

CHAPTER 26 Auditory Pathway Representations of Speech Sounds in Humans

Daniel A. Abrams and Nina Kraus

INTRODUCTION

An essential function of the central auditory system is the neural encoding of speech sounds. The ability of the brain to translate the acoustic events in the speech signal into meaningful linguistic constructs relies in part on the representation of the acoustic structure of speech by the central nervous system. Consequently, an understanding of how the nervous system accomplishes this task would provide important insight into the basis of language perception and cognitive function.

One of the challenges faced by researchers interested in this subject is that speech is a complex acoustic signal that is rich in both spectral and temporal features. In everyday listening situations, the abundance of acoustic cues in the speech signal provides enormous perceptual benefits to listeners. For example, it has been shown that listeners are able to shift their attention between different acoustic cues when perceiving speech from different talkers to compensate for the inherent variability in the acoustic properties of speech between individuals (Nusbaum and Morin, 1992).

There are two basic approaches that researchers have adopted for conducting experiments on speech perception and underlying physiology. One approach uses “simple” acoustic stimuli, such as tones and clicks, as a means to control for the complexity of the speech signal. While simple stimuli enable researchers to reduce the acoustics of speech to its most basic elements, the auditory system is nonlinear (Sachs and Young, 1979; Sachs et al., 1983; Rauschecker, 1997; Nagarajan et al., 2002), and therefore, responses to

simple stimuli generally do not accurately predict responses to actual speech sounds. A second approach uses speech and speech-like stimuli (Song et al., 2006). There are many advantages to this approach. First, these stimuli are more ecologically valid than simple stimuli. Second, a complete description of how the auditory system responds to speech can only be obtained by using speech stimuli, given the nonlinearity of the auditory system. Third, long-term exposure to speech sounds and the subsequent use of these speech sounds in linguistic contexts induces plastic changes in the auditory pathway, which may alter neural representation of speech in a manner that cannot be predicted by simple stimuli. Fourth, when speech stimuli are chosen carefully, the acoustic properties of the signal can still be well controlled.

This chapter reviews the literature that has begun to elucidate how the human auditory system encodes acoustic features of speech. This chapter is organized into five sections, with each section describing what is currently known about how the brain represents a particular acoustic feature present in speech (Table 26.1). These acoustic features of speech were chosen because of their essential roles in normal speech perception. Each section contains a description of the acoustic feature and an elaboration of its relevance to speech perception, followed by a review and assessment of the data for that acoustic feature. An important consideration is that the acoustic features described in this chapter are not mutually exclusive. For example, one section of this chapter describes the neural encoding of “periodicity,” which refers to acoustic events that occur at regular time intervals. Many features in the speech signal are periodic; however,

Tab. 1

TABLE 26.1 Title

Acoustic features in speech	Feature's role in the speech signal	Brainstem measure	Cortical measure
Formant structure	Ubiquitous in vowels, approximants, and nasals; essential for vowel perception	Frequency-following response	N100m source location; STS activity (fMRI)
Periodicity	Temporal cue for fundamental frequency and low formant frequencies (50–500 Hz)	Frequency-following response	N100m source location and amplitude; nonprimary auditory cortex activity patterns (fMRI)
Frequency transitions	Consonant identification; signal the presence of diphthongs and glides; linguistic pitch	Frequency-following response	Left vs. right STG activity (fMRI)
Acoustic onsets	Phoneme identification	ABR onset complex	N100m source location; N100 latency
Speech envelope	Syllable and low-frequency (<50 Hz) patterns in speech	N/A	N100m phase-locking

STS, superior temporal sulcus; fMRI, functional magnetic resonance imaging; STG, superior temporal gyrus; ABR, auditory brainstem response; N/A, not applicable.

AUTHOR:
 Please provide
 title for Table
 26.1.

describing the neurophysiologic encoding of all of the periodic features that are processed simultaneously in the speech stimulus in a study of the auditory system would be experimentally unwieldy. Consequently, for the sake of simplicity and to reflect the manner in which they have been investigated in the auditory neuroscience literature, some related acoustic features will be discussed in separate sections. Efforts will be made throughout the chapter to identify when there is overlap among acoustic features.

THE SIGNAL: BASIC SPEECH ACOUSTICS

The speech signal can be described according to a number of basic physical attributes (Johnson, 1997). An understanding of these acoustic attributes is essential to any discussion of how the auditory system encodes speech. The linguistic roles of these acoustic features are described separately within each section of the chapter.

Fundamental frequency. The fundamental frequency is a low-frequency component of speech that results from the periodic beating of the vocal folds. In Figure 26.1A, the frequency content of the naturally produced speech sentence “The young boy left home” is plotted as a function of time; greater amounts of energy at a given frequency are represented with red lines, while smaller amounts of energy are depicted in blue. The fundamental frequency can be seen as the horizontal band of energy in Figure 26.1A that is closest to the x-axis (i.e., lowest in frequency). The fundamental frequency is notated F_0 and provides the perceived pitch of an individual’s voice.

Harmonic structure. An acoustic phenomenon that is related to the fundamental frequency of speech is known as the harmonic structure of speech. Harmonics, which are integer multiples of the fundamental frequency, are present in ongoing speech. The harmonic structure of speech is displayed in Figure 26.1A as the regularly spaced horizontal bands of energy seen throughout the sentence.

Formant structure. Another essential acoustic feature of speech is the formant structure. Formant structure describes a series of discrete peaks in the frequency spectrum of speech that are the result of an interaction between the frequency of vibration of the vocal folds and the resonances within a speaker’s vocal tract. The frequency of these peaks, as well as the relative frequency between peaks, varies for different speech sounds. The formant structure of speech interacts with the harmonic structure of speech; the harmonic structure is represented by integer multiples of the fundamental frequency, and harmonics that are close to a resonant frequency of the vocal tract are formants. In Figure 26.1, the formant structure of speech is represented by the series of horizontal and occasionally diagonal red lines that run through most of the speech utterance. The word “left” has been enlarged in Figure 26.1B to better illustrate this phenomenon. The broad and dark red patches seen in this figure represent the peaks in the frequency spectrum of speech that are the result of an interaction between the frequency of vibration of the vocal folds and the resonances of a speaker’s vocal tract. The frequency of these peaks, as well as the relative frequency between peaks, varies for different speech sounds within the sentence. The lowest frequency formant is known as the first formant and is notated F_1 , while subsequent formants are notated F_2 , F_3 , etc.

Fig. 1

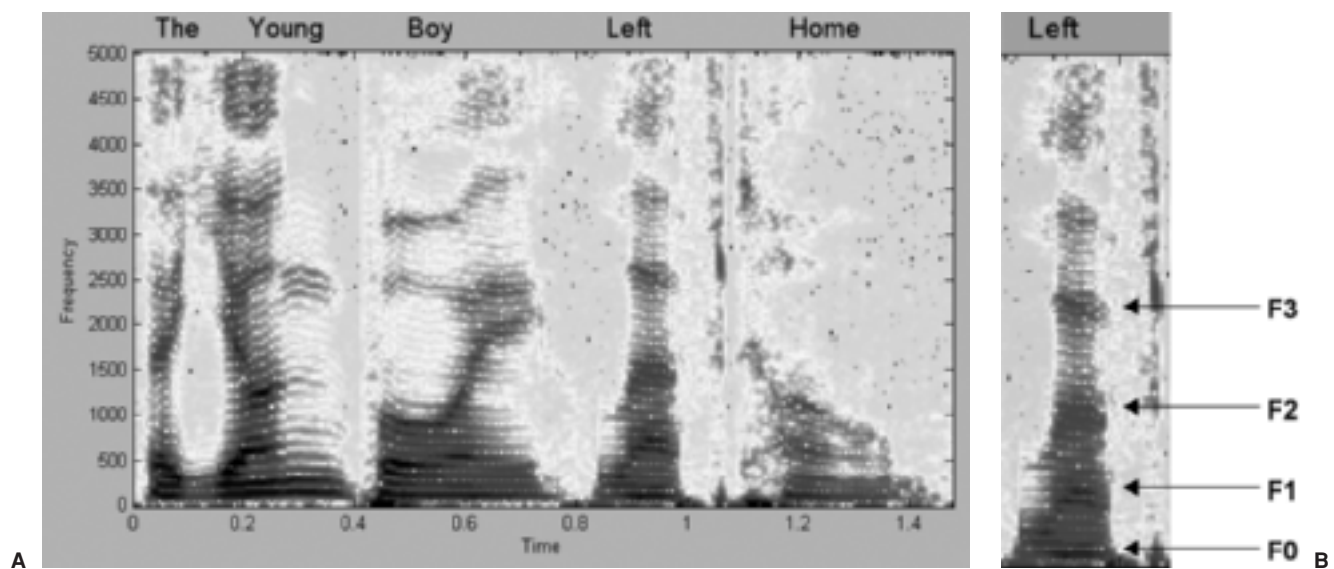


FIGURE 26.1 Spectrogram for the naturally produced speech sentence “The young boy left home.” **(A)** The complete sentence; **(B)** the word “left” is enlarged to illustrate the frequency structure; the fundamental frequency (F0) and formants (F1-F3) are represented in the spectrogram by broad red lines of energy.

