CHAPTER 28

# Auditory Pathway Representations of Speech Sounds in Humans

Daniel A. Abrams and Nina Kraus

## INTRODUCTION

An essential function of the human 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 way the central nervous system represents the acoustic structure of speech. Consequently, an understanding of how the nervous system accomplishes this task would provide important insights into the basis of language function and auditory-based cognition.

One of the challenges faced by researchers is that speech is a complex acoustic signal that is rich in both spectral and temporal features. In everyday listening situations, the abundance of acoustical cues in the speech signal provides enormous perceptual benefits to listeners. For example, listeners are able to shift their attention between different acoustical cues when perceiving speech from different talkers to compensate for the built-in variations in the acoustical properties (Nusbaum and Morin, 1992). This form of "listener flexibility" reflects a critical aspect of speech perception: The listener makes use of whatever spectral or temporal cues are available to help decode the incoming speech signal.

There are two basic approaches that researchers have adopted for conducting experiments on speech perception and the 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. Whereas simple stimuli enable researchers to reduce the acoustics of speech to its most basic elements, the auditory system is nonlinear (Sachs and Young, 1979) 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; Cunningham et al., 2002; Skoe and Kraus, 2010). There are many advantages to this approach. First, these stimuli have greater face validity for understanding speech processing. Second, a complete description of how the nonlinear auditory system responds to speech can only be obtained by using speech stimuli. Third, long-term exposure to speech sounds and their use linguistically produces plastic changes in the auditory pathways that 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 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 (see Table 28.1). These acoustic features of speech were chosen because they have essential roles in normal speech perception. Each section contains a description of the acoustical feature, an explanation of its importance in speech perception, followed by a review and assessment of the data for that acoustic feature.

An exciting aspect of brain function is the remarkable capacity of the brain to modify its functional properties following training. In the auditory domain, a growing body of research has shown that targeted training and remediation programs can provide substantial speech perception benefit to a number of populations, including both normal listeners and clinical populations with impaired auditory function. Given the prevalence of hearing deficits in industrialized societies and an aging population in most Western countries, targeted auditory training to maintain and improve speech perception, particularly in the presence of background noise, represents an important strategy for sustaining speech-based communication and cognitive skills (Lin et al., 2013). Importantly, behavioral improvements that result from training originate in changes in brain function, and it is of great interest to the field of auditory research to understand what aspects of brain function change in response to auditory-based training. These findings are of theoretical interest: Many auditory training paradigms constitute relatively complex tasks, exposing the listener to a host of acoustical features and tapping into a range of sensory and cognitive skills; therefore, an understanding of the specific brain changes that accompany training-based improvement provides a window on the particular acoustical features that are most important for improvement on the trained tasks. Thus, a final goal of this chapter is to highlight exciting recent research describing

TABLE 28.1

Acoustic Features of Speech and their Representations in the Central Auditory System

| Major Sections: Acoustic Features in Speech | Feature's Role in the Speech Signal | Brainstem Measure | Cortical Measure |
|---|---|---|---|
| 1. Formant structure | Ubiquitous in vowels, approximants, and nasals; essential for vowel perception | Frequency-following response | N100m source location; STS activity (fMRI) |
| 2. Periodicity | Temporal cue for the fundamental frequency and low formant frequencies (50–500 Hz) | Frequency-following response | N100m source location and amplitude; nonprimary auditory cortex activity patterns (fMRI) |
| 3. Frequency transitions | Consonant identification; signal the presence of diphthongs and glides; linguistic pitch | Frequency-following response | Left versus right STG activity (fMRI) |
| 4. Acoustic onsets | Phoneme identification | ABR onset complex | N100m source location; N100 latency |
| 5. Speech envelope | Syllable and low-frequency (<50 Hz) patterns in speech | N/A | N100m phase-locking |

changes in auditory brain function following speech and auditory training, with a focus on therapeutic training paradigms designed to improve speech perception in both clinical populations and normal hearing listeners.

An important consideration is that the acoustical 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 acoustical events that occur at regular time intervals. Many features in the speech signal are periodic; however, describing all of these simultaneously occurring periodic features would be experimentally unwieldy. For simplicity, and to show how these features were investigated, some related acoustical features will be discussed in separate sections. Throughout the chapter we have tried to identify when there is overlap among the acoustical 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 characteristics 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 component of speech results from the periodic beating of the vocal folds. In Figure 28.1A, 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 dark lines whereas smaller amounts of energy are depicted in white. The fundamental frequency can be seen as the horizontal band of energy in Figure 28.1A that is closest to the $x$-axis (i.e., lowest in frequency). The fundamental frequency is labeled F0 and provides the perceived pitch of an individual's voice.

### Harmonic Structure

An acoustical feature that is related to the fundamental frequency of speech is known as the harmonic structure. Speech harmonics, which are integer multiples of the fundamental frequency, are present in ongoing speech. The harmonic structure of speech is displayed in Figure 28.1A, as the regularly spaced horizontal bands of energy that are seen throughout the sentence.

