

# The scalp-recorded brainstem response to speech: Neural origins and plasticity

BHARATH CHANDRASEKARAN<sup>a,b,c</sup> AND NINA KRAUS<sup>a,b,d,e</sup>

<sup>a</sup>Auditory Neuroscience Laboratory, Northwestern University, Evanston, Illinois, USA

<sup>b</sup>Roxelyn and Richard Pepper Department of Communication Sciences, Northwestern University, Evanston, Illinois, USA

<sup>c</sup>Communication Neural Systems Research Group, Northwestern University, Evanston, Illinois, USA

<sup>d</sup>Department of Neurobiology and Physiology, Northwestern University, Evanston, Illinois, USA

<sup>e</sup>Department of Otolaryngology, Northwestern University, Evanston, Illinois, USA

## Abstract

Considerable progress has been made in our understanding of the remarkable fidelity with which the human auditory brainstem represents key acoustic features of the speech signal. The brainstem response to speech can be assessed noninvasively by examining scalp-recorded evoked potentials. Morphologically, two main components of the scalp-recorded brainstem response can be differentiated, a transient onset response and a sustained frequency-following response (FFR). Together, these two components are capable of conveying important segmental and suprasegmental information inherent in the typical speech syllable. Here we examine the putative neural sources of the scalp-recorded brainstem response and review recent evidence that demonstrates that the brainstem response to speech is dynamic in nature and malleable by experience. Finally, we propose a putative mechanism for experience-dependent plasticity at the level of the brainstem.

**Descriptors:** Language/speech, EEG/ERP, Sensation/perception

Speech is a stream of acoustic elements produced at an astounding average rate of three to six syllables per second (Laver, 1994). The ability to decode these elements in a meaningful manner is a complex task that involves multiple stages of neural processing. Models examining the neural bases of human speech perception have focused primarily on the cerebral cortex (Bennett & Hacker, 2006; Hickok & Poeppel, 2007; Näätänen, 2001; Poeppel & Hickok, 2004; Poeppel, Idsardi, & van Wassenhove, 2008; Scott & Johnsrude, 2003; Scott & Wise, 2004; Tervaniemi & Hugdahl, 2003). However, before speech can be perceived and integrated with long-term stored linguistic representations, relevant acoustic cues must be represented through a neural code and delivered to the auditory cortex with temporal and spectral precision by subcortical structures (Eggermont, 2001; Hickok & Poeppel, 2007; Poeppel & Hickok, 2004; Poeppel et al., 2008). Recent studies examining scalp-recorded evoked responses to speech stimuli have revealed that the auditory brainstem demonstrates considerable fidelity in representing the basic acoustic elements of speech (for a review, see Kraus & Nicol, 2005). These scalp-

recorded evoked responses thus offer a noninvasive method to test the integrity and functioning of subcortical structures in processing complex stimuli such as speech (Galbraith, Jhaveri, & Kuo, 1997; Glaser, Suter, Dasheiff, & Goldberg, 1976; Moushegian, Rupert, & Stillman, 1973; Russo, Nicol, Musacchia, & Kraus, 2004).

In the following review, we examine the morphology of the scalp-recorded brainstem response to speech stimuli. We then differentiate the scalp-recorded brainstem response from peripheral and cortical evoked responses. By reviewing recent studies that have demonstrated experience-dependent plasticity in the representation of various features of speech stimuli, we then shift the focus to the dynamic nature of the human auditory brainstem response. Finally, we explore a potential neurobiological mechanism that may underlie brainstem plasticity.

## Subcortical Encoding of Speech Features

The brainstem response to a consonant-vowel (CV) speech syllable is made up of two separate elements, the onset response and the frequency-following response (FFR; Akhoun et al., 2008; Russo et al., 2004). The onset response is a transient event that signals the beginning of the sound. In the case of consonants, the transient onset response marks the beginning portion of the consonant characterized by unvoiced, broadband friction (onset burst). The sustained FFR is synchronized to the periodicity

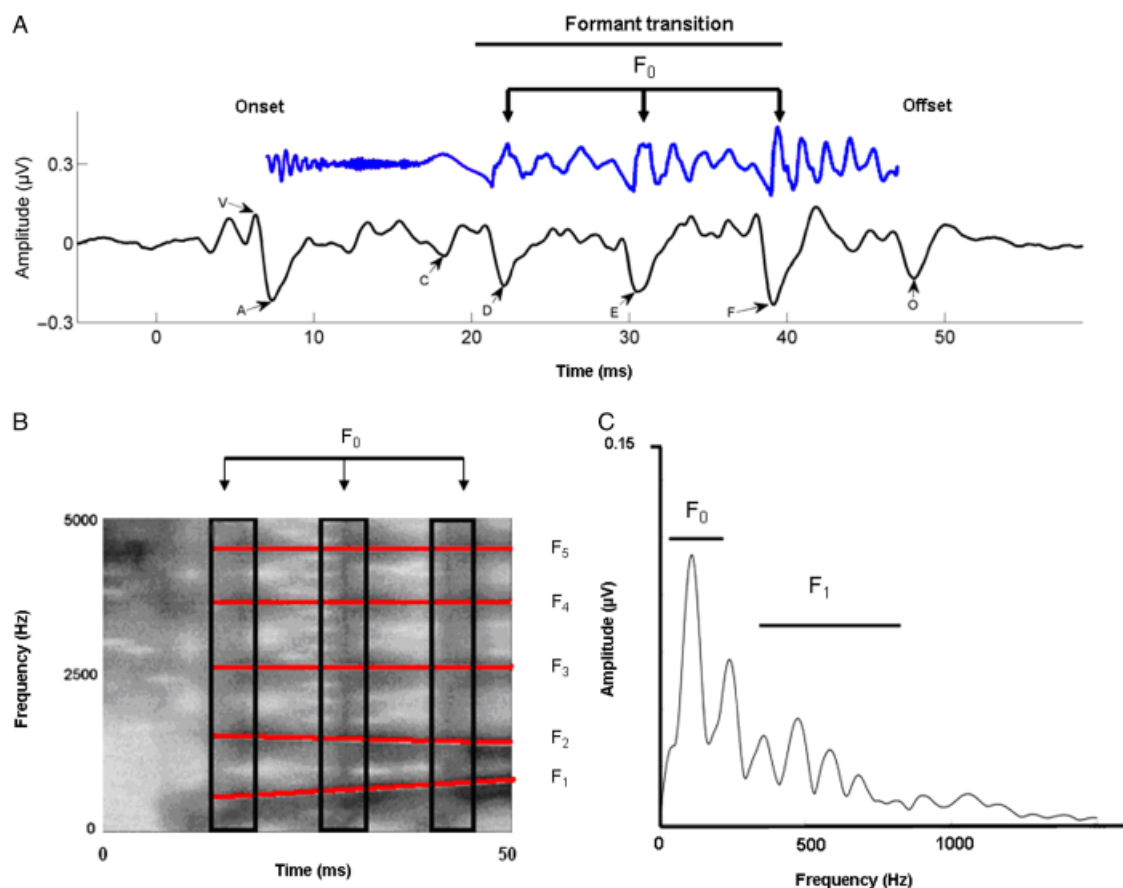
This work was supported by Grants NIH/NIDCD RO1-01510, F32DC008052, and NSF BCS-544846 and by the Hugh Knowles Center, Northwestern University. The authors acknowledge the anonymous reviewers for their useful comments. We also thank Trent Nicol, Erika Skoe, and Karen Banai for providing feedback on earlier versions of this article.