THE MEASURES OF BRAIN ACTIVITY

We begin by describing the neurophysiologic measures that have been used to probe auditory responses to speech and speech-like stimuli; comprehensive descriptions of these measures can be found elsewhere (Sato, 1990; Hall, 1992; Jezzard et al., 2001) as well as in various chapters in this text. Historically, the basic research on the neurophysiology of speech perception has borrowed a number of clinical tools to assess auditory system function.

Brainstem Responses

The auditory brainstem response (ABR) consists of small voltages originating from auditory structures in the brainstem in response to sound. Although these responses do not pinpoint the specific origin of auditory activity among the auditory brainstem nuclei, the great strength of the ABR (and auditory potentials in general) is that they precisely reflect the time-course of neural activity at the microsecond level. The ABR is typically measured with a single active electrode referenced to the earlobe or nose. Clinical evaluations using the ABR typically use brief acoustic stimuli, such as clicks and tones, to elicit brainstem activity. The ABR is unique among the auditory evoked potentials (AEPs) because of the remarkable reliability of this response, both within and across subjects. In the clinic, the ABR is used to assess the integrity of the auditory periphery and lower brainstem (Hall, 1992). The response consists of a number of

peaks, with wave V being the most clinically reliable. Deviations on the order of microseconds are deemed “abnormal” in the clinic and are associated with some form of peripheral hearing damage or with retrocochlear pathologies. Research using the ABR to probe acoustic processing of speech uses similar recording procedures but different acoustic stimuli.

Cortical Responses

CORTICAL EVOKED POTENTIALS AND FIELDS

Cortical evoked responses are used as a research tool to probe auditory function in normal and clinical populations. Cortical evoked potentials are small voltages originating from auditory structures in the cortex in response to sound. These potentials are typically measured with multiple electrodes, often referenced to a “common reference,” which is the average response measured across all electrodes. Cortical evoked “fields” are the magnetic counterpart to cortical evoked potentials; however, instead of measuring voltage across the scalp, the magnetic fields produced by brain activity are measured. Electroencephalography (EEG) is the technique by which evoked potentials are measured, and magnetoencephalography (MEG) is the technique by which evoked fields are measured. Similar to the ABR, the strength of assessing cortical evoked potentials and fields is that they provide detailed information about the time-course of activation and how sound is encoded by temporal response properties in large populations of auditory neurons, although this technique is limited in its spatial resolution.

614 Section III ■ Special Populations

Due to large inter- and intrasubject variability in cortical responses, they are not generally used clinically. Results from these two methodologies are generally compatible, despite some differences in the neural generators that contribute to each of these responses. Studies using both EEG and MEG are described interchangeably throughout this chapter despite the subtle differences between the measures. The nomenclature of waveform peaks is similar for EEG and MEG and typically involves an N or P, depicting a negative or positive deflection, followed by a number indicating the approximate latency of the peak. Finally, the letter “m” follows the latency for MEG results. For example, N100 and N100m are the labels for a negative deflection at 100 ms as measured by EEG and MEG, respectively.

Functional Imaging

Functional imaging of the auditory system is another often-used technique to quantify auditory activity in the brain. The technology that is used to measure these responses, as well as the results they yield, is considerably different from the previously described techniques. The primary difference is that functional imaging is an indirect measure of neural activity; that is, instead of measuring voltages or fields resulting from activity in auditory neurons, functional imaging measures hemodynamics, a term used to describe changes in metabolism as a result of changes in brain activity. The data produced by these measures produce a three-dimensional map of activity within the brain as a result of a given stimulus. The strong correlation between actual neural activity and blood flow to the same areas of the brain (Smith et al., 2002a) has made functional imaging a valuable investigative tool to measure auditory activity in the brain. The two methods of functional imaging described here are functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). The difference between these two techniques is that fMRI measures natural levels of oxygen in the brain because oxygen is consumed by neurons when they become active. PET, however, requires the injection of a radioactive isotope into a subject. The isotope emits positrons, which can be detected by a scanner, as it circulates in the subject’s bloodstream. Increases in neural activity draw more blood and, consequently, more of the radioactive isotope to a given region of the brain. The main advantage that functional imaging offers relative to evoked potentials and evoked fields is that it provides extremely accurate spatial information regarding the origin of neural activity in the brain. A disadvantage is the poor resolution in the temporal domain; neural activity is often integrated over the course of seconds, which is considered extremely slow given that speech tokens are as brief as 30 ms. Although recent work using functional imaging has begun describing activity in subcortical regions, the work described here will only cover studies of the temporal cortex.

ACOUSTIC FEATURES OF SPEECH

Periodicity

DEFINITION AND ROLE IN THE PERCEPTION OF SPEECH

Periodicity refers to regular temporal fluctuations in the speech signal between 50 and 500 Hz (Rosen, 1992). Important aspects of the speech signal that contain periodic acoustic information include the fundamental frequency and low-frequency components of the formant structure (note that encoding of the formant structure of speech is covered in a later section). The acoustic information provided by periodicity conveys both phonetic information as well as prosodic cues, such as intonation and stress, in the speech signal. As stated in Rosen’s paper, this category of temporal information represents both the periodic features in speech and the distinction between the periodic and aperiodic portions of the signal, which fluctuate at much faster rates.

This section will review studies describing the neural representation of relatively stationary periodic components in the speech signal, most notably the fundamental frequency. An understanding of the mechanism for encoding a simple periodic feature of the speech signal, the F0, will facilitate descriptions of complex periodic features of the speech signal, such as the formant structure and frequency modulations.

PHYSIOLOGIC REPRESENTATION OF THE PERIODICITY IN THE HUMAN BRAIN

Auditory Brainstem

The short-latency frequency-following response (FFR) is an electrophysiologic measure of phase-locked neural activity originating from brainstem nuclei that represents responses to periodic acoustic stimuli up to approximately 1,000 Hz (Smith et al., 1975; Stillman et al., 1978; Gardi et al., 1979; Galbraith et al., 2000). Based on the frequency range that can be measured with the FFR, a representation of the fundamental frequency can be measured using this methodology (Cunningham et al., 2001; King et al., 2002; Krishnan et al., 2004; 2005; Riecke et al., 2004; 2005; Wible et al., 2004; Johnson et al., 2005), as well as the F1 in some instances (encoding of F1 is discussed in detail in the Formant Structure section).

A number of studies have shown that F0 is represented within the steady-state portion of the brainstem response (i.e., FFR) according to a series of negative peaks that are temporally spaced in correspondence to the wavelength of the fundamental frequency. An example of F0 representation in the FFR can be seen in Figure 26.2, which shows the waveform of the speech stimulus /da/ (top), an experimental stimulus that has been studied in great detail, as well as the brainstem response to this speech sound (bottom). A cursory inspection of this figure shows that the primary periodic features of the speech waveform provided by the F0

Fig. 2

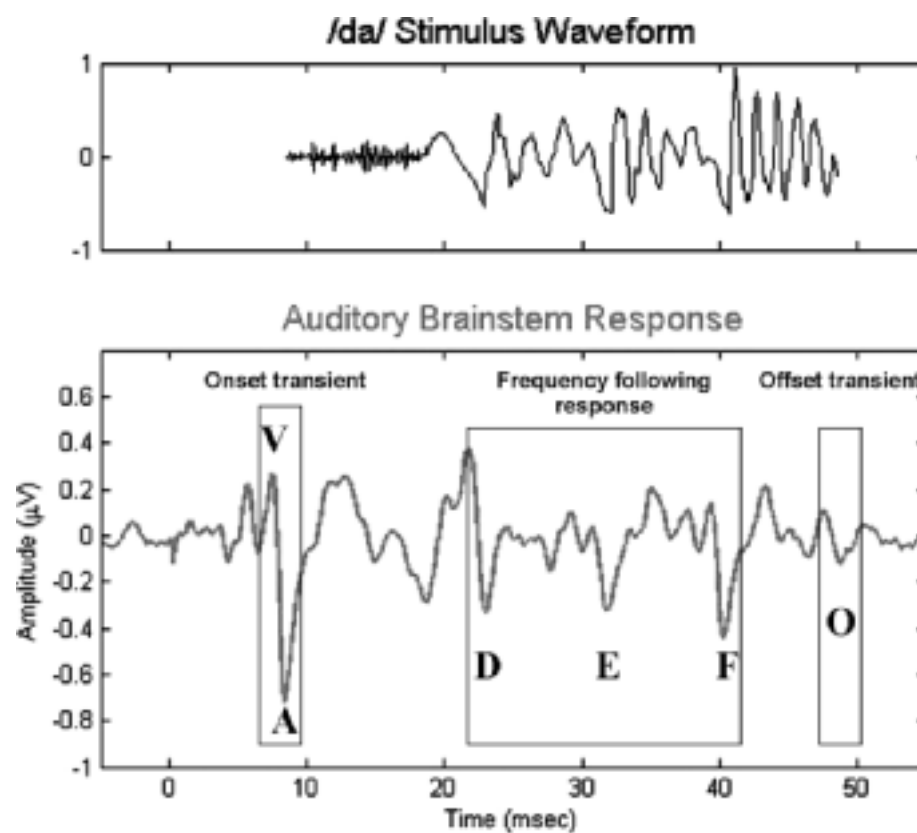


FIGURE 26.2 Acoustic waveform of the synthesized speech stimulus /da/ (above) and grand average auditory brainstem responses to /da/ (below). The stimulus has been moved forward in time to the latency of onset responses (peak V) to enable direct comparisons with brainstem responses. Peaks V and A reflect the onset of the speech sound, and peak O reflects stimulus offset. Peaks D, E, and F represent a phase-locked representation to the fundamental frequency of the speech stimulus, and the peaks between D, E, and F occur at the F1 frequency.

are clearly represented in peaks D, E, and F of the FFR brainstem response. Importantly, it has been shown that the FFR is highly sensitive to F0 frequency; this aspect of the brainstem response accurately “tracks” modulations in frequency (Krishnan et al., 2004), a topic that is discussed in depth in the Frequency Transitions section of this chapter.