### Formant Structure

Another essential acoustical feature of speech is the formant structure which describes a series of discrete peaks in the frequency spectrum of speech that are the result of an interaction between the frequency of the vocal-fold vibrations and the speaker's vocal tract resonance. The frequency of these peaks, as well as the relative frequency between peaks, varies for different speech sounds. The formant structure of speech depends on the harmonic structure of speech. Harmonic structure is represented by integer multiples of the fundamental frequency, and formants are harmonics that are close to a resonant frequency of the vocal tract. In Figure 28.1, the formant structure of speech is represented by the series of horizontal, and occasionally diagonal, lines that

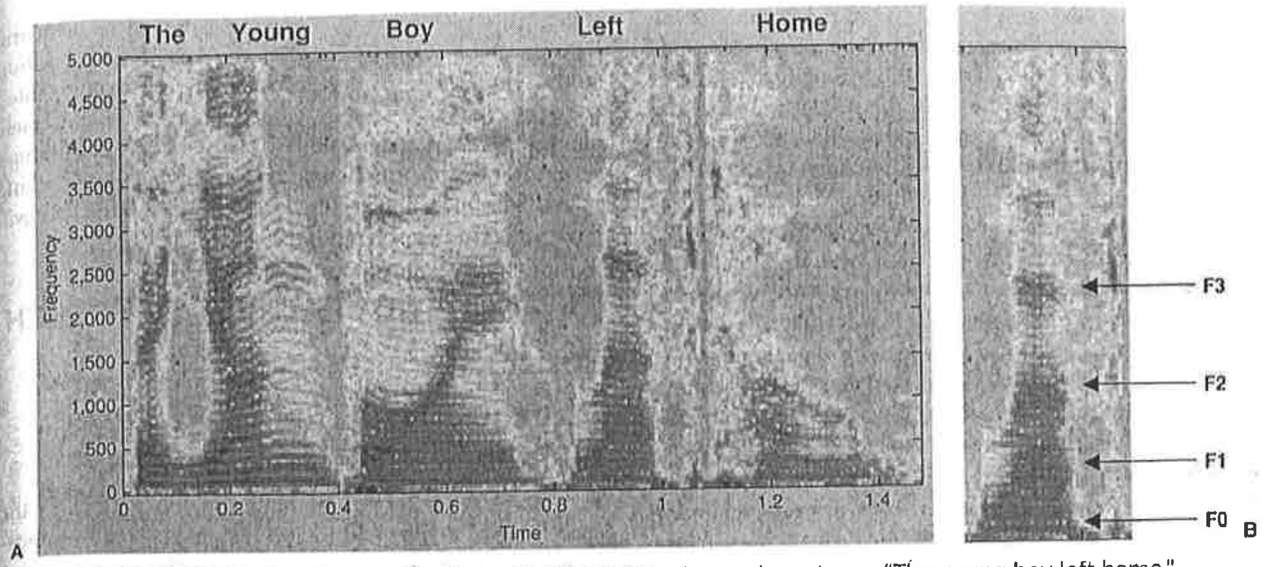

**FIGURE 28.1** Spectrogram for the naturally produced speech sentence "The young boy left home." (**A**) The complete sentence; (**B**) the word "left" is enlarged to illustrate the frequency structure: The fundamental frequency (F0) and formants (F1–F3) are represented in the spectrogram by *broad dark lines* of energy.

are darker than their neighbors that run through most of the speech utterance. The word "left" has been enlarged in Figure 28.1B to better illustrate this phenomenon. The broad and dark 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 F1, whereas subsequent formants are notated F2, F3, and so on. The frequencies of F1 and F2 in particular are important for vowel identity.

## THE MEASURES OF BRAIN ACTIVITY

We begin by describing the neurophysiological measures that have been used to probe auditory responses to speech and speech-like stimuli (comprehensive descriptions of these measures can be found elsewhere: Hall, 1992 as well as in chapters in this book). 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 neural activity in 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 ABRs (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 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 utilizes 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 neural activity auditory cortical structures 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, 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 of large populations of auditory neurons, though this technique is limited in its spatial resolution. Because of large inter- and intrasubject variability in cortical responses, these measures are not generally used clinically. Results from these two cortical 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: Typically, 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 is 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., 2002) 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, as 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 and precise 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 cover only studies of 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 to 500 Hz (Rosen, 1992). Important aspects of the speech signal that contain periodic acoustic information include the fundamental frequency and all 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 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.

### PHYSIOLOGICAL REPRESENTATION OF PERIODICITY IN THE HUMAN BRAIN

#### Auditory Brainstem

The short-latency frequency-following response (FFR) is an electrophysiological 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). Based on the frequency range that can be measured with the FFR, a representation of the fundamental frequency can be measured using this methodology (Krishnan et al., 2004; Russo et al., 2004; Skoe and Kraus, 2010), 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 brainstem response (i.e., FFR) according to a