Address reprint requests to: Nina Kraus, Ph.D., Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA. E-mail: nkraus@northwestern.edu

(repeating aspects) of the sound, with each cycle faithfully representing the temporal structure of the sound. Thus the sustained FFR reflects neural phase-locking with an upper limit of about 1000 Hz. For a CV syllable such as /da/, the onset response corresponds with the burst release of the stop consonant, the sustained FFR response reflects the transition period between the burst and the onset of the vowel, and the vowel itself. The typical response to a 40-ms syllable “da” is shown in Figure 1. The speech syllable has a sharp onset burst, a short period of formant transition, and a longer period associated with the vowel /a/. The brainstem response to “da” preserves all the elements of the stimulus crucial to the recognition of the speech syllable, the intention with which it is spoken (e.g., emotion), and speaker identity. In the response, there is a clear negative peak (A) that follows Wave V of the auditory brainstem response (ABR) that occurs with a lag of 6–10 ms relative to the onset of the stimulus (reflecting transmission delay between the ear and the rostral brainstem structures). The formant transition period is marked by Wave C, which marks the change from the burst to the pe-

riodic portion of the syllable, that is, the vowel. Waves D, E, and F represent the periodic portion of the syllable from which the fundamental frequency of the stimulus can be extracted (see Figure 1a). Finally, Wave O marks the offset of the stimulus. The waves described are highly replicable and occur with remarkable (nearly 100%) reliability in all subjects (Russo et al., 2004).

Table 1 summarizes results from a number of studies that have examined brainstem responses (in time and frequency domains) to segmental (vowel, consonant) and suprasegmental features of speech. The representation of speech is so robust that when brainstem responses to words are converted to audio files and played to participants, they can be correctly identified by normal hearing participants with greater-than-chance accuracy (Galbraith, Arbagey, Branski, Comerci, & Rector, 1995). FFRs have been shown to represent the first (F1) and second formants (F2) in two-tone vowel approximations as well as synthesized speech (Krishnan, 1999, 2002), with peaks in the response spectrum adjacent to the first and second formant frequencies in the stimulus (Krishnan, 2002). With respect to stop consonants, the brainstem



**Figure 1.** A: Time-amplitude waveform of a 40-ms synthesized speech stimulus /da/ is shown in blue (time shifted by 6 ms to be comparable with the response). The first 10 ms are characterized by the onset burst of the consonant /d/; the following 30 ms are the formant transition to the vowel /a/. The time-amplitude waveform of the time-locked brainstem response to the 40-ms /da/ is shown below the stimulus, in black. The onset response (V) begins 6–10 ms following the stimulus, reflecting the time delay to the auditory brainstem. The formant transition period is marked by wave C that marks the change from the burst to the periodic portion of the syllable, that is, the vowel. Waves D, E, and F represent the periodic portion of the syllable (frequency-following response) from which the fundamental frequency ( $F_0$ ) of the stimulus can be extracted. Finally, wave O marks stimulus offset. B: Broadband spectrogram of the 40-ms /da/. Darker areas indicate regions of greater energy. For the benefit of the reader, formants 1–5 are marked with red lines. The burst is characterized by a high-frequency energy during the first 10 ms, followed by formant transition into the vowel. Relative spacing between F1 and F2 frequencies relate to vowel identity. The black rectangles highlight the vertical bands in the spectrogram, which reflect glottal pulses. C: Fast Fourier transform analysis of the brainstem response to the stimulus /da/. The spectrum reveals clear representation of the fundamental and its harmonics. Also represented is the time-varying F1 (as an increase in the amplitude of the peaks encompassing the F1 range).

**Table 1.** Auditory Brainstem Response to Speech Features

Phonological feature	Acoustic correlate	Brainstem response	Reference	
Segmental	Vowel	Spectral peaks in harmonics adjacent to F1 and F2 and 2F1-F2 (distortion product)	Krishnan (1999, 2002)	
	Stop consonant	Onset burst	ABR wave V followed by negative wave “A”	Russo et al. (2004) Akhoun et al. (2008) King et al. (2002) Cunningham et al. (2001) Banai et al. (2005, 2009)
		Formant transition	Wave C of the FFR reflects formant transition period FFR spectra reflects formant structure  Latency shifts in FFR peaks in the formant transition period	Banai et al. (2009) Russo et al. (2004) Kraus and Nicol (2005) Krishnan (2002) Akhoun et al. (2008) Johnson et al. (2008) Hornickel (2009)
Suprasegmental	Tone	F0, harmonics	Stimulus-to-response correlations show phase-locking to time-varying F0 and harmonics	Krishnan et al. (2004, 2005) Xu et al. (2006) Wong et al. (2007) Song et al. (2008) Musacchia et al. (2007)
	Prosody, emotion	Paralinguistic F0, harmonics	Stimulus-to-response correlations show phase-locking to time-varying F0 and harmonics	Russo et al. (2008) Strait et al. (2009)

onset responses including Waves V and A reflect the onset of the burst (Akhoun et al., 2008; Russo et al., 2004). The first formant is clearly represented in the FFR spectrum, revealing a peak in the range consistent with the time-varying F1 (Figure 1c). Over 90% of participants demonstrated earlier latencies in the response peaks for /ga/ (which has the highest F2 and F3), relative to /da/ and /ba/, within the formant transition period of the FFR (Johnson et al., 2008a; Hornickel et al., 2009). Thus F2 and F3, despite being outside the phase-locking capabilities of the auditory brainstem, are still reflected in the response as timing differences. These latency differences may be driven by differential stimulation of the basilar membrane as a function of the frequency content of the consonant (Ulfendahl, 1997).

