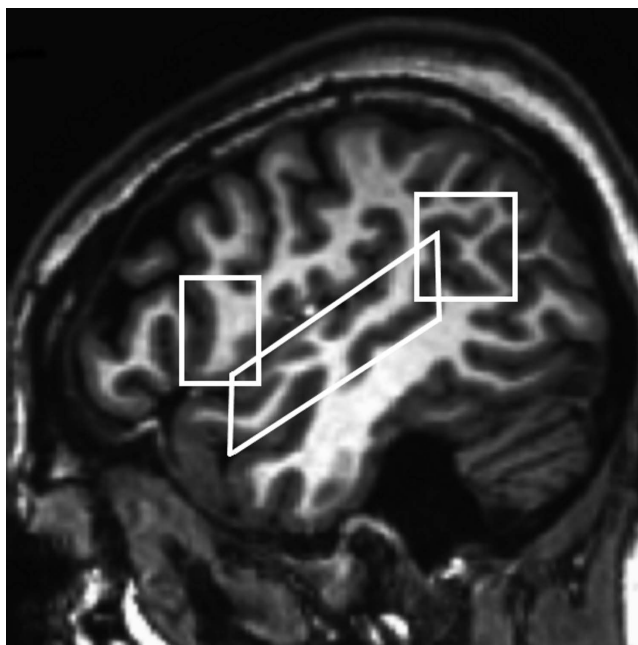


Figure 13.1 Sagittal view of left lateral cortex acquired using structural magnetic resonance imaging. White boxes highlight three areas involved in components of speech processing: the posterior two-thirds of inferior frontal gyrus (Broca's area); superior temporal sulcus and gyrus, including primary auditory cortex; and inferior parietal lobe.

suggested that specialized mechanisms for speech-related processing might be at work (Burton, Blumstein, & Small, 2000; Mummery, Ashburner, Scott, & Wise, 1999), but such studies left open the possibility that preferential activity for speech might be due to the greater acoustic complexity of speech syllables compared to the tone stimuli that had been used. However, a number of subsequent imaging studies comparing speech and nonspeech perception have controlled for acoustic complexity and have found a preferential response to speech in the left posterior superior temporal sulcus (Narain et al., 2003; Giraud et al., 2004; Hugdahl et al., 2003; Benson et al., 2006; Rimol et al., 2006). Such findings are bolstered by the results of Dehaene-Lambertz et al. (2005) and Meyer et al. (2005), who found greater activation in left posterior temporal cortex when distorted (sine-wave) speech was perceived by listeners as language, relative to when it was not. More generally, studies have tended to find that greater sound familiarity and greater acoustic complexity are associated with superior temporal activity that is, respectively, more posterior or more anterior (Price, 2012).

Despite the tendency in some studies for greater left temporal activation than right, and the predilection to discuss more fully the left hemisphere findings, the activation across imaging studies of speech perception has been consistently bilateral (Hickok & Poeppel, 2000). Further evidence for bilateral activation comes from fMRI studies of speech perception in normals and two aphasic patients listening to speech presented at varying rates (Mummery et al., 1999). In normal subjects, increasing rates of speech correlate with greater bilateral superior temporal activation in response to speech stimuli. Participants with left temporal infarction who perform well on single word comprehension tests have right superior temporal activity correlated with the rate of speech presentation, but no significant left temporal activity. Thus, their spared speech comprehension

ability could be due to involvement of the right hemisphere, which, in normal controls, is more characteristically involved in prelexical processing of speech.

Based on patterns of bilaterality, some researchers have postulated different roles for the left and right superior temporal cortices in speech signal analysis. One such hypothesis claims that left hemisphere is specialized for processing quickly changing frequency information, such as the rapid transitions that occur in stop consonants, such as /p t k b d g/ (Hesling et al., 2005; Hickok & Poeppel, 2000; Nicholls, 1996; Poeppel, 2003). For example, Belin and colleagues (1998) found significantly greater activation for left auditory cortex than right when subjects were presented with speech-like rapid (e.g., 40 msec) acoustic changes that had identical temporal structure to frequency transitions between consonants and vowels. In contrast, the right hemisphere has been associated with processing of slower-changing information found in some speech sounds, such as fricatives and nasals, compared to the rapid frequency transitions found in stops (Allard & Scott, 1975; Gage et al., 1998) or spectral information (Zatorre, 1997). Slower-changing information may be useful for processing that occurs over longer windows of time, such as prosody or the melody of a sentence. Evidence consistent with this sort of difference in hemispheric processing has been found in a number of subsequent studies (Husain et al., 2006; Rimol et al., 2005; Zaehle et al., 2004).

A related view is that the left hemisphere is more specialized for processing that requires greater temporal precision and the right hemisphere for computations that require greater spectral resolution (Zatorre, Belin, & Penhune, 2002). In still another proposal, double filtering by frequency, it is argued that an attentional filter determines the relevant frequency properties for analysis of the acoustic signal (Ivry & Robertson, 1998). Parts of the signal that occupy comparatively high portions of the frequency spectrum, relative to an anchoring point, are processed in the left hemisphere, whereas relatively low-frequency components of the signal are processed in the right hemisphere.

In summary, a number of proposals seek to explain the apparent differences in the roles of the left and right hemispheres in early acoustic/phonetic analysis. Whether these processes are speech specific remains to be determined. It may be possible for the contralateral hemisphere to compensate for loss when damage occurs; thus, it may be difficult to see evidence of hemispheric differences in aphasia (see also chapter 14, this volume).

Inferior Parietal Regions

Once the initial acoustic analysis is performed, the resulting sound representation must make contact with lexical-semantic information. Inferior parietal regions have consistently been implicated in such processes, often appearing to act in concert with frontal regions, and with the angular and supramarginal gyri appearing to play somewhat different roles.