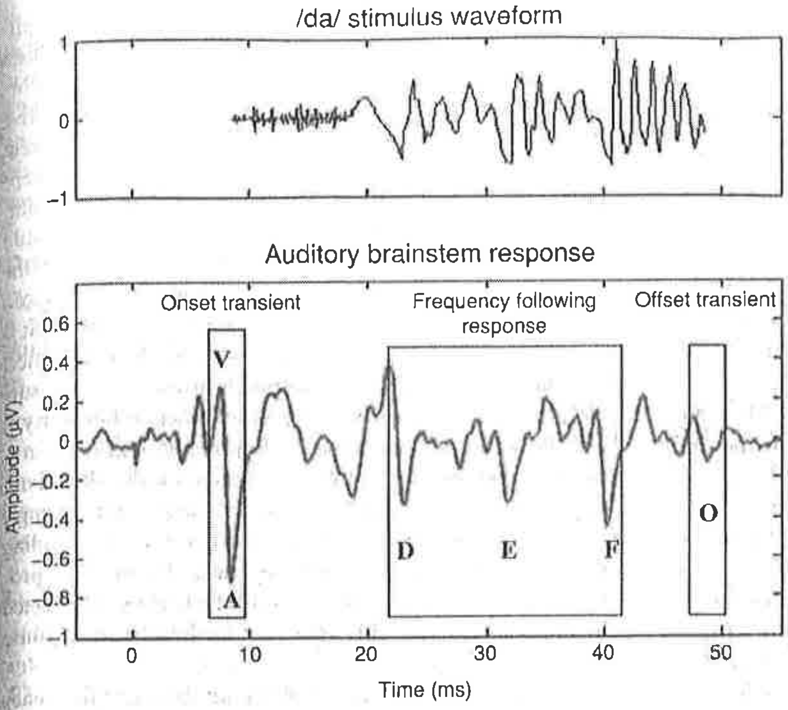

**FIGURE 28.2** Acoustic waveform of the synthesized speech stimulus /da/ (*above*) and grand average auditory brainstem responses to /da/ (*below*). The stimulus has been moved forward in time to the latency of onset responses (peak V) to enable direct comparisons with brainstem responses. Peaks V and 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.

series of peaks that are temporally spaced corresponding to the wavelength of the fundamental frequency. An example of F0 representation in the FFR can be seen in Figure 28.2, 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 are clearly represented in negative-going 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 which 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 (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 acoustical representation in the auditory brainstem. The source portion of the brainstem's response to speech is the representation of the F0, whereas 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 whereas filter class representation is impaired (Banai et al., 2009; Hornickel et al., 2012a; King et al., 2002). The converse, impairments in brainstem encoding of source (F0) but not filter components, is a characteristic of individuals with poor hearing in noise (Anderson et al., 2011). These data suggest that the acoustical 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.

### Cortex

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). Therefore, cortical phase-locking to the fundamental frequency of speech, which is near or greater than 100 Hz, is poor, and it is generally thought that the brainstem's phase-locked 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), which is 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 an adult or a child produces the vocalization, even though there is a considerable difference in the F0 of the adult's and child's vocalizations. To address how auditory cortical responses reflect relatively large variations in F0 between listeners, 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 F0 does not provide essential acoustic information relevant to the semantic meaning of the speech sound, it may be 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.

### Electrophysiological Changes due to Training

The brain's representation of periodicity has been shown to be malleable following auditory-based training. The goal of one study was to train the perception of speech in the presence of background noise, an environmental sound source which negatively impacts speech perception in normal individuals and has even more severe perceptual consequences in individuals with hearing impairments. In this study, a group of 28 normal hearing young adults were trained on a commercially available computer program entitled "Listening and Communication Enhancement" (LACE) (Sweetow and Sabes, 2006), which trains listeners on a number of auditory tasks including comprehension of degraded speech, auditory mnemonic and cognitive skills, and communication strategies (Song et al., 2012). After 4 weeks of training, participants showed improvements in measures of speech perception in noise as measured by LACE as well as independent measures of speech perception in noise, including the Hearing in Noise Test (Nilsson et al., 1994) and the Quick Speech in Noise Test (Killion et al., 2004). An age-matched group of normal hearing, untrained listeners showed no improvements in speech in noise perception.

Neural correlates of these behavioral improvements were explored by measuring ABRs to a synthetic /da/ stimulus in both quiet and in the presence of background noise. Results showed that behavioral improvements in trained listeners were accompanied by enhanced brainstem representation of periodicity, as measured by the spectral magnitude of the F0 and the second harmonic (H2), in responses measured in the presence of background noise. An important consideration is the breadth of auditory and cognitive skills trained by LACE and the specificity of these brainstem results. The LACE program broadly trains speech perception in noise, and consequently the brainstem representation of any number of acoustical features in speech could have shown training-related effects. Nevertheless, only the F0 and H2 features of the brainstem response were enhanced following LACE training. The interpretation of this result is that the brain's coding of periodicity is a particularly critical element for the perception of speech in noise. On the surface, this may be surprising: The fundamental frequency is not always necessary for speech comprehension. For example, the fundamental frequency is systematically filtered out of all telephone signals. Nevertheless, these results strongly suggest that in challenging listening conditions, including the perception of speech in noise, periodic features may provide important acoustical benefit to the listener as reflected by the sharpening of this feature in the brainstem response to speech in noise.

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 children with learning disabilities (LDs) despite the abnormal representations of speech-sound onset and first formant frequency. Yet, its role appears to be essential in hearing speech in noise. 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 differing F0s between speakers are filtered out by the time the signal reaches the level of auditory cortex. Results from speech in noise training indicate that improvements in speech perception in noise result in systematic enhancement of periodic aspects of the speech signal, including the F0 and H2 components.

## 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 for a more complete acoustical description of the formant structure). The formant structure is a dominant acoustic feature of sonorants, a class of speech sounds that includes vowels, approximants (e.g., /l/ and /ɹ/), 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 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). Because of the special role of formants for vowel perception, much of the research regarding the formant structure of speech uses vowel stimuli.