Recent studies have also demonstrated robust representation of time-varying F0 and harmonics in the FFR (Krishnan, Swaminathan, & Gandour, 2008; Krishnan, Xu, Gandour, & Cariani, 2004, 2005; Song, Skoe, Wong, & Kraus, 2008; Wong, Skoe, Russo, Dees, & Kraus, 2007). Time-varying F0 and harmonics are linguistically relevant in tonal languages such as Mandarin Chinese and Thai (Yip, 2002). Stimulus-to-response correlations and response autocorrelation functions suggest that the ability to represent time-varying pitch contours is present even in nontonal language speakers, although their pitch representation is poorer relative to tone-language speakers (Krishnan et al., 2005). In general, the ability to track time-varying pitch may be beneficial in speaker identification and processing speech prosody (Russo et al., 2008).

Taken together, the two key elements of the brainstem response to speech, the onset response and the FFR, represent speech features with remarkable fidelity. Both elements are recorded from the scalp and are presumed to faithfully reflect activity from an ensemble of neural elements within the central auditory pathway. It has been argued that these two elements of

the brainstem response may reflect different neural streams within the brainstem nuclei (Akhoun et al., 2008). In a study examining the brainstem response to the syllable /ba/, Akhoun et al. found the response characteristics of the onset and the sustained portions differed with intensity. Both the FFR and the onset latencies shifted with increased stimulus intensity (i.e., earlier latencies), but to different extents, with the FFR showing a greater rate of change. Furthermore, under noisy listening conditions, the FFR portion that corresponds with the vowel appears to be less affected than the onset and offset response (Cunningham, Nicol, King, Zecker, & Kraus, 2002; Russo et al., 2004). Since the auditory brainstem response represents both the source (F0) and filter (onset, offset, and formant transition) characteristics of speech signals, it has been proposed that the “what” and “where” cortical processing streams (Belin & Zatorre, 2000; Kaas & Hackett, 1999; Lomber & Malhotra, 2008; Romanski et al., 1999) may have brainstem origins (Kraus & Nicol, 2005). From a clinical perspective, it has been shown that the auditory brainstem representation of the filter may be more impaired in some individuals with learning impairments such as developmental dyslexia (Hornickel et al., 2009; Banai et al., 2009; Banai, Nicol, Zecker, & Kraus, 2005; Cunningham, Nicol, Zecker, Bradlow, & Kraus, 2001; King, Warrier, Hayes, & Kraus, 2002; Wible, Nicol, & Kraus, 2004) than the representation of the source. Interestingly, the reverse trend is seen in children with autism spectrum disorders (Russo et al., 2008). One of the speech-related symptoms of autism spectrum disorders is impaired speech prosody. In a subset of this clinical population, it has been demonstrated that the representation of F0 information in the brainstem response is impaired (Russo et al., 2008).

There are, hence, significant theoretical and clinical motivations for understanding the sources of the scalp-recorded brainstem response to speech stimulation. In the next few sec-

tions, we examine the sources of the sustained brainstem response. Much is known about the origin of the brainstem onset responses (ABR Waves I to V) to click stimuli (Hood, 1998; Jewett, 1994; Jewett, Romano, & Williston, 1970; Jewett & Williston, 1971). Although the onset response to complex stimuli such as speech has been less studied, given the response latency (typically 5–10 ms), it is clear that these responses are of brainstem origin. The current review focuses on examining the neural sources of the sustained FFR that mimics the periodicity of the input stimulus (see Figure 2 for a schematic of brainstem and cortical structures of the auditory pathway).

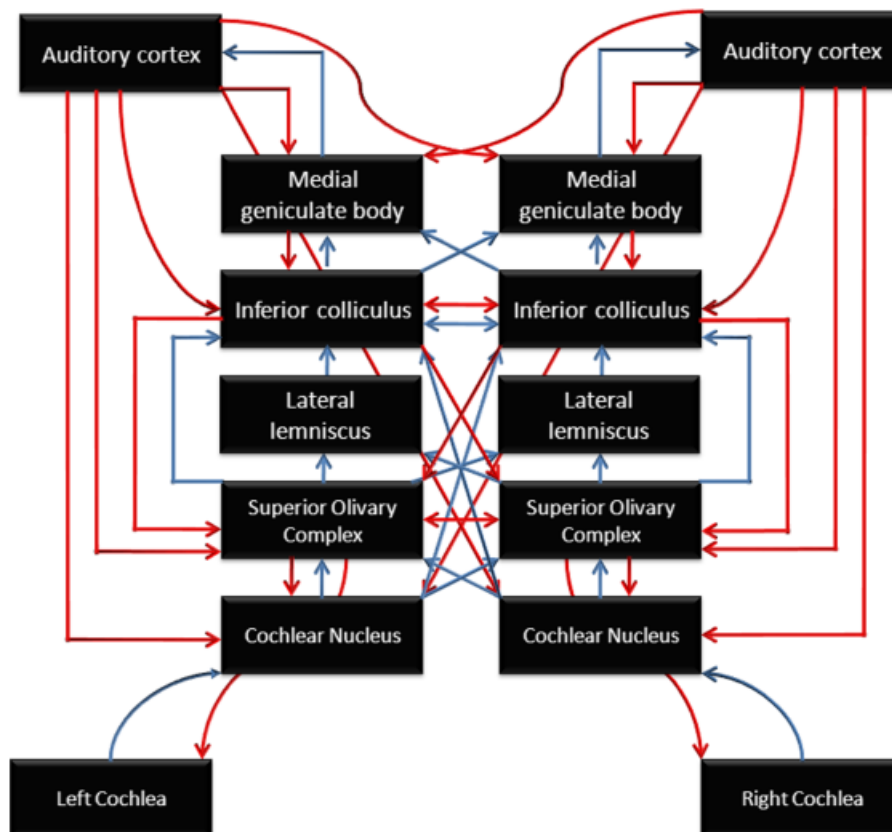
Although it is generally agreed that there are multiple generators of the scalp-recorded FFR (Galbraith, 1994; Stillman, Crow, & Moushegian, 1978), there is much less consensus on the exact sources of the FFR (Gardi, Merzenich, & McKean, 1979; Moushegian et al., 1973; Smith, Marsh, & Brown, 1975; Sohmer, Pratt, & Kinarti, 1977).

### Origins of the FFR: Ruling Out the Cochlea

The initial experiments examining the FFR sought to delineate this evoked activity from the cochlear microphonic (CM), the preneural, electrical potential originating from the hair cells in response to acoustic stimulation (Wever & Bray, 1930). Although FFRs, like the CM, accurately reproduce input acoustic stimulation, there are clear differences between the two potentials.