In Demonet et al. (1992, 1994b), participants made semantic judgments on words and phonological judgments on nonwords, and showed widespread activation in frontal, temporal, and inferior parietal regions. The results indicated that the angular gyri (acting together with left temporal regions) are more involved with semantic processing at the word level, while the supramarginal gyri (together with left inferior frontal cortex) appear to be more involved with tasks requiring phonological decision making.

Other early studies found activation in inferior parietal regions for tasks requiring short-term storage of phonological information (Awh, Smith, & Jonides, 1995; Jonides et al., 1998; Paulesu, Frith, & Frackowiak, 1993; Paulesu et al., 1996). This activation was reported more consistently for tasks that involved stimuli requiring more extensive phonological coding (e.g., nonwords) than with word tasks in which subjects could use a combination of semantic and phonological

coding to store verbal material (Jonides et al., 1998). One method of maintaining information in a phonological store is through rehearsal. If this area were part of an auditory-motor (articulatory) integration network that includes inferior frontal regions, the concomitant activity in the inferior frontal lobe due to rehearsal and inferior parietal lobe due to temporary storage could be explained as a single network of regions that participate in speech processing (Hickok & Poeppel, 2000). In a number of studies (Myers et al., 2009; Ravizza et al., 2011; Zevin et al., 2010), for tasks in which working memory demands were minimized, fronto-parietal areas associated with verbal short-term memory were also engaged. The results of these and other studies (Davis et al., 2007; Elmer et al., 2011) suggest that this fronto-parietal network is involved in auditory and categorization processes that are not language specific (Price, 2012; see also chapters 11 and 14, this volume).

Gow et al.'s (2008) combined MEG, EEG, and MRI study found evidence that interplay between inferior parietal and superior temporal regions underlie a lexical-phonological perceptual phenomenon known as the Ganong effect (Ganong, 1980; Warren, 1970; Samuel & Pitt, 2003). This phenomenon occurs when an acoustically ambiguous language sound (e.g., a fricative sound intermediate between /s/ as in "sin" and /ʃ/ as in "shin") is inserted into a phoneme sequence that may or may not be interpretable as a word, depending on how the ambiguous sound is interpreted (e.g., Chri[s]mas forms a word while Chri[ʃ]mas does not). Listeners in such a situation will tend to interpret the sound in such a way that a real lexical item is perceived, providing evidence that top-down processes influence lower-level phonetic perception. Gow and colleagues' analysis indicated that lexical-level processing taking place in supramarginal gyrus affects phonetic-level processing in posterior superior temporal gyrus. Clarifying the interplay between brain regions is increasingly seen as crucial to an understanding of the human language system, a theme we will return to in the section on processing streams, below.

Inferior Frontal Regions

Activation of the left inferior frontal cortex (particularly the inferior frontal gyrus) has long been attributed to phonological processing in auditory tasks (Demonet et al., 1992; Demonet, Price, Wise, & Frackowiak, 1994a; Zatorre et al., 1992, 1996). Reviews of early PET evidence found that the patterns of activation did not converge as expected (compare Poeppel, 1996, and Demonet, Fiez, Paulesu, Petersen, & Zatorre, 1996), but subsequent researchers have tested phonological processing with increasingly specific cognitive tasks. The results of such studies suggest that there are indeed common underlying speech processes that may specifically activate these regions.

Based on PET data, Zatorre and colleagues (1992, 1996) argue that inferior frontal activation during particular speech tasks reflects segmentation processes in which listeners must separate speech sounds contained within syllables. For example, in pairs of spoken consonant-vowel-consonant sequences where the vowels differ (e.g., "fat"–"tid"), subjects making a "same/different" judgment on the final consonant must separate out the final consonant from a continuous acoustic signal. By using stimuli in which different vowels precede the consonant within the pairs, the need for segmentation is ensured because the consonant transitions vary as a function of preceding vowel context. Furthermore, Zatorre and colleagues (1992, 1996) argued that in performing this segmentation, listeners are required to access articulatory representations. In this account, such recoding of acoustic information to articulatory gestures is said to require Broca's area involvement because of its traditional association with articulatory deficits in neuropsychological studies of aphasic patients. Results of experiments comparing speech discrimination to either pitch discrimination or passive listening have shown activation of the left frontal cortex in the phonetic discrimination task. However, Zatorre and colleagues

(1992, 1996) found no such activation in Broca's area under passive listening conditions to the same stimuli.

The location of activation in the speech discrimination tasks was in the most posterior and superior aspect of Broca's area. This subregion of Broca's area is similar to that reported in several other studies of auditory phonological processing that involved segmentation of speech sounds and comparison of stimuli for a decision (Burton et al., 2000; Demonet et al., 1992; Demonet et al., 1994a). These studies compared activation for tasks involving phoneme monitoring for sequences of sounds or consonant segmentation to passive or sensory tasks. This area differs from the regions cited in studies of semantic tasks that have argued for Broca's area activation (Pol-drack et al., 1999). Functional specialization of different subregions of Broca's area seems likely to explain—at least in part—why multiple functions have historically been attributed to Broca's area as a whole (Dronkers, 1998).