## PHYSIOLOGICAL 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 responses (FFRs) to three steady-state vowels were measured and the spectral content of the responses was 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 was greater than for 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) which also demonstrated a predominant representation of F1. 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- and 2,100-Hz puretones as well as a two-tone composite signal comprising the component puretones (i.e., simultaneous presentation of the 600- 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 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 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 has a different representation relative 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 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 naturally differ in F1–F2 relationships. Results indicated that the location of the N100m source reflects the relationship of the F1–F2 formant frequencies. 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).

Although these studies provide evidence that the cortex represents the formant structure of vowels in a manner that is (a) unrelated to the tonotopic map and (b) 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). Second, approximately six neural regions contribute to the N100 and therefore it represents a highly complex neural response. Consequently, the N100 described in these studies of phonemotopic maps should not be viewed as an exact representation of well-described, and highly localized, auditory maps in animal models (Schreiner, 1998). This is particularly relevant given that the clear tonotopic gradient in auditory cortex is no longer apparent when puretone stimuli are presented above 50 dB 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 if 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 auditory cortex as early as ~100 ms poststimulus 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 sounds, that is, laughs and coughs (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 original amplitude envelope of the speech signal modulated the altered spectral information. 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 structures of the superior temporal plane. 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. In addition, 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 et al., 2002).

An interesting consideration is how cortical asymmetries in response to the acoustic features of speech relate to well-established cerebral asymmetries for higher-order language processing, such as phonemic and semantic processing (Geschwind and Galaburda, 1985), which are strongly lateralized to the left hemisphere. Although 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 physiological 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.

## Electrophysiological Changes due to Training

Musical training can enhance the brainstem's representation of formant frequencies, and this enhancement is related to important aspects of speech perception. For example, it was recently shown that adult musicians have greater differentiation of brainstem responses for consonant–vowel stimuli that vary according to F2 frequency (Parbery-Clark et al., 2012; Strait et al., in press). Specifically, musicians showed more pronounced brainstem timing differences in response to /da/, /ga/, and /ba/ stimuli compared to nonmusicians, and brainstem differentiation of these stimuli correlated with standardized measures of speech perception in noise. This finding is important for a number of reasons. First, it shows that musicians' goal-directed attention to spectrotemporal features in music promotes neural differentiation of subtle variants in formant structure in speech as well as perceptual benefits for speech in noise. This result is also significant with regard to efficacy of therapy: Whereas many forms of auditory perceptual training fail to generalize to untrained stimuli (Burk and Humes, 2008; Halliday et al., 2012), results from the music literature have consistently shown that musical training generalizes to speech perception tasks in children (Moreno et al., 2009; Thompson et al., 2004) and adults (Schon et al., 2004; Thompson et al., 2004) as well as the neural encoding of speech (Moreno et al., 2009; Schon et al., 2004; Strait and Kraus, 2014). Importantly, results from the Parbery-Clark study show that musical training influences the neural differentiation of subtle formant frequency characteristics, which is fundamental to the identification and discrimination of phoneme contrasts (Peterson and Barney, 1952).

In summary, the brainstem encodes lower formant frequencies, which are critical to vowel perception, with phase-locked responses. Moreover, the representation of these formants is enhanced following long-term musical training, and the strength of these representations is correlated with perceptual benefit for speech in noise. 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 ~100 ms poststimulus onset as measured by N100m source location. In addition, functional imaging data provide evidence that the STS, a nonprimary auditory 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 associative 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 are ubiquitous in 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 for 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). Formant transitions have also 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): (a) The movement of a formant from the initiation of the consonant until the beginning of the vowel in a consonant–vowel combination, (b) the movement of a formant from one vowel to another vowel (i.e., in a diphthong), and (c) 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).

### Physiological Representation of Frequency Transitions in the Human Brain

#### Auditory Brainstem

The 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" was 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 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 the 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 acoustical cue for consonant identification.

A similar methodology was used in another study by Krishnan and colleagues to investigate the role of language experience on auditory brainstem encoding of pitch (Krishnan et al., 2005). FFRs to the "yi" stimuli described above were measured in native Mandarin speakers as well as native speakers of American English, to whom the pitch alterations 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.

An 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 of 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 efferent pathway-mediated tuning of afferent sensory input. Although the reverse hierarchy theory does not explicitly discuss plasticity of the brainstem, this theory could account for the findings of Krishnan. Specifically, because of the importance of extracting lexical information present in pitch contours, native Mandarin speakers are "experts" 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 required this level of modification.

An interesting question that was addressed in a subsequent study 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 (Xu et al., 2006). Results show that a lifetime of experience using F0 to extract linguistic meaning specifically affects auditory responses to familiar speech sounds and does not generalize to all periodic acoustic signals. However, data from the Kraus Lab suggests 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., 2007). This finding provides evidence that expertise associated with one type of acoustic signal (i.e., music) can provide 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. The 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 (a) Mandarin speech sounds that contained modulations of the fundamental frequency that signal lexical meaning and (b) English speech sounds which 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; 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 acoustical 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. As it is known that neurons in auditory cortex do not phase-lock to frequencies greater than approximately 100 Hz (Creutzfeldt et al., 1980), and the formant structure of speech consists of frequencies almost exclusively above 100 Hz, the cortical representation of frequency modulation as measured by evoked potentials is abstract (i.e., not represented with time-locked responses) relative to those 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 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 STG and 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, modulations in the fundamental frequency of speech are faithfully encoded in the FFR. Moreover, these 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, and 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, whereas the right-hemisphere insula is activated when F0 modulations do not provide lexical information. These cortical findings would appear to be contradictory: The former indicates asymmetric activation by left-hemisphere structures is based on physical parameters of the speech signal, irrespective of linguistic content, whereas 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.