A hypothesis regarding the brainstem’s encoding of different aspects of the speech signal has been proposed in a recent paper (Kraus and Nicol, 2005). Specifically, it is proposed that the source (referring to vocal fold vibration) and filter aspects (vocal musculature in the production of speech) of a speech signal show dissociation in their acoustic representation in the auditory brainstem. The source portion of the brainstem’s response to speech is the representation of the F0, while the filter refers to all other features, including speech onset, offset, and the representation of formant frequencies. For example, it has been demonstrated that brainstem responses are correlated within source and filter classes but are not correlated between classes (Russo et al., 2004). Moreover, in a study of children with language-learning disabilities whose behavioral deficits may be attributable to central auditory processing disorders, it has been shown that source representation in the auditory brainstem is normal, while filter class representation is impaired (Cunningham et al., 2001; King et al., 2002; Hayes et al., 2003; Wible et al., 2004; 2005). These data suggest that the acoustic representations of source and filter aspects of a given speech signal are differentially processed and provide evidence for neural

specialization at the level of the brainstem. Additionally, it is proposed that this scheme may constitute brainstem origins for cortical “what,” “where” pathways (Kraus and Nicol, 2005).

Cortex

It has been shown that neurons in the auditory cortex respond robustly with time-locked responses to slow rates of stimulation (< ~25 Hz) and generally do not phase-lock to frequencies greater than approximately 100 Hz (Creutzfeldt et al., 1980; Eggermont, 1991; Steinschneider et al., 1998; Lu et al., 2001). Therefore, cortical phase-locking to the fundamental frequency of speech, which is greater than 100 Hz, is poor, and it is generally thought that the brainstem’s phase-locked (i.e., linear) representation of F0 is transformed at the level of cortex to a more abstract representation. For example, it has been shown that cortical neurons produce sustained, nonsynchronized discharges throughout a high-frequency (>50 Hz) stimulus (Lu et al., 2001), resulting in a more abstract representation of the stimulus frequency compared to time-locked neural activation.

An important aspect of F0 perception is that listeners native to a particular language are able to perceive a given speech sound as invariant regardless of the speaker’s F0, which varies considerably among men (F0 = ~100 Hz), women (F0 = ~200 Hz), and children (F0 = up to 400 Hz). For example, the speech sound “dog” is categorized by a listener to mean the exact same thing regardless of whether

616 Section III ■ Special Populations

an adult or a child produces the vocalization, even though there is a considerable difference in the acoustic properties of the adult's and child's vocalization with respect to the fundamental frequency. To address how auditory cortical responses reflect relatively large variations in F0 between speakers, N100m cortical responses were measured with MEG for a set of Finnish vowel and vowel-like stimuli that varied in F0, while keeping all other formant information (F1-F4) constant (Makela et al., 2002). Results indicated that N100m responses were extremely similar in spatial activation pattern and amplitude for all vowel and vowel-like stimuli, irrespective of the F0. This is a particularly intriguing finding given that N100m responses differed when 100-, 200-, and 400-Hz puretone stimuli were presented to the same subjects in a control condition. The similarity of the speech-evoked brain responses, which were independent of the F0 frequency, suggests that variances in F0 may be filtered out of the neural representation by the time it reaches the cortex. The authors suggest that the insensitivity of cortical responses to variations in the F0 may facilitate the semantic categorization of the speech sound. In other words, since the F0 does not provide essential acoustic information relevant to the semantic meaning of the speech sound, it may be the case that the cortex does not respond to this aspect of the stimulus in favor of other acoustic features that are essential for decoding word meaning.

In summary, periodicity of the fundamental frequency is robustly represented in the FFR of the ABR. Moreover, the representation of the fundamental frequency is normal in learning-disabled children despite the abnormal representations of speech-sound onset and first formant frequency. This disparity in the learning-disabled auditory system provides evidence that different features of speech sounds may be served by different neural mechanisms and/or populations. In the cortex, MEG results show that cortical responses are relatively insensitive to changes in the fundamental frequency of speech sounds, suggesting that the differences between F0s between speakers are filtered out by the time they reach the level of auditory cortex.

Formant Structure

ROLE IN THE PERCEPTION OF SPEECH

Formant structure describes a series of discrete peaks in the frequency spectrum of speech that are the result of an interaction between the frequency of vibration of the vocal folds and the resonances within a speaker's vocal tract (see introduction of this chapter for a more complete acoustic description of the formant structure). The formant structure is a dominant acoustic feature of sonorants, a class of speech sounds that includes vowels, approximants, and nasals. The formant structure has a special role in the perception of vowels in that formant frequencies, particularly the relationship between F1 and F2 (Peterson and Barney, 1952), are the primary phonetic determinants of vowels. For example, the essential acoustic difference between /u/ and /i/ is a positive

shift in F2 frequency (Peterson and Barney, 1952). Due to the special role of formants for vowel perception, much of the research regarding the formant structure of speech uses vowel stimuli.

PHYSIOLOGIC REPRESENTATION OF FORMANT STRUCTURE IN THE HUMAN BRAIN

Auditory Brainstem

The question of how the human auditory brainstem represents important components of the formant structure was addressed in a study by Krishnan (2002). In this study, brainstem (FFR) responses to three steady-state vowels were measured, and the spectral content of the responses were compared to that of the vowel stimuli. All three of the stimuli had approximately the same fundamental frequency; however, the first two formant frequencies were different in each of the vowel stimuli. Results indicate that at higher stimulus intensities, the brainstem FFR accurately represents F1 and F2; however, the representation of F1 has an increased representation relative to F2. The author indicates the similarity between this finding and a similar result in a classic study of vowel representation in the auditory nerve of anesthetized cats (Sachs and Young, 1979) in which the predominance of the representation to F1 was also demonstrated. These data provide evidence that phase-locking serves as a mechanism for encoding critical components of the formant structure not only in the auditory nerve, but also in the auditory brainstem.

Auditory Cortex

A number of studies have described the representation of formant structure in the human cortex as a means of investigating whether a cortical map of phonemes, termed the "phonemotopic" map, exists in the human brain. Specifically, researchers want to know if the phonemotopic map is independent of the tonotopic map or, alternatively, whether phonemes are more simply represented according to their frequency content along the tonotopic gradient in auditory cortex. To this end, investigators have measured cortical responses to vowel stimuli, a class of speech sounds that differ acoustically from one another according to the distribution of F1-F2 formant frequencies. Vowel stimuli also offer the advantage of exhibiting no temporal structure beyond the periodicity of the formants.

The method that has been used to investigate the relationship between the tonotopic map in human auditory cortex and the representation of formant structure has been to compare cortical source locations for tones and specific speech sounds with similar frequency components. For example, in one study (Diesch and Luce, 1997), N100m source location was measured in response to separately presented 600-Hz and 2,100-Hz puretones, as well as a two-tone composite signal comprising the component puretones (i.e., simultaneous presentation of the 600-Hz and 2,100-Hz puretones). These responses were compared to isolated formants,

defined as the first and second formant frequencies of a vowel stimulus, complete with their harmonic structure, separated from the rest of the frequency components of the stimulus (i.e., F0, higher formant frequencies). These isolated formants had the same frequency as the tonal stimuli (i.e., 600 Hz and 2,100 Hz). Finally, a two-formant composite signal, which constituted a vowel, was also presented. Results indicated that the N100m source in response to the vowel stimulus was different in location from that predicted by both the puretone responses and by the superposition of responses to the component single-formant stimuli. These data indicate that formant structure is spatially represented in human cortex differently than the linear sum of responses to the component formant stimuli and suggest that formant structure is represented orthogonal to the tonotopic map. The authors of this work hypothesize that the different spatial representation of the vowel stimuli reflects the additional acoustic components of the vowel stimuli, including the harmonic and formant structures. The authors of this work refrain from stating a potentially more intriguing conclusion; that is, does the spatial representation of the vowel stimuli in some way reflect the behavioral experience of the subjects with these speech sounds? For example, it is possible that a larger, or different, population of cortical neurons is recruited for sounds that are familiar or have significant ecologic importance relative to the population recruited for puretones or single formant frequencies and that the source location for the vowels reflects this phenomenon.