A related question is whether listeners are always required to access segmental information in speech discrimination tasks. In segmentation tasks like those used by Zatorre and colleagues (1992, 1996), in which there are multiple differences among the segments in the stimuli, it appears likely that subjects must compare entire segments in order to make a same/different judgment. In contrast, if segments differ by only one phonetic feature, such as voicing, there may be no need for segmentation, since the subject only has to perceive a single phonetic difference between the stimulus pairs to make a decision. Burton et al. (2000) investigated this issue using an overt speech discrimination task in which subjects were required to make a same/different judgment about phonetic segments in initial position. It was expected that Broca's area would be involved only when the subject had to perform a task requiring articulatory recoding. Significantly more frontal activation was seen for tasks requiring overt segmentation (e.g., “dip–doom” pairs) compared to those that did not (e.g., “dip–tip” pairs). Thus, phonetic judgments may invoke different neural mechanisms depending on task demands.

Importantly, Burton and colleagues (2000) found significant superior temporal gyrus activation (with a trend toward left-lateralization) regardless of whether the speech task required segmentation. Thus, frontal areas might not typically be recruited in the processing of speech for purposes of at least some speech discrimination tasks (i.e., those that do not require segmentation) and may not necessarily be invoked on a more global level for auditory language comprehension, a situation in which listeners may not need to identify individual sounds to accomplish word recognition. These results also suggest that posterior brain structures participate in initial perceptual analysis of the signal that is necessary for the subsequent mapping of acoustic/phonetic patterns onto higher levels of language, such as meaning.

Although functional neuroimaging studies of speech focusing on frontal activity have typically concentrated on segmental phonological tasks, imaging studies have suggested that suprasegmental information such as pitch contour may also activate inferior frontal cortex in discrimination tasks (Gandour et al., 2000; Gandour, Wong, & Hutchins, 1998; Gandour, 2007). Pitch cues can be used to distinguish lexical meaning in tone languages, such as Mandarin Chinese or Thai. Cross-linguistic studies provide the opportunity to compare perception of the same pitch contrasts under circumstances in which the pairs of speech stimuli are linguistically distinctive (e.g., two Thai words varying in lexical tones that have different meanings) to cases where they are not (e.g., English). When a pitch discrimination condition was compared to a filtered speech control task in which semantic and phonological information was eliminated, but other suprasegmental information was preserved, Gandour and colleagues found that only Thai speakers showed inferior frontal activation. Similar to English speakers, Chinese listeners with experience in tonal language, but not the particular tone distinctions in Thai, did not show activation in Broca's area. These findings indicate that Broca's area subserves not only segmental, but also suprasegmental processing.

In response to a number of findings in both spoken and written language, Bornkessel-Schlesewsky and Schlewsky have developed a model called the extended argument dependency model (Bornkessel & Schlewsky, 2006; Bornkessel-Schlesewsky & Schlewsky, 2009). This model is informed by—and attempts to account for—the fact that qualitatively similar phenomena can elicit different processing consequences in different languages, or the reverse. The basic intuition behind the model is that the flow of language processing is broadly similar cross-linguistically, but differences arise due to the kinds of information that are available, relevant, or prominent in different languages. For example, word order is relatively fixed in English but flexible in Czech; case is rudimentary or nonexistent in languages like English and Mandarin Chinese but much more important in German and Russian. This leads to differences in how languages encode information such as agency (i.e., “who did what to whom”). The result is that the same sentence—modulo translation—can elicit one kind of processing consequence in one language (e.g., a particular ERP waveform) but not in another.

Processing Streams

In addition to providing a better understanding of how specific brain regions contribute to linguistic processing, recent work has also led to a growing appreciation of how brain regions operate together. Much of the current work on processing “streams” in language comprehension has its origins in the visual processing literature. Frequently these are tied to some degree to known anatomical pathways. For example, two visual processing pathways in macaque, one “ventral” and the other “dorsal,” appear to have different functions (Ungerleider & Mishkin, 1982). The ventral stream projects from visual cortex (striate and prestriate) to inferior temporal areas and is important for object recognition (“what”), while the dorsal stream connects visual cortex to inferior parietal areas and is important for spatial localization (“where”) and visual-motor transformations (“how”). These two pathways may therefore reflect a fundamental distinction between perception- or sensory-oriented processing on the one hand and action-oriented processing on the other (Goodale & Milner, 1992).

Rauschecker and colleagues (Rauschecker, 1998; Rauschecker & Tian, 2000) proposed that, like visual processing, auditory processing is also segregated into dorsal and ventral streams. Support for these “ventral/what” and “dorsal/where” streams has been found in connectivity studies of structure/function relations (Saur et al., 2008) and in statistical relations among regions (i.e., “functional connectivity”) (Obleser et al., 2007; Londei et al., 2010; Leff et al., 2008; Schofield et al., 2009; Osnes et al., 2011; Eickhoff et al., 2009; Nath & Beauchamp, 2011). There is also evidence that the anatomical pathways subserving these “what” and “how” functions converge in frontal regions (Romanski et al., 1999; Scott & Johnsrude, 2003). Much subsequent work has attempted to elucidate the structure of these processing streams and to clarify their roles in language processing (e.g., Bornkessel-Schlesewsky & Schlewsky, 2013; Bornkessel-Schlesewsky, Schlewsky, Small, & Rauschecker, 2014; Friederici, 2012; Hickok & Poeppel, 2004, 2007; Rauschecker & Scott, 2009).

Although anatomical and functional details of these pathways are not yet fully understood, the general dorsal versus ventral distinction is now widely recognized, and much ongoing work seeks to clarify their respective roles. For example, DeWitt and Rauschecker’s (2012) meta-analysis of functional imaging studies related to auditory processing found specific evidence of hierarchically organized linguistic processing along the ventral pathway. Processing related to progressively more temporally complex linguistic sounds (i.e., phonemes and words) appears to be localized to areas progressively further along the ventral stream, with left mid-superior temporal gyrus implicated in short-timescale sound form (phoneme-level) processing, and processing related to the integration

of such items into longer forms (words) occurring in left anterior superior temporal gyrus (STG). Accordingly, there is a tendency toward greater invariance further along this processing stream. Poeppel et al. (2008) also argue that auditory analysis in at least two different timescales, corresponding to approximately the feature/segment and syllable/word levels, occurs during speech processing.