### Electrophysiological Changes due to Training

There is ample evidence that multiple forms of auditory therapy and training have enhancing effects on the neural representation of frequency transitions in speech, including transitions of the fundamental and formant frequencies. Consistent with neural enhancement of formant structure discussed previously, musical training also strengthens brainstem representations of frequency transitions, including representations of both the fundamental and formant frequencies. As discussed previously, one study showed that adult musicians have enhanced brainstem representations in response to tonal permutations of the Mandarin word "mi," which are characterized by contours to the fundamental frequency (Wong et al., 2007). It is hypothesized that this neural benefit is the result of years of attention to pitch variations in musical stimuli, and again it is significant that this neural advantage generalizes from the music domain to speech. In another study, it was shown that musical training also enhances brainstem representations of formant transitions in speech. For example, young children (3 to 5 years of age) with at least a year of musical training showed earlier brainstem responses to the formant transition portion of a consonant–vowel stimulus compared to age-matched listeners, with the greatest effects of musicianship being evident in the presence of background noise (Strait et al., 2013). Studies examining other forms of auditory training have also shown strengthening of brainstem responses to formant transitions in speech. In one study, two groups of older adults (mean age = 62 years) participated in different training paradigms matched for time and computer use: One group was trained on an adaptive computer-based auditory training program that combined bottom-up perceptual discrimination exercises with top-down cognitive demands whereas an active control group was trained on a general educational stimulation program (Anderson et al., 2013). Results for the auditory training group showed improved resiliency of speech-evoked brainstem responses in background noise, and this resiliency was most pronounced for the formant transition period of the consonant–vowel stimulus. This neural effect in the auditory training group was accompanied by significant improvement in a number of auditory behavioral and cognitive measures, including speech in noise, auditory memory, and processing speed. Importantly, the active control group failed to show improvements on both the neural and behavioral measures. A third study examined brainstem plasticity for yet another type of auditory therapy, in this case the use of assistive listening devices for use by children with reading impairments in the classroom (Hornickel et al., 2012b). The theoretical basis for providing these listening devices to this population is that children with reading impairments have impaired speech perception in noise relative to age-matched children (Bradlow et al., 2003). Importantly, assistive listening devices provide substantial improvements with regard to the signal-to-noise ratio of the teacher voice relative to classroom background noise. Results from this study showed that after using assistive listening devices for one academic year, children with reading impairments showed greater consistency of brainstem responses in the formant transition period of a consonant–vowel stimulus. These children also showed behavioral improvements on standardized measures of phonologic processing and reading ability. A control group, composed of reading-impaired children who did not use assistive listening devices, failed to show improvements in either of these neural or behavioral measures.

Taken together, results from these studies show that the neural representation of frequency transitions in speech is highly malleable in response to very different kinds of auditory training, including musical training, adaptive auditory-based computer programs, and the use of assistive listening devices. This suggests that therapies that sharpen "top-down" brain mechanisms, such as goal-directed attention to auditory stimuli, and "bottom-up" signals, as provided by assistive listening devices, can focus and improve the efficiency of neural mechanisms serving the tracking of frequency modulations. Moreover, the relative abundance of studies showing training effects for neural responses of frequency transitions further suggests that the brain's representation of this acoustical feature is particularly plastic, reflecting a critical auditory mechanism underlying rapid improvement in important auditory skill acquisition.

## 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. Although 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 neurophysiological literature. Consequently, acoustic onsets are discussed here separately, despite some overlap with acoustic features (e.g., 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 F2 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 latter is that of the 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 primary acoustic cue that enables differentiation between consonants that are otherwise extremely similar (e.g., /da/ vs. /ta/, /ba/ vs. /pa/, /ga/ vs. /ka/).

### Physiological Representation of Acoustic Onsets in the Human Brain

#### Auditory Brainstem

The brainstem response to speech-sound onsets has been studied extensively (Banai et al., 2005; Russo et al., 2004;

Wible et al., 2004). The first components of the speech-evoked ABR reflect the onset of the stimulus (Figure 28.2). 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 function, including literacy. These studies have investigated brainstem responses to speech in normal children and children with language-based LDs, a population that has consistently demonstrated perceptual deficits in auditory tasks using both simple (Tallal and Piercy, 1973; Wright et al., 1997) and complex (Kraus et al., 1996; Tallal and Piercy, 1975) 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). 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 is an important neural indicator for distinguishing LD from typically developing (control) subjects. A number of studies have provided compelling evidence that the representation of speech onset is abnormal in a significant proportion of subjects with LD (Banai et al., 2009). For example, brainstem responses to the speech syllable /da/ were measured for a group of 33 normal children and 54 children with LD, and 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 subjects with LD compared to normal subjects. The authors of this study indicate that 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. 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. In yet another study, it was shown that the timing of children's brainstem onset responses to speech sounds correlated with standardized measures of reading and phonologic abilities across a wide range of reading abilities (Banai et al., 2009).