Worden and Marsh (1968) provided several lines of evidence that argue for a neural (as opposed to a preneural) basis for the FFRs. The onset of the FFR, unlike the CM, shows a delay of 5–10 ms even for simple sinusoidal tones, suggesting a site of origin rostral to the cochlea. Allaying doubts that the FFR simply reflected stimulus-related artifacts, Moushegian and colleagues (1973) argued that, for a typical ear-canal length of 2.7 cm, stimulus artifact would be generated at a latency of 0.029 ms, much earlier than the typical FFR latency of 5–10 ms. Also, the FFR (unlike CM) demonstrates small but appreciable amplitude and phase fluctuations that suggest that the responses are not a perfect replica of the input stimulus (Worden & Marsh, 1968). Further, a precise phase correspondence is seen between the scalp-recorded FFRs and unit activities in the cochlear nucleus (CN), trapezoid body, and superior olivary complex (SOC) in cats, suggesting that the FFR is an ensemble response reflecting phase-locked activity from multiple generator sites within the auditory brainstem (Marsh, Brown, & Smith, 1974). The CM can still be recorded under anoxia; the FFR shows reductions in amplitudes consistent with other neural evoked responses. Also, the CM is not sensitive to changes in rate of stimulation; the FFR shows latency shifts with increasing rates (Worden & Marsh, 1968).

In an article provocatively titled “Auditory Frequency-Following Responses: Neural or Artifact?” Marsh, Worden, and Smith (1970) used near-field recording techniques to demonstrate that sectioning the auditory cranial nerve eliminated FFRs but preserved the CM. Likewise, FFRs recorded from the CN



**Figure 2.** Schematic illustration of the auditory system. Blue arrows correspond to the ascending (bottom-up) pathways; red arrows correspond to the descending projections. The frequency-following response (FFR) reflects ensemble phase-locked responses from a number of subcortical auditory structures, including the cochlear nucleus, superior olivary complex, lateral lemniscus, and the inferior colliculus. Although the putative sources are subcortical, the FFR can be reliably recorded from the scalp.

using a cryoprobe were eliminated when the CN was cooled, and completely recovered when the temperature returned to normal; the CM was unaffected. Further, Marsh et al. (1970) observed binaural interaction (larger FFR amplitude relative to monaural stimulation) in the SOC, the site rostral to the CN in the auditory pathway. When the left CN was cooled, the response at the SOC was similar to right monaural stimulation. Taken together, these early studies supported a neural origin for the FFR and clearly delineated the phase-locked activity reflected by the FFR from CM or stimulus-artifact-related activities. Of more direct relevance to the scalp-recording technique used in humans, Smith and colleagues (1975) showed a drastic reduction of the scalp-recorded FFR in cats when the inferior colliculus (IC) was cooled. The phase-locked activity in the SOC (caudal to the IC) was preserved. The scalp-recorded FFRs regained their original amplitude when the IC was warmed. Further, depth recordings at the IC showed a mean latency shift of 5.2 ms, which was comparable with that of the FFRs from the scalp. In concordance with these data, no scalp-recorded FFRs were elicited from human participants with upper brainstem lesions (Sohmer et al., 1977). In contrast, the cochlear microphonic potential was still recordable. Lesion experiments by Gardi, Merzenich, et al. (1979) suggest that ablating the cochlear nucleus caused the largest reduction in the amplitude of the scalp-recorded FFR.

Two distinct pathways from the cochlear nucleus to the IC have been implicated in the generation of the FFR; a direct pathway to the contralateral IC via the lateral lemniscus (LL), and an ipsilateral pathway via the SOC and the LL (Marsh et al., 1974). The above mentioned studies suggest that the scalp-recorded FFR reflects activity from multiple generator sites in the brainstem. To reconcile the different brainstem generators, Stillman et al. (1978) utilized horizontal (earlobe to earlobe) and vertical (vertex to earlobe) electrode montages to examine the FFRs in human participants. They delineated two different frequency-following potentials, FFP1 and FFP2, in the FFR responses. FFP1 was represented well by both electrode montages; FFP2 was represented well by only the vertical montage. Stillman and colleagues proposed that the two electrode mon-

tages reflected phase-locked neural activities from different regions of the brainstem. Recording FFRs from the two montages in a missing-fundamental experiment, Galbraith (1994) found that the FFRs generated by the vertical montage represented the missing fundamental; the FFRs from the horizontal montage did not. Based on these findings, Galbraith et al. suggest that the missing fundamental is created before the sound reaches the auditory cortex, but not at the caudal brainstem structures (cochlear nucleus). They thus implicate the rostral brainstem structures in the generation of the missing fundamental. Taken together, these experiments suggest that the horizontal montage reflects more caudal brainstem structures (presumably the CN); the vertical montage reflects more rostral brainstem activity (presumably the lateral lemniscus or IC). Table 2 summarizes key differences between the cochlear microphonic and the auditory brainstem response.

### Origins of the FFR: Ruling Out the Cortex

Most experiments examining the FFR have opted for the vertical montage (vertex to mastoid or earlobe) because the responses are more robust relative to the horizontal montage. Even though the FFRs are recorded from vertex (Cz), there are numerous reasons that suggest that the scalp-recorded FFRs reflect brainstem activity rather than cortical activity. Across the board, studies that have examined processing of speech syllables using the FFR report amplitudes in the range of nanovolts. To our knowledge, no studies on humans have reported scalp-recorded FFR amplitudes greater than 1  $\mu$ V. In a comprehensive study of FFR amplitudes across various stimulus frequencies, Hoorman, Falkenstein, Hohnsbein, and Blanke (1992) found that the FFR amplitudes were largest between 320 and 380 Hz. The mean FFR amplitude in these frequency ranges was about 400 nV. In contrast, typical cortical responses are much larger, in the range of microvolts. It can be argued that the typical FFR recording site (vertex) is far away from the brainstem, and hence the responses are smaller in size relative to cortical responses, whose source is much closer to the scalp.

**Table 2.** Distinctions among Cochlear Microphonics (CM), Auditory Brainstem Responses, and Cortical Evoked Potentials

	CM	ABR/FFR	Cortical EPs
Origin	Preneural, cochlea	CN, LL, IC	Cortex, may also reflect MGB activity
Recording characteristics			
Polarity	Eliminated by alternating polarity	Responses present to alternating polarity	Responses present to alternating polarity
Rate	Unaffected by increasing rate	Latency shifts with stimulus rate	Eliminated at fast rates
Stimulus level	No latency shifts	Latency shifts with stimulus intensity	Latency shifts with stimulus intensity
Recording montage	Within ear canal	Two-channel (horizontal, vertical)	Multiple recording channels
Response characteristics			
Fidelity	Reflects stimulus	Reflects stimulus fine-structure and envelope	Reflects gross stimulus envelope
Onset latency	< 1 ms	5–10 ms	> 50 ms
Size (range)	Microvolts	Nanovolts	Microvolts
Latency variability	Not variable	Normal variability of < 1 ms	Large variability in latency (10–25 ms)
Maturation	Very early maturation	Adultlike responses by school age	Protracted development, not adultlike until late adolescence
Subject characteristics			
Arousal	Unaffected by subject state	Can be recorded in sleeping subjects	Reduced or eliminated in sleeping subjects
Attention	Unaffected by attention	Largely unaffected by attention	Attention-modulated
Plasticity	Unaffected by experience	Experience modulates responses	Experience modulates responses

CN: cochlear nucleus, LL: lateral lemniscus, IC: inferior colliculus, MGB: medial geniculate body (see Figure 2).