In addition to supporting auditory-motor mapping, the dorsal pathway also appears to be involved in syntactic processing (Bornkessel-Schlesewsky & Schlesewsky, 2013; Bornkessel-Schlesewsky, Schlesewsky, Small, & Rauschecker, 2014; Friederici et al., 2006; Wilson et al., 2010), a qualitatively different function. In light of these and other findings, Friederici (2012) presents a model with two separate dorsal streams, one joining temporal and premotor cortex via inferior parietal cortex and the superior longitudinal fasciculus, the other linking temporal cortex with Brodmann area 44 via the arcuate fasciculus. The ventral pathway may support simple syntactic processing, in addition to its more generally recognized role in auditory-semantic mapping. Friederici's (2012) model incorporates two ventral streams to accommodate this. At the same time, Friederici (2012) acknowledges some limitations of her model—for example, in not specifying a clear role for the right hemisphere in language processing. The work discussed earlier that attempts to account for apparent hemispheric differences in early acoustic/phonetic analysis (Hesling et al., 2005; Hickok & Poeppel, 2000; Ivry & Robertson, 1998; Nicholls, 1996; Poeppel, 2003; Zatorre, Belin, & Penhune, 2002) may suggest at least a partial solution to that problem. More generally, as we learn more about the functional and structural complexity of the human language system, no doubt further refinements leading to still better models will be possible.

Taylor & Francis *Motor Region Involvement in Speech Perception*

As noted earlier, the motor theory of speech perception (Lieberman et al., 1967; Liberman & Mattingly, 1985; Galantucci et al., 2006) posits that the percepts of speech are the articulatory gestures used in producing that speech or, more specifically, the neural commands that produced those gestures. Another way of expressing this idea is that understanding speech requires simulation of that speech on the part of the listener. It has been debated whether motor simulation is strictly necessary for speech perception or might simply be helpful in some situations. Though the former view appears untenable (Hickok et al., 2011; Pulvermüller & Fadiga, 2010; Rogalsky et al., 2011; Scott, McGettigan, & Eisner, 2009, 2013; Skipper et al., 2006; Tremblay & Small, 2011a, 2011b), evidence does suggest that motor regions may play a role in speech perception in particular contexts, such as under challenging perceptual conditions, or when useful visual input is present (Callan et al., 2010; Gow & Segawa, 2009; Hasson et al., 2007; Osnes et al., 2011; Skipper et al., 2005; Tremblay & Small, 2011b).

This topic has received renewed attention with the discovery of so-called “mirror neurons.” These neurons, first discovered in macaque, were so named because they fire selectively when the animal carries out certain kinds of manual or oral actions, as well as when the animal observes others executing those same actions (Gallese et al., 1996; Rizzolatti et al., 1996). Therefore, this “mirror” property could in principle enable a perceiver to relate others' actions with the perceiver's own motor plans. Mirror neurons were first found in a ventral premotor region of macaque, area F5, and were later discovered in a parietal region, area 7b (Gallese et al., 2002). Broca's area has sometimes been considered the human homologue of area F5, and these findings have led to speculation about the role a human fronto-parietal mirror neuron system might play in action understanding in general, language processing in particular, and in the evolution of language itself (Arbib, 2005; Rizzolatti & Arbib, 1998).

Two pertinent topics in this line of research are the role of visual information in speech processing, and the likelihood that substantially overlapping networks are involved in speech production and perception (e.g., Poeppel & Monahan, 2008; Pulvermüller & Fadiga, 2010; Rossi et al., 2011; Scott, McGettigan, & Eisner, 2013; Skipper et al., 2006; Skipper et al., 2007). It has long been known that speech perception can be aided (Sumby & Pollack, 1954), or even altered, by accompanying visible facial movements. The latter situation occurs, for instance, in the well-known McGurk effect, in which auditory “pa” and visual “ka” syllables fuse to induce the perception of a “ta” syllable (McGurk & MacDonald, 1976).

Skipper, Nusbaum, and Small (2006) propose a model in which available visual information (the latter of which can include both facial movements and co-speech gestures) aids in the perception of speech by helping the listener compute a hypothesized motor plan for the heard utterance. This computed motor plan can influence what the listener perceives. Such computations are performed by a dorsal-stream mirror neuron system whose basic “building blocks” are paired inverse and forward models. While inverse models map perceived actions to motor plans hypothesized to cause such actions, forward models map (simulated) motor commands to predicted sensory consequences. One function of such a system is to delimit the set of possible interpretations of the incoming language signal. This is consistent with the view that activity in inferior frontal or premotor regions during speech processing may be due at least in part to the operation of top-down mechanisms that constrain the output of bottom-up processing in temporal cortex (Dehaene-Lambertz et al., 2005; Price, 2012; Zekveld et al., 2006).

Conclusion

Although differing in some details, particularly with regard to the nature of sublexical representations, models of word recognition have broadly converged in overall structure: sublexical information is extracted from the speech signal and mapped onto a lexical level via some kind of activation method. Because of the similarities of these models, they are often difficult to evaluate against each other solely on the basis of psycholinguistic evidence. Nevertheless, functional neuroimaging evidence suggests that different brain areas are recruited for the processing of distinct types of sublexical information, and provides support for the basic framework of these models. More generally, functional neuroimaging studies of normal subjects are converging upon a number of brain regions associated with different aspects of speech processing, and are offering greater insight into how these regions operate in concert.