Additional studies have attempted to better describe the acoustic representation of vowels in the human brain. In one study, Obleser et al. (2003) addressed the neurophysiology underlying a classic study of speech acoustics in which it was shown that the distinction of vowels is largely carried by the frequency relationship of F1 and F2 (Peterson and Barney, 1952). To this end, cortical source locations were measured in response to German vowels that differ naturally in their F1-F2 relationships. Results indicated that the location of the N100m source reflects the frequency relationship of the F1-F2 formant components. This finding was replicated in a second study using 450 natural speech exemplars of three Russian vowels; again, the spectral distance between F1 and F2 was reflected in the dipole location of N100m responses (Shestakova et al., 2004). In both studies, the authors suggested that cortical sensitivity to F1-F2 differences can be explained by inhibitory response patterns in the auditory cortex; the closer the F1 frequency is to F2, the greater the reciprocal neural inhibition, which, in turn, influences the location of the dipole source as measured by MEG (Obleser et al., 2003).

While these studies provide evidence that the cortex represents the formant structure of vowels in a manner that is (1) unrelated to the tonotopic map and (2) organized according to the perceptually essential formant frequencies, these findings require a number of caveats. First, the source locations described in these studies represent the center of gravity, as a single point in three-dimensional space in the

cortex, of the neural contributors to a given N100m response (Naatanen and Picton, 1987). Because it is known that the N100 response has as many as six separate cortical generators, the N100m sources for even a simple cortical map (i.e., the tonotopic map), let alone a complex map such as the putative phonemotopic map, represent at least a partial abstraction of the underlying anatomy and should not be viewed as an exact representation of well-described auditory maps in animal models (Schreiner, 1998). This is particularly relevant given that the clear tonotopic gradient in the auditory cortex is no longer apparent when puretone stimuli are presented above 50 dB sound pressure level (SPL) (Schreiner, 1998), such as the levels used in the MEG experiments described in this section. In addition, it has not yet been definitively shown that the neural representations of phonemes described in these studies truly constitute a “phonemotopic” map. The presence of a phonemotopic map suggests behavioral relevance of phoneme stimuli beyond their acoustic attributes. None of the studies described here have tested whether cortical responses to the F1-F2 components for nonnative vowel sounds show similar sensitivity as native phonemes. Despite these limitations, these studies provide consistent evidence that a perceptually critical aspect of the formant structure of vowels, the F1-F2 relationship, is represented in a spatial map in the auditory cortex as early as approximately 100 ms after stimulus onset.

Another line of evidence has used functional imaging to show the particular regions of the temporal cortex that are sensitive to the formant structure of speech sounds relative to other natural and vocally generated (e.g., laughs, coughs) sounds (Belin et al., 2000). Cortical responses to natural vocal stimuli were compared to vocal stimuli in which the formant structure of speech was replaced by white noise and scrambled vocal sounds. All stimuli were matched for overall RMS energy. In both of these experimental conditions, the altered spectral information was modulated by the original amplitude envelope of the speech signal. Results from this experiment indicated that all stimuli activated regions along the superior temporal sulcus (STS), a cortical region consisting of unimodal auditory and multimodal areas that is hypothesized to be a critical speech processing center subsequent to more rudimentary acoustic processing in the superior temporal gyrus. However, responses to the natural vocal stimuli were significantly larger and more widespread throughout the STS, particularly in the right hemisphere, than for the spectrally manipulated vocal stimuli. These data indicate that the formant structure of speech deeply affects activity patterns in the STS, a speech-selective region of temporal cortex, even when the temporal components of the signals are held constant. Moreover, these data suggest a right hemisphere bias for processing the formant structure, which supports the more general hypothesis that the right hemisphere is dominant for resolving spectral components in acoustic signals (Zatorre and Belin, 2001; Zatorre et al., 2002).

An interesting consideration is how cortical asymmetries in response to the acoustic features of speech relate to

AUTHOR:
Please define
RMS.

618 Section III ■ Special Populations

well-established cerebral asymmetries for higher order language processing, such as phonemic and semantic processing (Geschwind and Galanter, 1985; Binder et al., 1997; Hickok and Poeppel, 2004), which are strongly lateralized to the left hemisphere. While a direct link between these forms of asymmetry has not been established, a plausible scenario is that the acoustic-level asymmetries precede and serve as the input to phonemic and semantic processing in left hemisphere language regions. If this is the case, it remains to be seen what physiologic advantage a right hemisphere preference for formant structure processing (Belin et al., 2000) might offer given that phonemic and semantic processing of speech stimuli takes place in the opposite hemisphere, thereby requiring transmission through the corpus callosum. Future studies investigating acoustic-level asymmetries and their interface with higher order language asymmetries would provide essential information regarding the functional neuroanatomy of speech perception.

In summary, the brainstem encodes lower formant frequencies, which are critical to vowel perception, with phase-locked responses. Converging evidence indicates that the cortex encodes a perceptually essential aspect of the formant structure of speech. Specifically, the F1-F2 relationship is spatially mapped in the cortex at approximately 100 ms after stimulus onset as measured by N100m source location. In addition, functional imaging data provide evidence that the STS, a nonprimary region of temporal cortex, is more responsive to speech stimuli that contain formant structure than speech in which the formant structure has been replaced with other sounds. Together, these results suggest that both primary and nonprimary regions of temporal cortex are sensitive to aspects of the formant structure that are essential for normal perception.

Frequency Transitions

ACOUSTIC DESCRIPTION AND ROLE IN THE PERCEPTION OF SPEECH

Frequency transitions of the fundamental and formant frequencies permeate ongoing speech. In English, modulation of the fundamental frequency typically does not provide segmental cues; rather, it provides suprasegmental cues such as the intent (e.g., question or statement) and emotional state of the speaker. In other languages, such as Mandarin and Thai, modulations to the fundamental frequency provide phonetic cues. Formant transitions, on the other hand, are critical to speech perception of English in that they serve as a cue for consonant identification and signal the presence of diphthongs and glides (Lehiste and Peterson, 1961). Moreover, formant transitions also have been shown to play a role in vowel identification (Nearey and Assmann, 1986). The movements of formant frequencies can be distilled to three basic forms that occur during an ongoing sequence of phonemes (taken from Lehiste and Peterson [1961]): (1) the movement of a formant from the initiation of the consonant until the beginning of the vowel in a consonant-vowel combi-

nation, (2) the movement of a formant from one vowel to another vowel (i.e., in a diphthong), and (3) formant movement from a vowel until vowel termination for a vowel-consonant combination. The frequency modulations that occur during formant transitions can occur at relatively fast rates (~40 ms) while spanning large frequency ranges (>2,000 Hz in F2 transitions).

PHYSIOLOGIC REPRESENTATION OF FREQUENCY TRANSITIONS IN THE HUMAN BRAIN

Auditory Brainstem

The short-latency FFR is able to “track,” or follow, frequency changes in speech. This phenomenon was demonstrated in a study of FFR tracking of the fundamental frequency (F0) in Mandarin speech sounds (Krishnan et al., 2004). In this study, FFR to four different tonal permutations of the Mandarin word “yi” were measured in a group of native Mandarin speakers. Specifically, synthetic stimuli consisted of “yi” pronounced with (1) a flat F0 contour, (2) a rising F0 contour, (3) a falling F0 contour, and (4) a concave F0 contour that fell and then rose in frequency. In Mandarin, which is a “tonal” language, these four stimuli are different words; the F0 contour provides the only acoustic cue to differentiate them. Results indicated that FFR represented the fundamental frequency modulations for all of the stimulus conditions, irrespective of the form of the frequency contour. These data indicate that the FFR represents phase-locked activity in the brainstem for rapidly changing frequency components in speech, an essential acoustic cue for consonant identification.

A similar methodology was used in another study by Krishnan et al. (2005) to investigate the role of language experience on auditory brainstem encoding of pitch. FFRs to the “yi” stimuli described earlier were measured in native Mandarin speakers as well as native speakers of American English, to whom the stimuli bear no linguistic value. Results from this study indicate greater FFR pitch strength and pitch tracking in the Chinese subjects compared to the native English speakers across all four of the Mandarin tones. The FFR of the Chinese subjects also indicated increased harmonic representation of the fundamental frequency (i.e., larger neural representation of the harmonic content of the F0) compared to the English speakers. These data indicate that responses from the auditory brainstem reflect the behavioral experience of a listener by enhancing the neural representation of linguistically relevant acoustic features.

A hypothesis proposed by Ahissar and Hochstein (2004) may explain how experience engenders plasticity at low levels of sensory systems. Their “reverse hierarchy” theory proposes that when a naïve subject attempts to perform a perceptual task, the performance on that task is governed by the “top” of a sensory hierarchy. As this “top” level of the system masters performance on the task, over time, lower levels of the system are modified and refined to provide more precise encoding of sensory information. This can be thought of as an efferent pathway-mediated tuning of afferent sensory input. While the reverse hierarchy theory does not

explicitly discuss plasticity of the brain, this theory could account for the findings of Krishnan. Specifically, due to the importance of extracting lexical information present in pitch contours, native Mandarin speakers are “expert” at encoding this acoustic feature, which is accomplished, at least in part, by extreme precision and robustness of sensory encoding in low levels of the auditory system such as the brainstem. Native English speakers, who are not required to extract lexical meaning from pitch contours, are relative novices at this form of pitch tracking, and consequently, their brainstems have not acquired this level of modification.