Taken together, these data suggest that the brainstem responses to acoustic transients can differentiate not only a subpopulation of LDs from normal subjects, but also within 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 sizable proportion of the LD population have led to the incorporation of this experimental paradigm into the clinical evaluation of LD and central auditory processing disorders.

### Auditory Cortex

Cortical encoding of spectral features of speech-sound onsets has been reported in the literature (Obleser et al., 2006) and indicates that a spectral contrast at speech onset, resulting from consonant place of articulation (i.e., front produced consonant /d/ or /t/ vs. back produced consonant /g/ or /k/), is mapped along the anterior–posterior axis in 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 contrasts, 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, many of which have utilized 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; Sharma et al., 2000). Additionally, perception of these phonetic contrasts was also measured using the same continua as a means of addressing whether cortical responses reflected categorical perception of the phonemes.

Neurophysiological 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. 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 interpeak latency of the two peaks ($r = -0.80$). The onset of double peaks in cortical responses with a VOT of 40 ms is consistent with neurophysiological responses measured directly from the auditory cortex of humans (Steinschneider et al., 1999), and an important consideration is that the onset of the double peak occurred at 40 ms for both /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/, whereas a VOT of ~60 ms was required for correct identification of /ga/ from /ka/. Taken together, these data indicate that cortical responses reflect the actual VOT irrespective of the categorical perception of the phonetic contrasts.

### 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 and complex acoustic stimuli at the level of the auditory cortex. Three studies have linked abnormal neural synchrony for acoustic transients at the auditory brainstem to abnormal representations of sounds in the cortex. In one study, 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 auditory cortex's susceptibility to background noise in both normal and LD subjects (Wible et al., 2005). Specifically, the longer the duration between onset peaks V and A, the more degraded the 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 (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). 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, whereas temporal attributes of speech onsets, as manifested by the VOT, are precisely encoded with double-peaked N100 responses.

### Electrophysiological Changes due to Training

A survey of the brainstem and cortical literatures indicates that there is relatively scant evidence that the brain's representation of acoustic onsets is malleable following auditory-based training and therapy, and the primary evidence for plasticity of this feature is from a study of very young children. This study, which was previously described in the Formant Transition section, showed that a year or more of musical training in young children (3 to 5 years of age) resulted in decreased brainstem onset latencies in response to a consonant–vowel stimulus (Strait et al., 2013). Sound onsets are considered to be particularly rudimentary sound features, and the fact that the brainstem's response to acoustical onsets does not appear to be plastic following training (except in very young children) strongly suggests that this neural feature is established early in development and remains largely static irrespective of the experience of the individual. However, subcortical encoding of acoustic onsets does undergo substantial developmental changes across the lifespan, irrespective of training (Anderson et al., 2012; Skoe et al., in press).

# The Speech Envelope

## Definition and Role in the Perception of Speech

The speech envelope refers to temporal fluctuations in the speech signal under 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), 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. The perceptual significance of the speech envelope has been investigated using a number of methodologies (Drullman et al., 1994; Shannon et al., 1995) and, taken together, these data indicate that the speech envelope is both necessary and sufficient for normal speech recognition.

## Physiological 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 neurophysiological activity below ~100 Hz (Hall, 1992). Since speech envelope frequencies are between 2 and 50 Hz, any linear representation of speech envelope timing in brainstem responses is removed with brainstem filtering.

In one EEG study, responses from the auditory cortex to conversational, clearly enunciated, and time-compressed (i.e., rapid) speech sentences were measured in children (Abrams et al., 2008). Results indicate that human cortex synchronizes its response to the contours of the speech envelope across all three speech conditions and that responses measured from right-hemisphere auditory cortex showed consistently greater phase-locking and response magnitude compared to left-hemisphere responses. An MEG study showed similar results; however, in this study, it was shown that these neurophysiological measures of speech envelope phase-locking correlated with subjects' ability to perceive the speech sentences: As speech sentences become more difficult to perceive, the ability of the cortex to phase-lock to the speech sentence was more impaired (Ahissar et al., 2001). These results are in concert with results from the animal literature, which show that neurons of 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).

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. Briefly, neurophysiological results indicated that for both /ba/-/pa/ and /ga/-/ka/ phonetic contrasts, cortical N100 responses precisely represented the acoustic attributes of the VOT. In addition, it was shown that neural responses were 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) whereas phase-locking fails to correlate with perception in the other study (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 ~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 to form phonetic category boundaries. The question of if and how category boundaries are established irrespective of auditory phase-locking requires additional investigation.

## CONCLUSIONS

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 to decode a linguistic message. Furthermore, how the human brain is innately and dynamically programmed to utilize 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.