Knowledge about the function of brain regions in components of normal speech processing has contributed to understanding the nature of some aphasic deficits. Specifically, aphasic individuals with damage to anterior structures may have difficulty with tasks that require explicit segmentation of the speech signal. For example, they may have trouble with phoneme identification tasks, which require such segmentation. However, because other auditory speech comprehension tasks may not require explicit segmentation of speech information, these patients may have relatively good speech comprehension. In contrast, aphasic individuals with damage in the posterior temporo-parietal junction may have difficulty performing an identification task because of difficulty integrating the auditory and articulatory information in the inferior parietal component of the network. Thus, both types of patients may do poorly on tasks involving explicit identification of speech sounds, but for different reasons, which is congruent with evidence from neuroimaging studies of control subjects and is consistent with the framework of models of word recognition that we have presented here.

A number of challenges remain in understanding the nature of breakdown of speech processing in aphasia. It is unlikely that an aphasic person would have damage only affecting as

specific an area as described by the functional neuroimaging studies. For example, in the frontal lobe, few people with chronic aphasia have damage only affecting Broca's area (Dronkers, 2000), and it is even less likely that such a lesion would only damage a subregion of Broca's area. Thus, understanding other functions, such as the role of verbal working memory (cf. Jacquemot & Scott, 2006), will be crucial to understanding how damage in a particular area affects language comprehension.

Acknowledgments

This research was supported in part by the National Institutes of Health under grant NIH DC R01-3378 to the senior author (SLS). Their support is gratefully acknowledged.

Note

1 In addition to psycholinguistic models, a number of computational models have been developed to perform automatic speech recognition in many cases using hidden Markov models (Deng & Erler, 1992; Krogh & Riis, 1999; Watrous, 1990). Many of these neural network models employ the same types of subphonetic (e.g., features) and phonemic representations to achieve high levels of success in recognizing isolated words. However, because these models typically have not attempted to account for a wide range of psycholinguistic data, we will not discuss them further.

References

- Allard, F., & Scott, B.L. (1975). Burst cues, transition cues, and hemispheric specialization with real speech sounds. *Quarterly Journal of Experimental Psychology*, 27, 487.
- Arbib, M.A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28, 105-124.
- Awh, E., Smith, E.E., & Jonides, J. (1995). Human rehearsal processes and the frontal lobes: PET evidence. *Annals of the New York Academy of Sciences*, 769, 97-117.
- Bandettini, P.A. (2009). What's new in neuroimaging methods? *Annals of the New York Academy of Sciences*, 1156(1), 260-293.
- Basso, A., Casati, G., & Vignolo, L. (1977). Phonemic identification defect in aphasia. *Cortex*, 13, 85-95.
- Bates, E., Wilson, S., Saygin, A.P., Dick, F., Sereno, M., Knight, R.T., & Dronkers, N.F. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience*, 6(5), 448-450.
- 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. (1988). Classical syndromes of aphasia. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology*. New York: Elsevier.
- Benson, R.R., Richardson, M., Whalen, D.H., & Lai, S. (2006). Phonetic processing areas revealed by sine-wave speech and acoustically similar non-speech. *Neuroimage*, 31, 342-353.
- Binder, J.R., Rao, S.M., Hammeke, T.A., Yetkin, F.Z., Jesmanowicz, A., Bandettini, P.A., Wong, E.C., Estkowski, L.D., Goldstein, M.D., Haughton, V.M., & Hyde, J.S. (1994). Functional magnetic resonance imaging of human auditory cortex. *Annals of Neurology*, 35, 662-672.
- Blumstein, S.E. (1998). Phonological aspects of aphasia. In M. Sarno (Ed.), *Acquired aphasia*. New York: Academic Press.
- Blumstein, S.E., Baker, E., & Goodglass, H. (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, 15, 19-30.
- Blumstein, S.E., Cooper, W., Zurif, E., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia*, 15, 371-383.
- Blumstein, S.E., Milberg, W., Brown, T., Hutchinson, A., Kurowski, A., & Burton, M.W. (2000). The mapping from sound structure to the lexicon in aphasia: Evidence from rhyme and repetition priming. *Brain and Language*, 72, 75-99.
- Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113(4), 787.

- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass*, 3(1), 19–58.
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2013). Reconciling time, space and function: A new dorsal-ventral stream model of sentence comprehension. *Brain and Language*, 125(1), 60–76.
- Bornkessel-Schlesewsky, I., Schlewsky, M., Small, S.L., & Rauschecker, J.P. (2014). Neurobiological roots of language in primate audition: Common computational properties. *Trends in Cognitive Sciences*, in revision.
- Burton, M., Blumstein, S., & Small, S.L. (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience*, 12, 679–690.
- Burton, M., & Small, S.L. (1999). An introduction to fMRI. *The Neurologist*, 5, 145–158.
- Callan, D., Callan, A., Gamez, M., Sato, M.A., Kawato, M. (2010). Premotor cortex mediates perceptual performance. *Neuroimage*, 51, 844–858.
- Caplan, D., & Waters, G.S. (1995). Aphasic disorders of syntactic comprehension and working memory capacity. *Cognitive Neuropsychology*, 12, 637–649.
- Carpenter, R., & Rutherford, D. (1973). Acoustic cue discrimination in adult aphasia. *Journal of Speech and Hearing Research*, 16, 534–544.
- Davis, M.H., Coleman, M.R., Absalom, A.R., Rodd, J.M., Johnsrude, I.S., Matta, B.F., Owen, A.M., & Menon, D.K. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 16032–16037.
- Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage*, 24, 21–33.
- Demonet, J.-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.-L., Wise, R., Rascol, A., & Frackowiak, R.S. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, 115, 1753–1768.
- Demonet, J.-F., Fiez, J. A., Paulesu, E., Petersen, S.E., & Zatorre, R.J. (1996). PET studies of phonological processing: A critical reply to Poeppel. *Brain and Language*, 55, 352–379.
- Demonet, J.-F., Price, C., Wise, R., & Frackowiak, R.S. (1994a). Differential activation of right and left posterior sylvian regions by semantic and phonological tasks: A positron emission tomography study in normal human subjects. *Neuroscience Letters*, 182, 25–28.
- Demonet, J.-F., Price, C., Wise, R., & Frackowiak, R.S. (1994b). A PET study of cognitive strategies in normal subjects during language tasks: Influence of phonetic ambiguity and sequence processing on phoneme monitoring. *Brain*, 117, 671–682.
- Deng, L., & Erler, K. (1992). Structural design of hidden Markov model speech recognizer using multivalued phonetic features: Comparison with segmental speech units. *Journal of the Acoustical Society of America*, 92, 3058–3067.
- DeWitt, I., & Rauschecker, J.P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, 109(8), E505–E514.
- Dhankhar, A., Wexler, B.E., Fulbright, R.K., Halwes, T., Blamire, A.M., & Shulman, R.G. (1997). Functional magnetic resonance imaging assessment of the human brain auditory cortex response to increasing word presentation rates. *Journal of Neurophysiology*, 77, 476–483.
- Dronkers, N.F. (1998). Symposium: The role of Broca's area in language. *Brain and Language*, 65, 71–72.
- Dronkers, N.F. (2000). The pursuit of brain-language relationships. *Brain and Language*, 71, 59–61.
- Dronkers, N.F., Wilkins, D.P., Van Valin, R.D., Redfern, B.B., & Jaeger, J.J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, 92(1), 145–177.
- Eickhoff, S.B., Heim, S., Zilles, K., & Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1896), 2399–2421.
- Elmer, S., Meyer, M., Marrama, L., & Jancke, L. (2011). Intensive language training and attention modulate the involvement of fronto-parietal regions during a nonverbal auditory discrimination task. *European Journal of Neuroscience*, 34, 165–175.
- Frauenfelder, U., & Floccia, C. (1998). The recognition of spoken word. In A. Friederici (Ed.), *Language comprehension*. New York: Springer.
- Friederici, A.D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16(5), 262–268.
- Friederici, A.D., Bahlmann, J., Heim, S., Schubotz, R.I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7), 2458–2463.

- 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.
- Galantucci, B., Fowler, C.A., Turvey, M.T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13, 361–377.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119(2), 593–609.
- Gallese, V., Fogassi, L., Fadiga, L. & Rizzolatti, G. (2002). Action representation and the inferior parietal lobule. In W. Prinz & B. Hommel (Eds.), *Attention and performance*. New York: Oxford University Press.
- Gandour, J. (2007). Neural substrates underlying the perception of linguistic prosody. *Tones and Times*, 2, 3–25.
- Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G.D. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience*, 12, 207–222.
- Gandour, J., Wong, D., & Hutchins, G. (1998). Pitch processing in the human brain is influenced by language experience. *Neuroreport*, 9, 2115–2119.
- Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110–125.
- Giraud, A.L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M.O., Preibisch, C., & Kleinschmidt, A. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex*, 14, 247–255.
- Goodale, M.A., & Milner, A.D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25.
- Gordon, J., & Baum, S. (1994). Rhyme priming in aphasia: The role of phonology in lexical access. *Brain and Language*, 47, 661–683.
- Gow, D.W. Jr., & Caplan, D. (1996). An examination of impaired acoustic-phonetic processing in aphasia. *Brain and Language*, 52, 386–407.
- Gow, D.W. Jr., & Segawa, J.A. (2009). Articulatory mediation of speech perception: A causal analysis of multi-modal imaging data. *Cognition*, 110(2), 222–236.
- Gow, D.W. Jr., Segawa, J.A., Ahlfors, S.P., & Lin, F.H. (2008). Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. *Neuroimage*, 43(3), 614–623.
- Hasson, U., Skipper, J.I., Nusbaum, H.C., & Small, S.L. (2007). Abstract coding of audiovisual speech: Beyond sensory representation. *Neuron*, 56, 1116–1126.
- Hein, G., & Knight, R.T. (2008). Superior temporal sulcus—it's my area: Or is it? *Journal of Cognitive Neuroscience*, 20(12), 2125–2136.
- Hesling, I., Dilharreguy, B., Clément, S., Bordessoules, M., & Allard, M. (2005). Cerebral mechanisms of prosodic sensory integration using low-frequency bands of connected speech. *Human Brain Mapping*, 26(3), 157–169.
- Hickok, G., Costanzo, M., Capasso, R., & Miceli, G. (2011). The role of Broca's area in speech perception: Evidence from aphasia revisited. *Brain and Language*, 119, 214–220.
- Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4, 131–138.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67–99.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.
- Hugdahl, K., Thomsen, T., Erslund, L., Rimol, L. M., & Niemi, J. (2003). The effects of attention on speech perception: An fMRI study. *Brain and Language*, 85, 37–48.
- Husain, F.T., Fromm, S.J., Pursley, R.H., Hosey, L.A., Braun, A.R., & Horwitz, B. (2006). Neural bases of categorization of simple speech and nonspeech sounds. *Human Brain Mapping*, 27, 636–651.
- Ivry, R., & Robertson, L. (1998). *The two sides of perception*. Cambridge, MA: MIT Press.
- Jacquemot, C., & Scott, S.K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences*, 10(11), 480–486.
- 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.
- Jusczyk, P. (1986). Speech perception. In K. Boff, K. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance: Cognitive processes and performance* (Vol. 2). New York: Wiley.
- Klatt, D. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, NJ: Lawrence Erlbaum.