An interesting question that was not addressed in this study but was proposed as a discussion item is whether native Mandarin speakers are better than English speakers at pitch tracking the F0 exclusively for familiar speech sounds or whether Mandarin speakers’ superior performance would extend to all periodic acoustic signals, including nonnative speech sounds. This question would address whether a lifetime of experience using F0 to extract linguistic meaning generally improves the auditory system’s ability to track all types of pitches or, alternatively, whether this phenomenon is exclusive to pitches present in familiar speech sounds. Data from our lab suggest that another form of long-term auditory experience, musicianship, contributes to enhanced neural encoding of speech sounds in the auditory brainstem relative to nonmusicians (Wong et al., 2004). This finding provides evidence that expertise associated with one type of acoustic signal (i.e., music) provides a general augmentation of the auditory system that is manifested in brain responses to another type of acoustic signal (i.e., speech) and indicates that auditory experience can modify basic sensory encoding.

Auditory Cortex

Similar to Krishnan’s work involving the brainstem, multiple studies have investigated cortical processing of F0 pitch contours and its relationship to language experience (Gandour et al., 1998; Klein et al., 2001; Wang et al., 2004). Most convincing of these studies is that by Wong et al. (2004). In this study, native Mandarin and native English speakers underwent PET scanning during passive listening and while performing a pitch discrimination task. Stimuli consisted of (1) Mandarin speech sounds that contained modulations of the fundamental frequency signaling lexical meaning and (2) English speech sounds that also contained modulations to the fundamental frequency; however, F0 modulations never provide lexical information in English. Imaging results indicated that native Mandarin speakers showed significant activation of the left anterior insular cortex, adjacent to Broca’s area, only when discriminating Mandarin speech sounds (but not when engaged in passive listening); the homologous right anterior insula was activated when this group discriminated English speech sounds, as well as when native English speakers discriminated both Mandarin and English speech sounds. These data suggest that the left anterior insula is involved in auditory processing of modulations to the fundamental frequency only when those modulations are

associated with lexical processing. Moreover, these data suggest that the neural processing of acoustic signals is context dependent and is not solely based on the acoustic attributes of the stimuli.

In addition to studies of the neural representation of F0 modulations, a number of studies have also addressed the cortical representation of formant frequency modulation in humans. It is known that neurons in the auditory cortex do not phase-lock to frequencies greater than approximately 100 Hz (Creutzfeldt et al., 1980; Eggermont, 1991; Steinschneider et al., 1998; Lu et al., 2001), whereas the formant structure of speech consists of frequencies almost exclusively above 100 Hz. Consequently, the cortical representation of frequency modulation as measured by evoked potentials is abstract (i.e., not represented with time-locked responses) relative to that described for the auditory brainstem. One cortical mechanism that has received considerable attention for the processing of rapid formant modulations is that of asymmetric processing in the left hemisphere auditory cortex. A more general hypothesis proposes that the left hemisphere auditory cortex is specialized for all forms of rapid acoustic stimuli and serves as an early acoustic analysis stage at the level of the cortex. A significant piece of evidence in support of this hypothesis was provided in a study of cortical activation patterns for rapid and slow formant frequency modulations (Belin et al., 1998). In this study, nonspeech sounds containing temporal and spectral characteristics similar to speech sounds were presented to listeners as they were PET scanned. Nonspeech sounds were used so that any cortical asymmetry could not be associated with well-known asymmetries for language processing. Results indicated that the left superior temporal gyrus (STG), including primary auditory cortex, showed greater activation than the right STG for rapid (40 ms) formant frequency transitions but not for slow (200 ms) transitions. In addition, a left hemisphere region of prefrontal cortex was asymmetrically activated for the rapid formant transition, which was corroborated in a separate fMRI study that used nearly identical acoustic stimuli (Temple et al., 2000). These data suggest that left hemisphere auditory regions preferentially process rapid formant modulations present in ongoing speech.

In summary, results measured from the auditory brainstem indicate that modulations in the fundamental frequency of speech are faithfully encoded in the FFR. Moreover, these particular brainstem responses appear to be shaped by linguistic experience, a remarkable finding that indicates that cognitive processes (e.g., language) influence basic sensory processing. In the cortex, a mechanism for encoding frequency modulation is the specialization of left hemisphere auditory regions. Results indicate that rapid frequency changes in speech-like stimuli preferentially activate the left hemisphere relative to slower frequency changes. In addition, the anterior insular cortex is activated for the processing of F0 modulations; the left hemisphere insula is specifically activated when F0 modulations provide lexical information to a native speaker, while the right hemisphere

620 Section III ■ Special Populations

insula is activated when F0 modulations do not provide lexical information. These cortical findings would appear to be contradictory; the former indicates that asymmetric activation by left hemisphere structures is based on physical parameters of the speech signal, irrespective of linguistic content, while the latter suggests that linguistic context is essential for left asymmetric insular processing of F0 modulations. However, Wong et al. (2004) stated that these results can be reconciled if the insular activity shown in their study occurs after the “acoustically specialized” cortical activity described by Belin et al. (1998) and Temple et al. (2000). If this were true, it would indicate two independent levels of cortical asymmetry: one based on the acoustic attributes of the signal and one based on the linguistic relevance to the listener. This hypothesis needs to be tested in future studies.

Acoustic Onsets

ACOUSTIC DESCRIPTION AND ROLE IN THE PERCEPTION OF SPEECH

Acoustic onsets are defined here as the spectral and temporal features present at the beginning (the initial ~40 ms) of speech sounds. While the acoustics of phonemes are only slightly altered based on their location in a word (i.e., beginning, middle, or end of a word), an emphasis has been put on acoustic onsets in the neurophysiologic literature. Consequently, acoustic onsets are discussed here separately, despite some overlap with acoustic features (i.e., frequency transitions) discussed previously.

Onset acoustics of speech sounds vary considerably in both their spectral and temporal attributes. In some cases, the spectral features of the onset are essential for perception (e.g., the onset frequency of F3 for discriminating /da/ vs. /ga/), whereas in other cases, temporal attributes of onsets are the critical feature for perception. A frequently studied acoustic phenomenon associated with the temporal attributes of speech-sound onset is that of voice onset time (VOT), which is present in stop consonants. The VOT is defined as the duration of time between the release of a stop consonant by speech articulators and the beginning of vocal fold vibration. The duration of the VOT is the acoustic cue that enables differentiation between consonants that are otherwise extremely similar (e.g., /da/ vs. /ta/, /ba/ vs. /pa/, /ga/ vs. /ka/).

PHYSIOLOGIC REPRESENTATION OF ACOUSTIC ONSETS IN THE HUMAN BRAIN

Auditory Brainstem

The brainstem response to speech-sound onset has been studied extensively (Cunningham et al., 2001; King et al., 2002; Russo et al., 2004; 2005; Wible et al., 2004; 2005; Banai et al., 2005; Johnson et al., 2005; Kraus and Nicol, 2005). The first components of the speech-evoked ABR reflect the onset of the brainstem response to the stimulus (Fig. 26.2). Speech onset is represented in the brainstem response at approximately 7 ms in the form of two peaks, positive peak V and negative peak A.

Findings from a number of studies have demonstrated that the brainstem’s response to acoustic transients is closely linked to auditory perception and to language-based cortical function such as literacy. These studies have investigated brainstem responses to speech in normal children and children with language-based learning disabilities (LD), a population that has consistently demonstrated perceptual deficits in auditory tasks using both simple (Tallal and Piercy, 1973; Reed, 1989; Hari and Kiesila, 1996; Wright et al., 1997; Hari et al., 1999; Nagarajan et al., 1999; Ahissar et al., 2001; Benasich and Tallal, 2002; Witton et al., 2002) and complex (Tallal and Piercy, 1975; Kraus et al., 1996; Bradlow et al., 1999; 2003; Ramus et al., 2003) acoustic stimuli. A general hypothesis proposes a causal link between basic auditory perceptual deficits in LDs and higher level language skills, such as reading and phonologic tasks (Tallal et al., 1993), although this relationship has been debated (Mody et al., 1997; Schulte-Korne et al., 1998; Bishop et al., 1999; Ramus et al., 2003). In support of a hypothesis linking basic auditory function and language skills, studies of the auditory brainstem indicate a fundamental deficiency in the synchrony of auditory neurons in the brainstem for a significant proportion of language-disabled subjects.

The brainstem’s response to acoustic transients in speech features prominently in distinguishing LD from normal (control) subjects. A number of studies have provided compelling evidence that the representation of speech onset (Cunningham et al., 2000; King et al., 2002; Wible et al., 2004; 2005; Banai et al., 2005) is abnormal in a significant proportion of LD subjects. For example, brainstem responses to the speech syllable /da/ were measured for a group of 33 normal and 54 LD children; a “normal range” was established from the results of the normal subjects (King et al., 2002). Results indicated that 20 LD subjects (37%) showed abnormally late responses to onset peak A. Another study showed a significant difference between normal and LD subjects based on another measure of the brainstem’s representation of acoustic transients (Wible et al., 2004). Specifically, it was shown that the slope between onset peaks V and A to the /da/ syllable was significantly smaller in LD subjects compared to normal subjects. The authors of this study indicate that the diminished V/A slope demonstrated by LDs is a measure of abnormal synchrony to the onset transients of the stimulus and could be the result of abnormal neural conduction by brainstem generators. The suggestion of abnormal neural conduction is consistent with anatomic findings of deficient axonal myelination in the temporal cortex of LD subjects (Klingberg et al., 2000). In another study (Banai et al., 2005), LD subjects with abnormal brainstem timing for acoustic transients were more likely to have a more severe form of LD, manifested in poorer scores on measures of literacy, compared to LD subjects with normal brainstem responses.