The primary goals of this chapter are to describe our current understanding of neural representation of speech as well as training-related changes to these representations. By exploring these two topics concurrently it is argued that we have provided complementary perspectives on auditory function: The initial descriptions of brainstem and cortical representations of these speech features are thought to reflect "bottom-up" function of the auditory system with minimal consideration for the dynamic interactions provided by top-down connections in the auditory system (Xiao and Suga, 2002); in contrast, the descriptions of training-related changes to these representations provide information regarding how "top-down" cognitive and brain mechanisms sharpen these auditory representations (reviewed in Kraus and Chandrasekaran, 2010). Evidence accumulated across studies provides a complicated, but compelling, account of the malleability of these auditory responses. Results show that brainstem representations of speech can be affected and sharpened by multiple forms of auditory-based experiences, from long-term musical experiences to relatively short-term auditory-cognitive training paradigms. Importantly, the relative plasticity of these different speech features appears to fall on a continuum: Acoustic onsets, which are largely static following all forms of auditory training, occupy one end of this continuum, whereas neural representations of formant transitions occupy the other end of this continuum, showing enhanced response properties following multiple training paradigms measured in a wide range of subject populations. Consistent with the animal literature (Recanzone et al., 1993), it is plausible that the relative plasticity of these features reflects the behavioral demands of each form of training, and a prediction of this hypothesis is that relatively static neural representations do not significantly contribute to the improvement on these tasks whereas more dynamic neural representations are important for improved performance.

To garner a greater understanding of how the central auditory system processes speech, it is important to consider subcortical and cortical auditory regions as reciprocally interactive. Indeed, auditory processing reflects an interaction of sensory, cognitive, and reward systems. Across the acoustic features described in this review, the brainstem appears to represent discrete acoustic events: 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 and 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 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 be likely achieved using high-resolution imaging technologies in concert with EEG and MEG technologies.

## FOOD FOR THOUGHT

Here, we have described what is currently known about brain representations of key elements of speech that are necessary for normal speech perception. Our review covers information garnered from multiple research methodologies, including brainstem- and cortical-evoked responses using EEG, which provide crucial information regarding the neural timing in response to specific speech features, as well as fMRI research, which provides complementary information regarding "where" in the brain this activity occurs. Furthermore, we have described the relative plasticity of these brain responses as a result of specific behavioral experiences, with an emphasis on musical training. The following are important questions for future research that will enable us to further understand the brain basis of speech perception as well as associated plasticity and impairments.

1. Both the auditory brainstem and cortical regions are highly sensitive to elements of speech structure. An important question is what is the relationship between the integrity of brainstem representations of speech structure and cortical regions beyond auditory cortex that are known to be critical for structural processing of speech? For example, the posterior temporal sulcus is considered "voice-selective cortex" (Belin et al., 2000) and has been proposed to be a critical gateway which enables speech information to access other brain networks that serve semantic, reward, and mnemonic processes (Belin et al., 2011). A better understanding of how lower levels of the auditory hierarchy (i.e., the auditory brainstem) impact voice selectivity in the posterior temporal sulcus would provide important information regarding the function of this extensive network.
2. While humans are drawn to the sounds of speech, it is seldom considered a "rewarding" stimulus. Perhaps for this reason little research has been conducted to study the brain networks that are used for pleasurable speech. For example, what parts of the auditory hierarchy are differentially activated in response to pleasurable compared to neutral speech? Would these pleasurable speech sounds provide altered neural responses across the entire auditory hierarchy, or alternatively would only specific regions of the brain show effects of pleasure?
3. Research described in this chapter has convincingly shown that speech in noise perception is greatly improved through musical training (Parbery-Clark et al., 2012; Song et al., 2012). An exciting question is what are the particular neural mechanisms that enable this effect of musicianship? What aspects of musical training facilitate these behavioral advantages, and how might we harness this information to train individuals of all ages to become better listeners in noisy environments?

## ACKNOWLEDGMENTS

We thank Trent Nicol for his comments on a previous draft of this chapter. This work is supported by F32DC010322, DC011095, and the Mosbacher Foundation (DAA) and the Hugh Knowles Center, NIH RO1DC10016, R01HD069414, NSF 0921275, NSF1057566, and NSF1015614 (NK).

## REFERENCES

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

Abrams DA, Nicol T, Zecker SG, Kraus N. (2006) Auditory brainstem timing predicts cerebral asymmetry for speech. *J Neurosci.* 26, 11131–11137.

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, Hochstein S. (2004) The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci.* 8, 457–464.

Anderson S, Parbery-Clark A, Yi HG, Kraus N. (2011) A neural basis of speech-in-noise perception in older adults. *Ear Hear.* 32, 750–757.

Anderson S, Parbery-Clark A, White-Schwoch T, Kraus N. (2012) Aging affects neural precision of speech encoding. *J Neurosci.* 32 (41), 14156–14164.

Anderson S, White-Schwoch T, Parbery-Clark A, Kraus N. (2013) Reversal of age-related neural timing delays with training. *Proc Natl Acad Sci USA.* 110, 4357–4362.

Banai K, Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N. (2009) Reading and subcortical auditory function. *Cereb Cortex.* 19, 2699–2707.

Banai K, Nicol T, Zecker SG, Kraus N. (2005) Brainstem timing: Implications for cortical processing and literacy. *J Neurosci.* 25, 9850–9857.

Belin P, Bestelmeyer PE, Latinus M, Watson R. (2011) Understanding voice perception. *Br J Psychol.* 102, 711–725.

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, et al. (1998) Lateralization of speech and auditory temporal processing. *J Cogn Neurosci.* 10, 536–540.

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.

Burk MH, Humes LE. (2008) Effects of long-term training on aided speech-recognition performance in noise in older adults. *J Speech Lang Hear Res.* 51, 759–771.

Creutzfeldt O, Hellweg FC, Schreiner C. (1980) Thalamocortical transformation of responses to complex auditory stimuli. *Exp Brain Res.* 39, 87–104.