- Krogh, A., & Riis, S. (1999). Hidden neural networks. *Neural Computation*, 11, 541–563.
- Kuhl, P.K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50, 93–107.
- Kuriki, S., Okita, Y., & Hirata, Y. (1995). Source analysis of magnetic field responses from the human auditory cortex elicited by short speech sounds. *Experimental Brain Research*, 104, 144–152.
- Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245–294.
- Leff, A.P., Schofield, T.M., Stephan, K.E., Crinion, J.T., Friston, K.J., & Price, C.J. (2008). The cortical dynamics of intelligible speech. *Journal of Neuroscience*, 28(49), 13209–13215.
- Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431.
- Liberman, A.M., Harris, K.S., Hoffman, H.S., & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358–368.
- Liberman, A.M., & Mattingly, I.G. (1985). The motor theory of speech revised. *Cognition*, 21, 1–36.
- Londei, A., D’Ausilio, A., Basso, D., Sestieri, C., Gratta, C.D., Romani, G.L., & Belardinelli, M.O. (2010). Sensory-motor brain network connectivity for speech comprehension. *Human Brain Mapping*, 31(4), 567–580.
- Luce, P.A., & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
- Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1), 71–102.
- Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101, 653–675.
- Martin, R.C., Breedin, S.D., & Damian, M.F. (1999). The relation of phoneme discrimination, lexical access, and short-term memory: A case study and interactive activation account. *Brain and Language*, 70, 437–482.
- Martin, R.C., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multiple-components view. *Neuropsychology*, 8, 506–523.
- McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.
- Metter, E.J., Hanson, W.R., Jackson, C.A., Kempler, D., Van Lancker, D., Mazziotta, J.C., & Phelps, M.E. (1990). Temporoparietal cortex in aphasia: Evidence from positron emission tomography. *Archives of Neurology*, 47(11), 1235.
- Metter, E.J., Kempler, D., Jackson, C., Hanson, W.R., Mazziotta, J.C., & Phelps, M.E. (1989). Cerebral glucose metabolism in Wernicke’s, Broca’s, and conduction aphasia. *Archives of Neurology*, 46, 27–34.
- Meyer, M., Zysset, S., von Cramon, D.Y., & Alter, K. (2005). Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex. *Cognitive Brain Research*, 24, 291–306.
- Miceli, G., Caltagirone, C., Gainotti, G., & Payer-Rigo, P. (1978). Discrimination of voice versus place contrasts in aphasia. *Brain and Language*, 6, 47–51.
- Miyake, A., Carpenter, P., & Just, M. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology*, 11, 671–717.
- Mummary, C.J., Ashburner, J., Scott, S.K., & Wise, R.J. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *Journal of the Acoustical Society of America*, 106, 449–457.
- Myers, E.B., Blumstein, S.E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, 20, 895–903.
- Narain, C., Scott, S.K., Wise, R.J., Rosen, S., Leff, A., Iversen, S.D., & Matthews, P.M. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, 13, 1362–1368.
- Nath, A.R., & Beauchamp, M.S. (2011). Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *Journal of Neuroscience*, 31(5), 1704–1714.
- Nicholls, M. (1996). Temporal processing asymmetries between the cerebral hemispheres: Evidence and implications. *Laterality*, 1, 97–137.
- Norris, D., McQueen, J.M., & Cutler, A. (2000). Merging phonetic and lexical information in phonetic decision-making. *Behavioral and Brain Sciences*, 23, 299–325.
- Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception and Psychophysics*, 57, 989–1001.

- Obleser, J., Wise, R.J., Dresner, M.A., & Scott, S.K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, 27(9), 2283–2289.
- Osnes, B., Hugdahl, K., & Specht, K. (2011). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *Neuroimage*, 54(3), 2437–2445.
- Oscar-Berman, M., Zurif, E., & Blumstein, S. (1975). Effects of unilateral brain damage on the processing of speech sounds. *Brain and Language*, 2, 345–355.
- Paulesu, E., Frith, C.D., & Frackowiak, R.J. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362, 342–345.
- Paulesu, E., Frith, U., Snowling, M., Gallagher, A., Morton, J., Frackowiak, R.S.J., & Frith, C.D. (1996). Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain*, 119, 143–157.
- Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M.A., & Raichle, M.E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, 331, 585–589.
- Pisoni, D.B. (1993). Long-term memory in speech perception: Some new findings on talker-variability, speaking rate, and perceptual learning. *Speech Communication*, 13, 109–125.
- Pitt, M.A., & Samuel, A.G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 699–725.
- Poeppel, D. (1996). A critical review of PET studies of phonological processing. *Brain and Language*, 55, 317–351.
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as ‘asymmetric sampling in time.’ *Speech Communication*, 41(1), 245–255.
- Poeppel, D., Idsardi, W.J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1071–1086.
- Poeppel, D., & Monahan, P.J. (2008). Speech perception cognitive foundations and cortical implementation. *Current Directions in Psychological Science*, 17(2), 80–85.
- Poeppel, D., Yellin, E., Phillips, C., Roberts, T.P., Rowley, H.A., Wexler, K., & Marantz, A. (1996). Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Brain Research: Cognitive Brain Research*, 4, 231–242.
- Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., & Gabrieli, J.D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage*, 10, 15–35.
- Price, C., Wise, R., Ramsey, S., Friston, K., Howard, D., & Patterson, K. (1992). Regional response differences within the human auditory cortex when listening to words. *Neuroscience Letters*, 146, 179–182.
- Price, C.J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, 62(2), 816–847.
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5), 351–360.
- Rauschecker, J.P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, 8(4), 516–521.
- Rauschecker, J.P., & Scott, S.K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.
- Rauschecker, J.P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences*, 97(22), 11800–11806.
- Ravizza, S.M., Hazeltine, E., Ruiz, S., Zhu, D.C., (2011). Left TPJ activity in verbal working memory: Implications for storage- and sensory-specific models of short term memory. *Neuroimage*, 55, 1836–1846.
- Riedel, K., & Studdert-Kennedy, M. (1985). Extending formant transitions may not improve aphasies’ perception of stop consonant place of articulation. *Brain and Language*, 24, 223–232.
- Rimol, L.M., Specht, K., Weis, S., Savoy, R., & Hugdahl, K. (2005). Processing of sub-syllabic speech units in the posterior temporal lobe: An fMRI study. *Neuroimage*, 26, 1059–1067.
- Rimol, L.M., Specht, K., & Hugdahl, K. (2006). Controlling for individual differences in fMRI brain activation to tones, syllables, and words. *Neuroimage*, 30, 554–562.
- Rinne, T., Alho, K., Alku, P., Holi, M., Sinkkonen, J., Virtanen, J., Bertrand, O., & Näätänen, R. (1999). Analysis of speech sounds is left-hemisphere predominant at 100–150 ms after sound onset. *Neuroreport*, 10, 1113–1117.
- Rizzolatti, G., & Arbib, M.A. (1998). Language within our grasp. *Trends in Neurosciences*, 21, 188–194.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3, 131–141.