Taken together, these data suggest that the brainstem responses to acoustic transients can not only differentiate a subpopulation of LD persons from normal subjects, but can

also differentiate the LD population in terms of the severity of the disability. Findings from the brainstem measures also indicate a link between sensory encoding and cognitive processes such as literacy. An important question is whether the link between sensory encoding and cognition is a causal one and, if so, whether brainstem deficits are responsible for cortical deficits (or vice versa). Alternatively, these two abnormalities may be merely coincident. Nevertheless, the consistent findings of brainstem abnormalities in a certain portion of the LD population have led to the incorporation of this experimental paradigm into the clinical evaluation of LD and central auditory processing disorders. The “BioMAP” (Biological Marker of Auditory Processing, Biologic Systems Corp., Mundelein, IL) measures and analyzes the brainstem response to speech and has been shown to be a reliable measure for the objective evaluation of children with learning and listening disorders.

Auditory Cortex

Cortical encoding of spectral features of speech-sound onsets has been reported in the literature, most recently in a paper by Obleser et al. (2006). In this paper, it was shown that a spectral contrast at speech onset, resulting from consonant place of articulation (e.g., front-produced consonant /d/ or /t/ vs. back-produced consonant /g/ or /k/), is mapped along the anterior-posterior axis in the auditory cortex as measured by N100m source location. This is significant because it indicates that phonemes differentially activate regions of auditory cortex according to their spectral characteristics at speech onset. It was also shown that the discrete mapping of consonants according to onset acoustics is effectively erased when the speech stimuli are manipulated to become unintelligible despite keeping the spectral complexity of the stimuli largely the same. This stimulus manipulation was accomplished by altering the spectral distribution of the stimuli. The authors argue that this latter finding indicates that the cortex is spatially mapping only those sounds that are intelligible to listeners. These data provide important evidence that cortical spatial representations may serve as an important mechanism for the encoding of spectral characteristics in speech-sound onsets. In addition to differences in spatial representations for place of articulation contrast, cortical responses also showed latency differences for these contrasts. Specifically, it was shown that front consonants, which have higher frequency onsets, elicited earlier N100m responses than back consonants. This finding is consistent with near-field recordings measured from animal models indicating earlier response latencies for speech onsets with higher frequency formants (McGee et al., 1996).

Cortical responses to temporal features of speech-sound onsets have also been reported in the literature, all of which have used VOT contrasts as stimuli. These studies were performed by measuring obligatory evoked potentials (N100 responses) to continua of consonant-vowel speech sounds that varied gradually according to VOT (Sharma and Dorman, 1999; 2000). Additionally, perception of these phonetic con-

trasts was also measured using the same continua as a means of addressing whether cortical responses reflected categorical perception of the phonemes. Neurophysiologic results indicated that for both /ba-/pa/ and /ga-/ka/ phonetic contrasts, one large negative peak was evident at approximately 100 ms in the response waveform for stimulus VOTs <40 ms. Importantly, a second negative peak in the response waveform emerged for stimulus VOTs of 40 ms, and this second peak occurred approximately 40 ms after the first peak and was thought to represent the onset of voicing in the stimulus. Moreover, as the VOT of the stimulus increased in duration, the lag between the second peak relative to the first increased proportionally, resulting in a strong correlation between VOT and the latency between the successive peaks ($r = \sim 0.80$). The onset of double peaks in cortical responses with a VOT of 40 ms is consistent with neurophysiologic responses measured directly from the auditory cortex of humans (Steinschneider et al., 1999). An important consideration is that the onset of the double peak occurred at 40 ms for both the /ba-/pa/ and /ga-/ka/ phonetic contrasts. In contrast, behavioral results require different VOTs to distinguish the /ba-/pa/ and /ga-/ka/ phonetic contrasts. Specifically, a VOT of ~ 40 ms was required for listeners to correctly identify /pa/ from /ba/, while a VOT of ~ 60 ms was required for correct identification of /ga/ from /ka/. Taken together, these data indicate that the cortical responses reflect the actual VOT at 40 ms irrespective of the categorical perception of the phonetic contrasts, which in the case of the /ga-/ka/ contrast requires 60 ms.

Brainstem-Cortex Relationships

In addition to linking precise brainstem timing of acoustic transients to linguistic function, it has also been shown that abnormal encoding of acoustic transients in the brainstem is related to abnormal auditory responses measured at the level of cortex. In addition to their imprecise representation of sounds at the auditory brainstem, a significant proportion of LDs have also consistently demonstrated abnormal representations of simple (Menell et al., 1999; Ahissar et al., 2000) and complex (Kraus et al., 1996; Bradlow et al., 1999; Ahissar et al., 2001; Wible et al., 2002; 2005; Banai et al., 2005) acoustic stimuli at the level of the auditory cortex. Three recent studies linked abnormal neural synchrony for acoustic transients at the auditory brainstem to abnormal representations of sounds in the cortex. In one study (Wible et al., 2005), it was shown that a brainstem measure of the encoding of acoustic transients, the duration of time between onset peaks V and A, was positively correlated to the auditory cortex's susceptibility to background noise in both normal and LD subjects. Specifically, the longer the duration between onset peaks V and A, the more degraded cortical responses became in the presence of background noise. In another study, it was shown that individuals with abnormal brainstem timing to acoustic transients were more likely to indicate reduced cortical sensitivity to acoustic change, as measured by the mismatch negativity (MMN) response

622 Section III ■ Special Populations

(Banai et al., 2005). Finally, a third study showed that brainstem timing for speech-sound onset and offset predicts the degree of cortical asymmetry for speech sounds measured across a group of children with a wide range of reading skills (Abrams et al., 2006). Thus, results from these studies indicate that abnormal encoding of acoustic onsets at the brainstem may be a critical marker for systemic auditory deficits manifested at multiple levels of the auditory system, including the cortex.

In summary, evidence from examining the ABR indicates that acoustic transients are encoded in a relatively simple fashion in the brainstem, yet they represent a complex phenomenon that is related to linguistic ability and cortical function. In the cortex, results indicate that spectral contrasts of speech onsets are mapped along the anterior-posterior axis in the auditory cortex, while temporal attributes of speech onsets, as manifested by the VOT, are precisely encoded with double-peaked N100 responses.

The Speech Envelope

DEFINITION AND ROLE IN THE PERCEPTION OF SPEECH

The speech envelope refers to the temporal fluctuations in the speech signal between 2 and 50 Hz. The dominant frequency of the speech envelope is at ~ 4 Hz, which reflects the average syllabic rate of speech (Steeneken and Houtgast, 1980). Envelope frequencies in normal speech are generally below 8 Hz (Houtgast and Steeneken, 1985), and the perceptually essential frequencies of the speech envelope are between 4 and 16 Hz (Drullman et al., 1994; van der Horst et al., 1999), although frequencies above 16 Hz contribute slightly to speech recognition (Shannon et al., 1995). The speech envelope provides phonetic and prosodic cues to the duration of speech segments, manner of articulation, the presence (or absence) of voicing, syllabication, and stress (van der Horst et al., 1999). The perceptual significance of the speech envelope has been investigated using a number of methodologies (Drullman et al., 1994; Shannon et al., 1995; Smith et al., 2002b), and taken together, these data indicate that the speech envelope is both necessary and sufficient for normal speech recognition.

PHYSIOLOGIC REPRESENTATION OF THE SPEECH ENVELOPE IN AUDITORY CORTEX

Only a few studies have investigated how the human brain represents the slow temporal information of the speech envelope. It should be noted that the representation of the speech envelope in humans has only been studied at the level of the cortex since measuring ABRs typically involves filtering out the neurophysiologic responses below ~ 100 Hz (Hall, 1992). Since speech envelope frequencies are between 2 and 50 Hz, any linear representation of the speech envelope in brainstem responses is removed with brainstem filtering.

In one MEG study, responses from the auditory cortex to natural and time-compressed (i.e., rapid) speech sen-

tences were measured while subjects listened for semantic incongruities in experimental sentences (Ahissar et al., 2001). Results indicate that the human cortex synchronizes its response to the contours of the speech envelope, a phenomenon known as “phase-locking,” and mimics the frequency content of the speech envelope, which the investigators called “frequency matching.” Moreover, it was shown that these two neurophysiologic measures correlate with subjects’ ability to perceive the speech sentences; as speech sentences become more difficult to perceive due to increased time compression, the ability of the cortex to phase-lock and frequency match is more impaired. These results are in concert with results from the animal literature, which show that cortical neurons of the primary auditory cortex represent the temporal envelope of complex acoustic stimuli (i.e., animal communication calls) by phase-locking to this temporal feature of the stimulus (Wang et al., 1995; Gehr et al., 2000; Nagarajan et al., 2002).

A second line of inquiry into the cortical representation of speech envelope cues was described previously in this chapter in the discussion of cortical responses to VOT (Sharma and Dorman, 1999; 2000; Sharma et al., 2000). Acoustically, VOT is a slow temporal cue in speech (40 to 60 ms; 17 to 25 Hz) that falls within the range of speech envelope frequencies. As discussed earlier, neurophysiologic results indicate that for both /ba/-/pa/ and /ga/-/ka/ phonetic contrasts, cortical N100 responses precisely represent the acoustic attributes of VOT. In addition, it was shown that neural responses are independent of the categorical perception of these phonetic contrasts (see the Acoustic Onsets section for a more detailed description of this study).