Due to the differences in size between cortical and brainstem responses, cortical responses require only about 75–100 averages to gain the typical morphology, whereas the scalp-recorded FFRs need >1,000 averages. Cortical responses are known to reduce in amplitude with stimulus repetition, a phenomenon termed repetition suppression or neural adaptation (Grill-Spector, Henson, & Martin, 2006). The FFR, on the other hand, is highly stable even with thousands of stimulus repetitions (Johnson et al., 2008b). These differences are in line with evidence that suggests that at the single neuron level, stimulus-specific adaptation is seen more in cortical neurons than subcortical neurons (Ulanovsky, Las, & Nelken, 2003). The FFR is highly repeatable. Peaks in the brainstem response that represent key features in the complex syllable /da/ are present in almost all individuals (Johnson et al., 2008b; Russo et al., 2004). Latency variability of 10–25 ms would be considered normal variability in cortical responses; variability of less than 1 ms in the auditory brainstem response to speech syllables has been associated with learning problems as well as difficulties in perceiving speech-in-noise (Banai et al., 2005; Cunningham et al., 2001; King, Nicol, McGee, & Kraus, 1999; Wible et al., 2004; Cunningham et al., 2002; Banai et al., 2009; Hornickel et al., 2009). The high repeatability of the brainstem response to speech syllables is in line with the idea that as one ascends up the auditory pathway, responses show less fidelity and more selectivity (Langner, 1992). Indeed, response properties of neurons in the auditory cortex are exceedingly complex, responsive to abstract categories such as conspecific vocalizations (Suga, O'Neill, Kujirai, & Manabe, 1983). Earlier auditory structures, on the other hand, represent the incoming stimuli as faithfully as possible (Langner, 1992; Nelken, 2004; Wang, Lu, Bendor, & Bartlett, 2008). Fine temporal precision at the level of the auditory cortex may not be advantageous because temporal precision would impair viewing long duration sounds (such as syllables) holistically in order to extract key information-bearing elements (Lu, Liang, & Wang, 2001; Wang et al., 2008). In fact, the complexity in the response property of auditory cortical neurons has led researchers to suggest that the auditory analog of the primary visual area (V1), which responds to simple visual features, may be the IC and not the cortex (Nelken, 2004).

A consistent finding across multiple researchers working with a variety of species is that the upper limit of temporal precision in representation reduces with each ascending step in the auditory pathway (Lu et al., 2001; Wang et al., 2008). Speech is characterized by fast temporal transitions. The ability of neurons to follow fast modulations reduces with each ascending auditory station (Langner, 1992). This reduction in temporal resolution may be related to temporal jitter due to multiple synapses with increasing ascent in the pathway (Burkard, Don, & Eggermont, 2006). In the auditory nerve, the upper limit of neural phase-locking varies from >5 kHz in cats to 3.5 kHz in guinea pigs (Johnson, 1980; Palmer & Russell, 1986). Units in the ventral cochlear nucleus can phase-lock up to 2 and 3.5 kHz, depending on the neural population (Winter & Palmer, 1990). In the guinea pig IC, there appears to be considerable variability, with units in some neural population phase-locking up to 1000 Hz (central nucleus); others show maximal phase-locking at 700 Hz (dorsal cortex); still others in the external nucleus can phase-lock only to about 320 Hz (Liu, Palmer, & Wallace, 2006). At the auditory cortex, recent studies have reported units capable of phase-locking up to about 250 Hz in anesthetized guinea pigs (Wallace, Shackleton, Anderson, & Palmer, 2005; Wallace, Shackleton, & Palmer, 2002). In awake monkeys, on the other hand, the upper

limit for cortical phase-locking has been shown to be closer to 100 Hz (Steinschneider, Arezzo, & Vaughan, 1980; Steinschneider, Fishman, & Arezzo, 2008). Although there are units in the cortex that appear to be capable of following the fundamental frequency in speech, there are fewer such units capable of synchronization, relative to earlier auditory structures (Bartlett & Wang, 2007; Wang et al., 2008). In fact, at each step of the auditory pathway, the number of units capable of synchronization reduces; at the medial geniculate nucleus (MGB) and the cortex, there are greater numbers of nonsynchronized and mixed units relative to synchronized units (Bartlett & Wang, 2007; Lu et al., 2001). This pattern is reversed at the level of the inferior colliculus, where there is preponderance of synchronized units (Batra, Kuwada, & Stanford, 1989). A recent study attempted to model the ABR and FFR by convolving the unitary brainstem response with discharge patterns from the auditory nerve model (Dau, 2003). Dau found that the modeled FFR at high intensity levels closely resembled responses elicited from human participants. This suggests that the scalp-recorded FFRs to simple stimuli can be predicted largely by examining the operating characteristics of the auditory periphery exclusively.

A universal aspect of neural maturation is that central structures take a longer time to mature relative to the periphery. In accordance with this concept, reliable, reasonably adultlike FFRs can be measured from neonates (Gardi, Salamy, & Mendelson, 1979). Although cortical evoked potentials (EPs) can be measured from infants, they do not resemble adult EPs. In fact, cortical potentials show a protracted development and do not completely mature until late adolescence (Sussman, Steinschneider, Gumenyuk, Grushko, & Lawson, 2008; Suzuki & Hirabayashi, 1987). Few studies have comprehensively examined the development of FFRs. A recent study compared click-evoked and speech-evoked responses from 3–4-year-olds and school-aged children (Johnson, Nicol, Zecker, & Kraus, 2008). They found that the 3–4-year-olds differed from older children more in their brainstem response to speech stimuli relative to click stimuli. Johnson, Nicol, Zecker, and Kraus (2008) suggested that the developmental trajectory for the FFR responses to speech may be influenced by the maturity of the corticofugal pathway. However, from a morphological perspective, the major peaks to the syllable /da/ are clearly preserved in both age groups. Cortical potentials do not show adultlike morphology even by adolescence (Sussman et al., 2008).