- Rogalsky, C., Love, T., Driscoll, D., Anderson, S.W., & Hickok, G. (2011). Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system. *Neurocase*, *17*(2), 178–187.
- 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*(12), 1131–1136.
- Rossi, E., Schippers, M., & Keysers, C. (2011). Broca's area: Linking perception and production in language and actions. In S. Han & E. Pöppel, (Eds.), *Culture and neural frames of cognition and communication*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Samuel, A.G., & Pitt, M.A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, *48*(2), 416–434.
- Saur, D., Kreher, B.W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., & Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*, *105*(46), 18035–18040.
- Schofield, T.M., Iverson, P., Kiebel, S.J., Stephan, K.E., Kilner, J.M., Friston, K.J., Crinion, J.T., Price, C.J., & Leff, A.P. (2009). Changing meaning causes coupling changes within higher levels of the cortical hierarchy. *Proceedings of the National Academy of Sciences*, *106*(28), 11765–11770.
- Scott, S.K., & Johnsrude, I.S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26*(2), 100–107.
- Scott, S.K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*(4), 295–302.
- Scott, S.K., McGettigan, C., & Eisner, F. (2013). The neural basis of links and dissociations between speech perception and production. In J.J. Bolhuis & M. Everaert (Eds.), *Birdsong, speech, and language: Exploring the evolution of mind and brain*. Cambridge, MA: MIT Press.
- Skipper, J.I., Nusbaum, H.C., & Small, S.L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage*, *25*, 76–89.
- Skipper, J.I., Nusbaum, H.C., & Small, S.L. (2006). Lending a helping hand to hearing: Another motor theory of speech perception. In M.A. Arbib (Ed.), *Action to language via the mirror neuron system*. Cambridge, MA: Cambridge University Press.
- Skipper, J.I., van Wassenhove, V., Nusbaum, H.C., & Small, S.L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, *17*, 2387–2399.
- Sumby, W.H., & Pollack, I. (1954). Visual contribution of speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212–215.
- Tremblay, P., & Small, S.L. (2011a). From language comprehension to action understanding and back again. *Cerebral Cortex*, *21*(5), 1166–1177.
- Tremblay, P., & Small, S.L. (2011b). On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *Neuroimage*, *57*, 1561–1571.
- Ungerleider, L., & Mishkin, M. (1982). Two cortical visual systems. In M.A. Goodale, D.J. Ingle, & R.J. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.
- Warren, R.M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392–393.
- Watrous, R.L. (1990). Phoneme discrimination using connectionist networks. *Journal of the Acoustical Society of America*, *87*, 1753–1772.
- Wilson, S.M., Dronkers, N.F., Ogar, J.M., Jang, J., Growdon, M.E., Agosta, F., Henry, M.L., Miller, B.L., & Gorno-Tempini, M.L. (2010). Neural correlates of syntactic processing in the nonfluent variant of primary progressive aphasia. *Journal of Neuroscience*, *30*(50), 16845–16854.
- Wise, R., Chollet, F., Hadar, U., Friston, K., Hoffner, E., & Frackowiak, E. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain*, *114*, 1803–1817.
- Zaehle, T., Wustenberg, T., Meyer, M., & Jancke, L. (2004). Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. *European Journal of Neuroscience*, *20*, 2447–2456.
- Zatorre, R. (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*. 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., Meyer, E., Gjedde, A., & Evans, A. (1996). PET studies of phonetic processing of speech: Review, replication, and reanalysis. *Cerebral Cortex*, *6*, 21–30.
- Zekveld, A.A., Heslenfeld, D.J., Festen, J.M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *Neuroimage*, *32*, 1826–1836.
- Zevin, J.D., Yang, J., Skipper, J.I., McCandliss, B.D., (2010). Domain general change detection accounts for “dishabituation” effects in temporal–parietal regions in functional magnetic resonance imaging studies of speech perception. *Journal of Neuroscience*, *30*, 1110–1117.

Taylor & Francis
Not for distribution