Cunningham J, Nicol T, King CD, Zecker SG, Kraus N. (2002) Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. *Hear Res.* 169, 97–111.

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.

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, MA: Allyn and Bacon.

Halliday LF, Taylor JL, Millward KE, Moore DR. (2012) Lack of generalization of auditory learning in typically developing children. *J Speech Lang Hear Res.* 55, 168–181.

Hornickel J, Anderson S, Skoe E, Yi HG, Kraus N. (2012a) Subcortical representation of speech fine structure relates to reading ability. *Neuroreport.* 23, 6–9.

Hornickel J, Zecker SG, Bradlow AR, Kraus N. (2012b) Assistive listening devices drive neuroplasticity in children with dyslexia. *Proc Natl Acad Sci USA.* 109, 16731–16736.

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.

Johnson K. (1997) *Acoustic and Auditory Phonetics.* Cambridge, MA: Blackwell Publishers Inc.

Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, Banerjee S. (2004) Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am.* 116, 2395–2405.

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.

Kraus N, Chandrasekaran B. (2010) Music training for the development of auditory skills. *Nat Rev Neurosci.* 11, 599–605.

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 J, Cariani P. (2005) Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res.* 25, 161–168.

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.

Lehiste I, Peterson GE. (1961) Transitions, glides, and diphthongs. *J Acoust Soc Am.* 33, 268–277.

Lin FR, Yaffe K, Xia J, Xue QL, Harris TB, Purchase-Helzner E, et al. (2013) Hearing loss and cognitive decline in older adults. *JAMA Intern Med.* 173, 293–299.

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.

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.

Moreno S, Marques C, Santos A, Santos M, Castro SL, Besson M. (2009) Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb Cortex.* 19, 712–723.

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.

Nearey TM, Assmann PF. (1986) Modeling the role of inherent spectral change in vowel identification. *J Acoust Soc Am.* 80, 1297–1308.

Nilsson M, Soli SD, Sullivan JA. (1994) Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am.* 95, 1085–1099.

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: Ohmasha 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.

Parbery-Clark A, Tierney A, Strait DL, Kraus N. (2012) Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience.* 219, 111–119.

Peterson GE, Barney HL. (1952) Control methods used in a study of the vowels. *J Acoust Soc Am.* 24, 175–184.

Recanzone GH, Schreiner CE, Merzenich MM. (1993) Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J Neurosci.* 13, 87–103.

Rosen S. (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci.* 336, 367–373.

Russo N, Nicol T, Musacchia G, Kraus N. (2004) Brainstem responses to speech syllables. *Clin Neurophysiol.* 115, 2021–2030.

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.

Schon D, Magne C, Besson M. (2004) The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology.* 41, 341–349.

Schreiner CE. (1998) Spatial distribution of responses to simple and complex sounds in the primary auditory cortex. *Audiol Neurootol.* 3, 104–122.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. (1995) Speech recognition with primarily temporal cues. *Science.* 270, 303–304.

Sharma A, Dorman M. (2000) Neurophysiologic correlates of cross-language phonetic perception. *J Acoust Soc Am.* 107, 2697–2703.

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, 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.

Skoe E, Kraus N. (2010) Auditory brain stem response to complex sounds: a tutorial. *Ear Hear.* 31 (3), 302–324.

Skoe E, Krizman J, Anderson S, Kraus N. (in press) Stability and plasticity of auditory brainstem function across the lifespan. *Cereb Cortex.* doi: 10.1093/cercor/bht311

Smith AJ, Blumenfeld H, Behar KL, Rothman DL, Shulman RG, Hyder F. (2002) 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.

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.

Song JH, Skoe E, Banai K, Kraus N. (2012) Training to improve hearing speech in noise: biological mechanisms. *Cereb Cortex.* 22, 1180–1190.

Steeneken HJ, Houtgast T. (1980) A physical method for measuring speech-transmission quality. *J Acoust Soc Am.* 67, 318–326.

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.

Strait DL, Kraus N. (2014) Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hear Res.* 308, 109–121.

Strait DL, O'Connell S, Parbery-Clark A, Kraus N. (in press) Musicians' enhanced neural differentiation of speech sounds arises early in life: developmental evidence from ages three to thirty. *Cereb Cortex.* doi:10.1093/cercor/bht103.

Strait DL, Parbery-Clark A, O'Connell S, Kraus N. (2013) Biological impact of preschool music classes on processing speech in noise. *Dev Cogn Neurosci.* 6, 51–60.

Sweetow RW, Sabes JH. (2006) The need for and development of an adaptive Listening and Communication Enhancement (LACE) program. *J Am Acad Audiol.* 17, 538–558.

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.

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.

Temple E, Poldrack RA, Protopapas A, Nagarajan S, Salz T, Tallal P, et al. (2000) Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. *Proc Natl Acad Sci USA.* 97, 13907–13912.

Thompson WF, Schellenberg EG, Husain G. (2004) Decoding speech prosody: do music lessons help? *Emotion.* 4, 46–64.

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.

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.

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.

Wong PC, Skoe E, Russo NM, Dees T, Kraus N. (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci.* 10, 420–422.

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.

Xu Y, Krishnan A, Gandour JT. (2006) Specificity of experience-dependent pitch representation in the brainstem. *Neuroreport.* 17, 1601–1605.

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