In summary, compared to cortical potentials, the scalp-recorded FFR is highly consistent, smaller in amplitude, less susceptible to adaptation with repetition, and demonstrates earlier maturation. Based on these facts, we conclude that the scalp-recorded FFRs most likely reflect brainstem sources. Furthermore, the FFR is capable of robustly representing the temporal structure in speech with a resolution on the order of 1 ms. From a neurophysiologic perspective, given that the auditory system shows a reduction in temporal resolution and in the number of units capable of synchronized responses with each ascending level, it is indeed reasonable to assume that the scalp-recorded FFR is of brainstem, rather than cortical, origin.

## Plasticity in the Brainstem Representation of Speech Features

### *Effect of Long-Term Experience*

Recent studies have demonstrated malleability in the brain stem representation of speech (for a review, see Kraus & Banai, 2005). Long-term and short-term auditory experiences have been

**Table 3.** *Experience-Dependent Plasticity in the Brainstem Response to Speech*

Type of experience	Brainstem plasticity	Reference
Short-term	Quiet-to-noise correlations of the FFR portion to /da/ increase following auditory perceptual training (children with learning impairments).	Russo et al. (2005)
Short-term	Improved brainstem representation of time-varying dipping pitch contour following sound-to-meaning auditory training.	Song et al. (2007)
Long-term	Enhanced brainstem representation of linguistic pitch contours (F0, harmonics) in tone language speakers (Mandarin) relative to nontone speakers (English).	Krishnan et al. (2005)
Long-term	Enhanced brainstem representation of linguistic pitch contours (stimulus-to-response correlation, autocorrelation magnitude) in musicians relative to nonmusicians.	Wong et al. (2007)
Long-term	Enhanced brainstem representation of pitch (larger spectral amplitude of F0) and faster timing in musicians relative to nonmusicians.	Musacchia et al. (2007)
Long-term	Enhanced representation of pitch (F0, harmonics) in older children (5–12-year-olds) relative to younger children (3–4-year-olds).	Johnson et al. (2008)
Long-term	Robust representation of acoustically complex portion of infant cry in musicians relative to nonmusicians.	Strait et al. (2009)

shown to enhance the brainstem responses to complex, behaviorally relevant sounds. Table 3 summarizes key results from a number of such studies. Krishnan et al. (2005), in a cross-language experiment, showed that long-term experience with linguistic pitch contours enhanced pitch representation as reflected by the FFRs. These authors found that native speakers of Mandarin had significantly better brainstem representation of linguistic pitch contours relative to native American English speakers. Such plasticity appears to be highly specific to the nature of the long-term experience, as only naturally occurring Mandarin tones and not linear approximations have been shown to elicit experience-dependent effects in native speakers (Xu, Krishnan, & Gandour, 2006). Furthermore, plasticity is not selective to speech stimuli as long as linguistic relevance is still maintained (Krishnan et al., 2008). Krishnan et al. (2008) conducted a cross-language study using iterative ripple noise (IRN) to simulate Mandarin tones. The IRN stimuli preserved the complexity of pitch information in the signal while being sufficiently nonspeech. Relative to English speakers, Mandarin participants represented pitch better at the level of the brainstem, suggesting that brainstem plasticity is not specific to speech. Rather, plasticity was found to be specific to dimensions that occurred in natural speech (Swaminathan, Krishnan, & Gandour, 2008).

Long-term experience with music has also been shown to provide an advantage across domains, that is, in the cortical (Chandrasekaran, Krishnan, & Gandour, 2009; Schon, Magne, & Besson, 2004) and brainstem representation of speech (Musacchia, Sams, Skoe, & Kraus, 2007; Strait, Skoe, Kraus, &

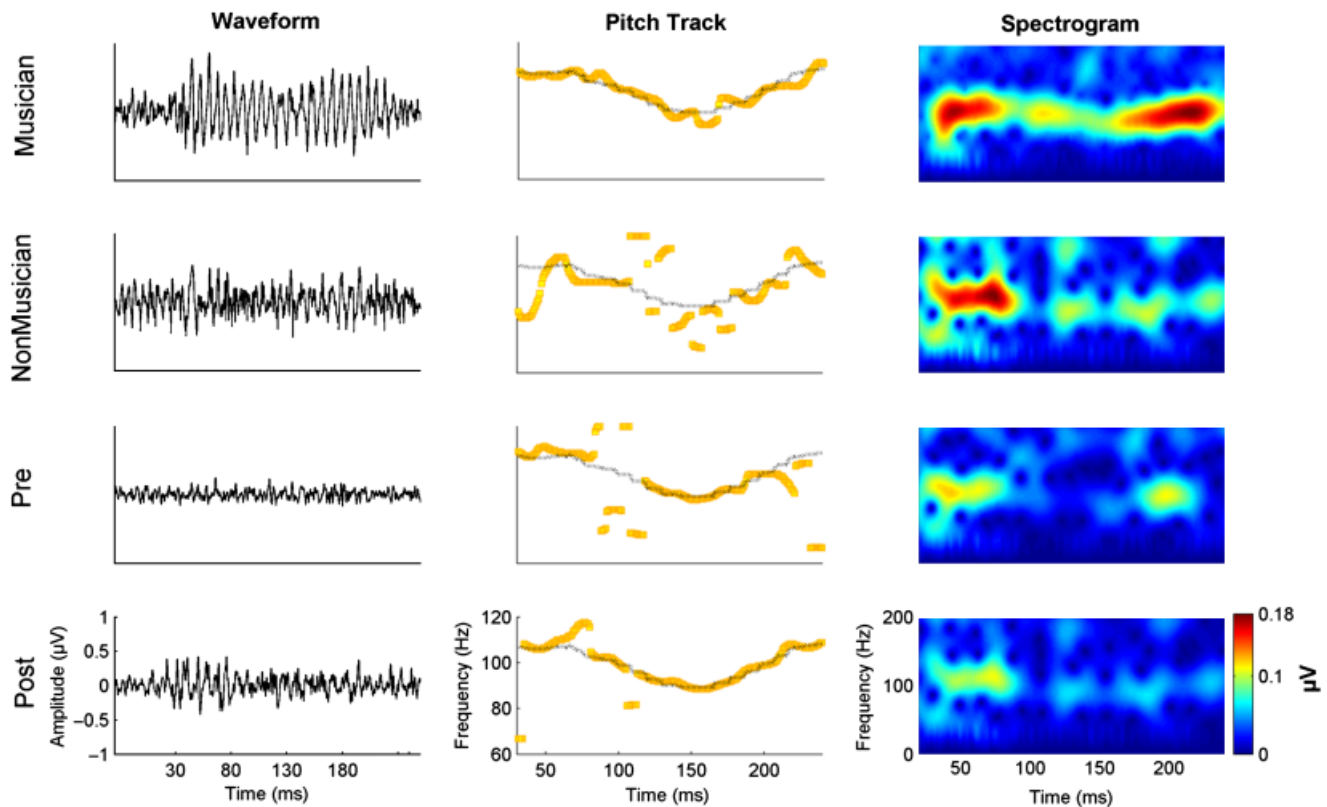
Ashley, 2009; Wong et al., 2007). Recent studies have compared FFRs from musicians and nonmusicians and showed an advantage for musicians in processing native speech sounds (Musacchia et al., 2007), nonnative linguistic pitch contours (Wong et al., 2007), and emotionally salient vocal sounds (Strait et al., 2009). These studies clearly demonstrate that plasticity at the level of the brainstem is not specific to the context of the long-term experience (see Figure 3).