On the surface, it may appear that the findings from these experiments contradict one another since cortical phase-locking to the speech envelope correlates with perception in one study (Ahissar et al., 2001), while phase-locking fails to correlate with perception in other studies (Sharma and Dorman, 1999; 2000; Sharma et al., 2000). These data are not, however, in contradiction to one another. In both cases, an a priori requirement for perception is phase-locking to the speech envelope; there is no evidence for perception in the absence of accurate phase-locking to the temporal envelope in either study. The primary difference between the studies is that, despite phase-locking to the temporal envelope in the /ka/ stimulus condition at a VOT of ~ 40 ms, reliable perception of /ka/ occurs at approximately 60 ms. This suggests that accurate phase-locking is required for perception; however, perception cannot be predicted by phase-locking alone. Presumably, in the case of the /ka/ VOT stimulus, there is another processing stage that uses the phase-locked temporal information in conjunction with additional auditory-linguistic information (e.g., repeated exposure to /ka/ stimuli with 60-ms VOT) as a means of forming phonetic category boundaries. The questions of if and how category boundaries are established, irrespective of auditory phase-locking, require additional investigation.

CONCLUSION

Speech is a highly complex signal composed of a variety of acoustic features, all of which are important for normal speech perception. Normal perception of these acoustic features certainly relies on their neural encoding, which has been the subject of this review. An obvious conclusion from these studies is that the central auditory system is a remarkable machine, able to simultaneously process the multiple acoustic cues of ongoing speech in order to decode a linguistic message. Furthermore, how the human brain is innately and dynamically programmed to use any number of these acoustic cues for the purpose of language, given the appropriate degree and type of stimulus exposure, further underscores the magnificence of this system.

A limitation of this chapter is that it has adopted a largely “bottom-up” approach to the acoustic encoding of speech sounds; neural encoding of acoustic signals is generally discussed as an afferent phenomenon with minimal consideration for the dynamic interactions provided by top-down connections in the auditory system (Xiao and Suga, 2002; Perrot et al., 2006). A notable exception to this includes work by Krishnan et al. (2004), which was described in the section on frequency modulation, in which the role of language experience was shown to affect sensory encoding in the auditory brainstem. Another limitation to this chapter is that it has also ignored the influence of other systems of the central nervous system, such as cognitive and emotional effects on auditory processing of speech, which most certainly have a role in shaping auditory activity.

To garner a greater understanding of how the central auditory system processes speech, it is important to con-

sider both subcortical and cortical auditory regions. Across the acoustic features described in this review, the brainstem appears to represent acoustic events in a relatively linear fashion. The fundamental frequency and its modulation are represented with highly synchronized activity as reflected by the FFR; speech-sound onset is represented with highly predictable neural activation patterns that vary within fractions of milliseconds. Alternatively, the cortex appears to transform many of these acoustic cues, resulting in more complex representations of acoustic features of speech. For example, many of the cortical findings described here are based on the spatial representation of acoustic features (i.e., the relationship between F1-F2 required for vowel identification; the differentiation of speech transients; the encoding of periodicity). Because cortical neurons are not able to phase-lock to high-frequency events, it is tempting to propose that the cortex has found an alternative method for encoding these features based on the activity of spatially distributed neural populations. The extent to which these acoustic features are truly represented via a spatial organization in cortex is a future challenge that will likely be achieved using high-resolution imaging technologies in concert with EEG and MEG technologies.

ACKNOWLEDGEMENTS

We would like to thank Karen Banai and Trent Nicol for their comments on a previous draft of this chapter. This work is supported by the National Institutes of Health grant R01 DC01510-10 and National Organization for Hearing Research grant 340-B208.

REFERENCES

- Abrams DA, Nicol T, Zecker S, Kraus N. (2006) Auditory brainstem timing predicts cerebral asymmetry for speech. *J Neurosci.* 26, 11131–11137.
- Ahissar M, Hochstein S. (2004) The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci.* 8, 457–464.
- Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM. (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc Natl Acad Sci USA.* 98, 13367–13372.
- Ahissar M, Protopapas A, Reid M, Merzenich MM. (2000) Auditory processing parallels reading abilities in adults. *Proc Natl Acad Sci USA.* 97, 6832–6837.
- Banai K, Nicol T, Zecker SG, Kraus N. (2005) Brainstem timing: implications for cortical processing and literacy. *J Neurosci.* 25, 9850–9857.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. (2000) Voice-selective areas in human auditory cortex. *Nature.* 403, 309–312.
- Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A, Masure MC, Samson Y. (1998) Lateralization of speech and auditory temporal processing. *J Cogn Neurosci.* 10, 536–540.
- Benasich AA, Tallal P. (2002) Infant discrimination of rapid auditory cues predicts later language impairment. *Behav Brain Res.* 136, 31–49.
- Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T. (1997) Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci.* 17, 353–362.
- Bishop DV, Bishop SJ, Bright P, James C, Delaney T, Tallal P. (1999) Different origin of auditory and phonological processing problems in children with language impairment: evidence from a twin study. *J Speech Lang Hear Res.* 42, 155–168.
- Bradlow AR, Kraus N, Hayes E. (2003) Speaking clearly for children with learning disabilities: sentence perception in noise. *J Speech Lang Hear Res.* 46, 80–97.
- Bradlow AR, Kraus N, Nicol TG, McGee TJ, Cunningham J, Zecker SG, Carrell TD. (1999) Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children. *J Acoust Soc Am.* 106, 2086–2096.
- Creutzfeldt O, Hellweg FC, Schreiner C. (1980) Thalamocortical transformation of responses to complex auditory stimuli. *Exp Brain Res.* 39, 87–104.

624 Section III ■ Special Populations

- Cunningham J, Nicol T, Zecker S, Bradlow A, Kraus N. (2001) Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin Neurophysiol.* 112, 758–767.
- Cunningham J, Nicol T, Zecker S, Kraus N. (2000) Speech-evoked neurophysiologic responses in children with learning problems: development and behavioral correlates of perception. *Ear Hear.* 21, 554–568.
- Diesch E, Luce T. (1997) Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation. *Psychophysiology.* 34, 501–510.
- Drullman R, Festen JM, Plomp R. (1994) Effect of temporal envelope smearing on speech reception. *J Acoust Soc Am.* 95, 1053–1064.
- Eggermont JJ. (1991) Rate and synchronization measures of periodicity coding in cat primary auditory cortex. *Hear Res.* 56, 153–167.
- Galbraith GC, Threadgill MR, Hemsley J, Salour K, Songdej N, Ton J, Cheung L. (2000) Putative measure of peripheral and brainstem frequency-following in humans. *Neurosci Lett.* 292, 123–127.
- Gandour J, Wong D, Hutchins G. (1998) Pitch processing in the human brain is influenced by language experience. *Neuroreport.* 9, 2115–2119.
- Gardi J, Merzenich M, McKean C. (1979) Origins of the scalp recorded frequency-following response in the cat. *Audiology.* 18, 358–381.
- Gehr DD, Komiya H, Eggermont JJ. (2000) Neuronal responses in cat primary auditory cortex to natural and altered species-specific calls. *Hear Res.* 150, 27–42.
- Geschwind N, Galaburda AM. (1985) Cerebral lateralization. Biological mechanisms, associations, and pathology: I. A hypothesis and a program for research. *Arch Neurol.* 42, 428–459.
- Hall JH. (1992) *Handbook of Auditory Evoked Responses.* Boston: Allyn and Bacon.
- Hari R, Kiesila P. (1996) Deficit of temporal auditory processing in dyslexic adults. *Neurosci Lett.* 205, 138–140.
- Hari R, Saaskilahti A, Helenius P, Uutela K. (1999) Non-impaired auditory phase locking in dyslexic adults. *Neuroreport.* 10, 2347–2348.
- Hayes EA, Warrier CM, Nicol TG, Zecker SG, Kraus N. (2003) Neural plasticity following auditory training in children with learning problems. *Clin Neurophysiol.* 114, 673–684.
- Hickok G, Poeppel D. (2004) Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition.* 92, 67–99.
- Houtgast T, Steeneken HJM. (1985) A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J Acoust Soc Am.* 77, 1069–1077.
- Jezzard P, Matthews PM, Smith SM. (2001) *Functional MRI: An Introduction to Methods.* Oxford, United Kingdom: Oxford University Press.
- Johnson K. (1997) *Acoustic and Auditory Phonetics.* Cambridge, MA: Blackwell Publishers Inc.
- Johnson KL, Nicol TG, Kraus N. (2005) Brain stem response to speech: a biological marker of auditory processing. *Ear Hear.* 26, 424–434.
- King C, Warrier CM, Hayes E, Kraus N. (2002) Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neurosci Lett.* 319, 111–115.
- Klein D, Zatorre RJ, Milner B, Zhao V. (2001) A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage.* 13, 646–653.
- Klingberg T, Hedehus M, Temple E, Salz T, Gabrieli JD, Moseley ME, Poldrack RA. (2000) Microstructure of temporo-parietal white matter as a basis for reading ability: evidence from diffusion tensor magnetic resonance imaging. *Neuron.* 25, 493–500.
- Kraus N, McGee TJ, Carrell TD, Zecker SG, Nicol TG, Koch DB. (1996) Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science.* 273, 971–973.
- Kraus N, Nicol T. (2005) Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends Neurosci.* 28, 176–181.
- Krishnan A. (2002) Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res.* 166, 192–201.
- Krishnan A, Xu Y, Gandour JT, Cariani PA. (2004) Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res.* 189, 1–12.
- Krishnan A, Xu Y, Gandour JT, Cariani P. (2005) Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res.* 25, 161–168.
- Lehiste I, Peterson GE. (1961) Transitions, glides, and diphthongs. *J Acoust Soc Am.* 33, 268–277.
- Lu T, Liang L, Wang X. (2001) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci.* 4, 1131–1138.
- Makela AM, Alku P, Makinen V, Valtonen J, May P, Tiitinen H. (2002) Human cortical dynamics determined by speech fundamental frequency. *Neuroimage.* 17, 1300–1305.
- McGee T, Kraus N, King C, Nicol T, Carrell TD. (1996) Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe. *J Acoust Soc Am.* 99, 3606–3614.
- Menell P, McAnally KI, Stein JF. (1999) Psychophysical sensitivity and physiological response to amplitude modulation in adult dyslexic listeners. *J Speech Lang Hear Res.* 42, 797–803.
- Mody M, Studdert-Kennedy M, Brady S. (1997) Speech perception deficits in poor readers: auditory processing or phonological coding? *J Exp Child Psychol.* 64, 199–231.
- Naatanen R, Picton T. (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology.* 24, 375–425.
- Nagarajan SS, Cheung SW, Bedenbaugh P, Beitel RE, Schreiner CE, Merzenich MM. (2002) Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. *J Neurophysiol.* 87, 1723–1737.
- Nagarajan SS, Mahncke H, Salz T, Tallal P, Roberts T, Merzenich MM. (1999) Cortical auditory signal processing in poor readers. *Proc Natl Acad Sci USA.* 96, 6483–6488.
- Nearey TM, Assmann PF. (1986) Modeling the role of inherent spectral change in vowel identification. *J Acoust Soc Am.* 80, 1297–1308.
- Nusbaum HC, Morin TM. (1992) Paying attention to differences among talkers. In: Tohkura Y, Sagisaka Y, Vatikiotis-Bateson E, eds. *Speech Perception, Production, and Linguistic Structure.* Tokyo: Ohmsha Publishing; pp 113–134.
- Obleser J, Elbert T, Lahiri A, Eulitz C. (2003) Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Brain Res Cogn Brain Res.* 15, 207–213.