#### *Effect of Short-Term Experience*

Short-term auditory training has been shown to improve the timing of the FFR to the syllable /da/. Children with learning problems who underwent an auditory training program exhibited brainstem responses that were more resistant to the deleterious effects of background noise (Russo, Nicol, Zecker, Hayes, & Kraus, 2005). A more recent study examined whether short-term training improves brainstem representation of lexical pitch contours (Song et al., 2008). Non-Mandarin-speaking participants in the Song et al. study underwent a short-term word-learning training program in which they learned to lexically incorporate Mandarin pitch contours embedded in nonwords. FFRs were recorded before and after training. The eight-session training program significantly improved the brainstem representation of the Mandarin dipping tone, suggesting that the adult brainstem is indeed malleable to short-term training (see Figure 3).

#### *Mechanisms Underlying Experience-Dependent Plasticity*

The outstanding question in all the above-mentioned studies is the neurobiological mechanism that underlies short-term and long-term brainstem plasticity. Some of these studies have implicated corticofugal tuning as a putative mechanism underlying brainstem plasticity, but such a proposal is difficult to test using noninvasive methods. Yet, there are good reasons to implicate a corticofugal tuning mechanism. First, there are massive efferent connections from the cortex to subcortical structures that could form the basis of feedback-related top-down projections (Kral & Eggermont, 2007). Efferent connections exist between layers of the auditory cortex that provide excitatory and inhibitory control over the inferior colliculus (Keuroghlian & Knudsen, 2007). Repeated stimulation by behaviorally relevant stimuli (Chowdhury & Suga, 2000), electrical stimulation of forebrain structures (Ma & Suga, 2008; Zhang & Suga, 2005), and auditory fear conditioning (Gao & Suga, 2000) have all been shown to induce plastic changes to the neuronal response properties in the IC in animals (for a review, see Suga, 2008; Suga, Xiao, Ma, & Ji, 2002). Importantly, these collicular changes are restricted when the forebrain structures are inactivated, suggesting that some kind of corticocollicular tuning shapes response properties of the IC. Cooling the cortex has dramatic effects on the response properties of collicular neurons (Nakamoto, Jones, & Palmer, 2008). The dorsal and external cortices of the inferior colliculus show longer response latencies relative to the central nucleus in guinea pigs (Liu et al., 2006). These authors suggest that the longer response latencies for neurons in dorsal and external cortices of the IC may reflect influences of the forebrain structures (MGB, auditory cortex). Indeed, consistent with this idea, the expression of protein Fos was found to be reduced in the dorsal and external cortices when the cortex was inhibited; no change in Fos expression was found in the central nucleus (Sun et al., 2007), suggesting corticofugal modulation of the dorsal and external cortices of the IC.



**Figure 3.** Effect of long-term and short-term auditory experiences on the frequency-following response. Frequency-following responses are from representative subjects in Wong et al. (2007) and Song et al. (2008). The left column shows FFR waveforms; the middle row shows trajectories (yellow line) of brainstem pitch tracking elicited by the same tone from the same subjects. The black line indicates the stimulus pitch contour. The right-most column shows spectrograms of the response from the same subjects. The top two rows are FFR responses from a musician (top) and nonmusician (bottom) elicited by a dipping pitch contour (Tone 3) in Wong et al. For the musician, the FFR waveform is more periodic; pitch track and spectrogram show that the fundamental frequency of the response closely follows the time-varying pitch contour. In contrast, the nonmusicians' pitch track and spectrogram show more deviations from the period of the F0 of the stimulus. Wong et al. showed that long-term experience with music modulates the brainstem representation of nonnative pitch contours. The bottom two rows are FFR responses to Tone 3, elicited from a representative subject in Song et al. obtained before and after a behavioral tone-learning paradigm. Relative to pretraining FFR (third row), the posttraining response (fourth row) shows a more faithful pitch track to the time-varying pitch contour. Song et al. demonstrated that even in the adult auditory brainstem, processing is enhanced by short-term auditory training.

Taken together, these animal studies strongly support the view that corticofugal modulation changes the neuronal properties of subcortical structures in a behaviorally relevant manner. Kral and Eggermont (2007) suggested that such top-down-driven corticofugal control mechanisms may drive plasticity after the critical period in development. According to these authors, the potential of bottom-up-driven plasticity reduces with auditory development, as synaptic mechanisms become less labile. Auditory representations are formed during this critical period in development, and these higher-order representations can guide plasticity in a top-down manner. This idea is consistent with the reverse-hierarchical theory in perception that suggests that there is a top-down-guided search for increased resolution in experts (Ahissar & Hochstein, 2004; Ahissar, Nahum, Nelken, & Hochstein, 2009; Nahum, Nelken, & Ahissar, 2008). Whereas the reverse hierarchy theory primarily addresses top-down processing within the cortex, evidence from animal and human studies argues for an extension of the reverse hierarchy theory well beyond the cortex (Luo, Wang, Kashani, & Yan, 2008; Suga, 2008). As individuals become “expert” listeners through long-term or short-term auditory experience, it is possible that they are able to utilize the corticofugal feedback mechanism in a more

efficient manner (Chandrasekaran et al., in press; Banai et al., 2009; Chandrasekaran et al., 2009; Song et al., 2008; Wong et al., 2007). In contrast, in individuals with reading and speech-in-noise processing deficits, faulty corticofugal shaping during development may result in deficient encoding (Chandrasekaran et al., in press). Consistent with the contemporary view of cognitive influences even at the lowest levels of auditory processing (Suga, 2008), we argue that there is a critical need to understand the complex, *bidirectional* interactions between higher-level cognitive processing and lower-level sensory encoding in expert listeners as well as those with auditory processing disorders. Cognitive and sensory processes are thus inextricably linked, and scalp-recorded brainstem responses may provide a comprehensive view of the consequences of these processes as a system in humans.