- Obleser J, Scott SK, Eulitz C. (2006) Now you hear it, now you don't: transient traces of consonants and their nonspeech analogues in the human brain. *Cereb Cortex*. 16, 1069–1076.
- Perrot X, Ryvlin P, Isnard J, Guenot M, Catenoix H, Fischer C, Mauguiere F, Collet L. (2006) Evidence for corticofugal modulation of peripheral auditory activity in humans. *Cereb Cortex*. 16, 941–948.
- Peterson GE, Barney HL. (1952) Control methods used in a study of the vowels. *J Acoust Soc Am*. 24, 175–184.
- Ramus F, Rosen S, Dakin SC, Day BL, Castellote JM, White S, Frith U. (2003) Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain*. 126, 841–865.
- Rauschecker JP. (1997) Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Otolaryngol Suppl*. 532, 34–38.
- Reed MA. (1989) Speech perception and the discrimination of brief auditory cues in reading disabled children. *J Exp Child Psychol*. 48, 270–292.
- Rosen S. (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci*. 336, 367–373.
- Russo NM, Nicol TG, Musacchia G, Kraus N. (2004) Brainstem responses to speech syllables. *Clin Neurophysiol*. 115, 2021–2030.
- Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N. (2005) Auditory training improves neural timing in the human brainstem. *Behav Brain Res*. 156, 95–103.
- Sachs MB, Young ED. (1979) Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. *J Acoust Soc Am*. 66, 470–479.
- Sachs MB, Voigt HF, Young ED. (1983) Auditory nerve representation of vowels in background noise. *J Neurophysiol*. 50, 27–45.
- Sato S. (1990) *Magnetoencephalography* New York: Raven Press.
- Schreiner CE. (1998) Spatial distribution of responses to simple and complex sounds in the primary auditory cortex. *Audiol Neurootol*. 3, 104–122.
- Schulte-Korne G, Deimel W, Bartling J, Remschmidt H. (1998) Auditory processing and dyslexia: evidence for a specific speech processing deficit. *Neuroreport*. 9, 337–340.
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. (1995) Speech recognition with primarily temporal cues. *Science*. 270, 303–304.
- Sharma A, Dorman MF. (1999) Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J Acoust Soc Am*. 106, 1078–1083.
- Sharma A, Dorman MF. (2000) Neurophysiologic correlates of cross-language phonetic perception. *J Acoust Soc Am*. 107, 2697–2703.
- Sharma A, Marsh C, Dorman M. (2000) Relationship between N1 evoked potential morphology and the perception of voicing. *J Acoust Soc Am*. 108, 3030–3035.
- Shestakova A, Brattico E, Soloviev A, Klucharev V, Huotilainen M. (2004) Orderly cortical representation of vowel categories presented by multiple exemplars. *Brain Res Cogn Brain Res*. 21, 342–350.
- Smith AJ, Blumenfeld H, Behar KL, Rothman DL, Shulman RG, Hyder F. (2002a) Cerebral energetics and spiking frequency: the neurophysiological basis of fMRI. *Proc Natl Acad Sci USA*. 99, 10765–10770.
- Smith JC, Marsh JT, Brown WS. (1975) Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr Clin Neurophysiol*. 39, 465–472.
- Smith ZM, Delgutte B, Oxenham AJ. (2002b) Chimaeric sounds reveal dichotomies in auditory perception. *Nature*. 416, 87–90.
- Song JH, Banai K, Russo NM, Kraus N. (2006) On the relationship between speech- and nonspeech-evoked auditory brainstem responses. *Audiol Neurootol*. 11, 233–241.
- Steeneken HJ, Houtgast T. (1980) A physical method for measuring speech-transmission quality. *J Acoust Soc Am*. 67, 318–326.
- Steinschneider M, Reser DH, Fishman YI, Schroeder CE, Arezzo JC. (1998) Click train encoding in primary auditory cortex of the awake monkey: evidence for two mechanisms subserving pitch perception. *J Acoust Soc Am*. 104, 2935–2955.
- Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA 3rd. (1999) Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J Neurophysiol*. 82, 2346–2357.
- Stillman RD, Crow G, Moushegian G. (1978) Components of the frequency-following potential in man. *Electroencephalogr Clin Neurophysiol*. 44, 438–446.
- Tallal P, Piercy M. (1973) Defects of non-verbal auditory perception in children with developmental aphasia. *Nature*. 241, 468–469.
- Tallal P, Piercy M. (1975) Developmental aphasia: the perception of brief vowels and extended stop consonants. *Neuropsychologia*. 13, 69–74.
- Tallal P, Miller S, Fitch RH. (1993) Neurobiological basis of speech: a case for the preeminence of temporal processing. *Ann N Y Acad Sci*. 682, 27–47.
- Temple E, Poldrack RA, Protopapas A, Nagarajan S, Salz T, Tallal P, Merzenich MM, Gabrieli JD. (2000) Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. *Proc Natl Acad Sci USA*. 97, 13907–13912.
- van der Horst R, Leeuw AR, Dreschler WA. (1999) Importance of temporal-envelope cues in consonant recognition. *J Acoust Soc Am*. 105, 1801–1809.
- Wang X, Merzenich MM, Beitel R, Schreiner CE. (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol*. 74, 2685–2706.
- Wang Y, Jongman A, Sereno JA. (2001) Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain Lang*. 78, 332–348.
- Wible B, Nicol T, Kraus N. (2002) Abnormal neural encoding of repeated speech stimuli in noise in children with learning problems. *Clin Neurophysiol*. 113, 485–494.
- Wible B, Nicol T, Kraus N. (2004) Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol Psychol*. 67, 299–317.
- Wible B, Nicol T, Kraus N. (2005) Correlation between brainstem and cortical auditory processes in normal and language-impaired children. *Brain*. 128, 417–423.
- Witton C, Stein JF, Stoodley CJ, Rosner BS, Talcott JB. (2002) Separate influences of acoustic AM and FM sensitivity on the phonological decoding skills of impaired and normal readers. *J Cogn Neurosci*. 14, 866–874.

626 Section III ■ Special Populations

- Wong PC, Parsons LM, Martinez M, Diehl RL. (2004) The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. *J Neurosci.* 24, 9153–9160.
- Wright BA, Lombardino LJ, King WM, Puranik CS, Leonard CM, Merzenich MM. (1997) Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature.* 387, 176–178.
- Xiao Z, Suga N. (2002) Modulation of cochlear hair cells by the auditory cortex in the mustached bat. *Nat Neurosci.* 5, 57–63.
- Zatorre RJ, Belin P. (2001) Spectral and temporal processing in human auditory cortex. *Cereb Cortex.* 11, 946–953.
- Zatorre RJ, Belin P, Penhune VB. (2002) Structure and function of auditory cortex: music and speech. *Trends Cogn Sci.* 6, 37–46.