### Summary

The scalp-recorded brainstem response to speech offers a unique window into understanding how the human brainstem represents key elements of the speech signal. The brainstem response to



speech has two dissociable components, the onset and a sustained frequency-following response. Together, these components faithfully represent key acoustical features, the source and filter characteristics of the speech signal. The neural sources of the FFR can be distinguished from preneural cochlear and cortical activity. Multiple lines of evidence including ablation and cooling studies, modeling and developmental data, in addition to the general phase-locking capabilities of the auditory brainstem,

strongly suggest a brainstem origin for the scalp-recorded FFR. Although the scalp-recorded onset response and the FFR probably reflect multiple sources (LL, CN, IC), they offer a noninvasive method to examine the sub-cortical encoding of speech features as well as the effect of experience on the representation of speech features. Furthermore, the dynamic nature of the brainstem response to speech offers a means to examine corticofugal modulation in the human species.

## REFERENCES

- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8, 457–464.
- Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364, 285–299.
- Akhoun, I., Gallégo, S., Moulin, A., Ménard, M., Vuillet, E., Berger-Vachon, C., et al. (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clinical Neurophysiology*, 119, 922–933.
- Banai, K., Hornickel, J. M., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Reading and subcortical auditory function. *Cerebral Cortex* (in press).
- Banai, K., Nicol, T., Zecker, S. G., & Kraus, N. (2005). Brainstem timing: Implications for cortical processing and literacy. *Journal of Neuroscience*, 25, 9850–9857.
- Bartlett, E., & Wang, X. (2007). Neural representations of temporally modulated signals in the auditory thalamus of awake primates. *Journal of Neurophysiology*, 97, 1005–1017.
- Batra, R., Kuwada, S., & Stanford, T. (1989). Temporal coding of envelopes and their interaural delays in the inferior colliculus of the unanesthetized rabbit. *Journal of Neurophysiology*, 61, 257–268.
- Belin, P., & Zatorre, R. J. (2000). 'What', 'where' and 'how' in auditory cortex. *Nature Neuroscience*, 3, 965–966.
- Bennett, M. R., & Hacker, P. M. S. (2006). Language and cortical function: Conceptual developments. *Progress in Neurobiology*, 80, 20–52.
- Burkard, R., Don, M., & Eggermont, J. (2006). *Auditory evoked potentials: Basic principles and clinical application*. Philadelphia: Lippincott Williams & Wilkins.
- Chandrasekaran, B., Hornickel, J. M., Skoe, E., Nicol, T., & Kraus, N. (in press). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: Implications for developmental dyslexia. *Neuron*.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*, 108, 1–9.
- Chowdhury, S. A., & Suga, N. (2000). Reorganization of the frequency map of the auditory cortex evoked by cortical electrical stimulation in the big brown bat. *Journal of Neurophysiology*, 83, 1856–1863.
- Cunningham, J., Nicol, T., King, C., Zecker, S. G., & Kraus, N. (2002). Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. *Hearing Research*, 169, 97–111.
- Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. (2001). Neurobiologic responses to speech in noise in children with learning problems: Deficits and strategies for improvement. *Clinical Neurophysiology*, 112, 758–767.
- Dau, T. (2003). The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. *Journal of Acoustical Society of America*, 113, 936–950.
- Eggermont, J. J. (2001). Between sound and perception: Reviewing the search for a neural code. *Hearing Research*, 157, 1–42.
- Galbraith, G., Arbagey, P., Branski, R., Comerci, N., & Rector, P. (1995). Intelligible speech encoded in the human brain stem frequency-following response. *NeuroReport*, 6, 2363–2367.
- Galbraith, G. C. (1994). Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. *Electroencephalography and Clinical Neurophysiology*, 92, 321–330.
- Galbraith, G. C., Jhaveri, S. P., & Kuo, J. (1997). Speech-evoked brainstem frequency-following responses during verbal transformations due to word repetition. *Electroencephalography and Clinical Neurophysiology*, 102, 46–53.
- Gao, E., & Suga, N. (2000). Experience-dependent plasticity in the auditory cortex and the inferior colliculus of bats: Role of the corticofugal system. *Proceedings of the National Academy of Sciences, USA*, 97, 8081–8086.
- Gardi, J., Merzenich, M., & McKean, C. (1979). Origins of the scalp recorded frequency-following response in the cat. *Audiology*, 18, 358–381.
- Gardi, J., Salamy, A., & Mendelson, T. (1979). Scalp-recorded frequency-following responses in neonates. *International Journal of Audiology*, 18, 494–506.
- Glaser, E. M., Suter, C. M., Dasheiff, R., & Goldberg, A. (1976). The human frequency-following response: Its behavior during continuous tone and tone burst stimulation. *Electroencephalography and Clinical Neurophysiology*, 40, 25–32.
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, 10, 14–23.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402.
- Hood, L. (1998). *Clinical applications of the auditory brainstem response*. San Diego: Singular.
- Hoormann, J., Falkenstein, M., Hohnsbein, J., & Blanke, L. (1992). The human frequency-following response (FFR): Normal variability and relation to the click-evoked brainstem response. *Hearing Research*, 59, 179–188.
- Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proceedings in the National Academy of Sciences USA*, 106, 13022–13027.
- Jewett, D. L. (1994). A Janus-eyed look at the history of the auditory brainstem response as I know it. *Electromyography and Clinical Neurophysiology*, 34, 41–48.
- Jewett, D. L., Romano, M. N., & Williston, J. S. (1970). Human auditory evoked potentials: Possible brain stem components detected on the scalp. *Science*, 167, 1517–1518.
- Jewett, D. L., & Williston, J. S. (1971). Auditory-evoked far fields averaged from the scalp of humans. *Brain*, 94, 681–696.
- Johnson, D. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *Journal of the Acoustical Society of America*, 68, 1115–1122.
- Johnson, K. L., Nicol, T., Zecker, S. G., Bradlow, A. R., Skoe, E., & Kraus, N. (2008a). Brainstem encoding of voiced consonant-vowel stop syllables. *Clinical Neurophysiology*, 119, 2623–2635.
- Johnson, K. L., Nicol, T., Zecker, S. G., & Kraus, N. (2008). Developmental plasticity in the human auditory brainstem. *Journal of Neuroscience*, 28, 4000–4007.
- Johnson, K. L., Nicol, T. G., Zecker, S. G., Bradlow, A., Skoe, E., & Kraus, N. (2008b). Brainstem encoding of voiced consonant-vowel stop syllables. *Clinical Neurophysiology*, 119, 2623–2635.
- Kaas, J. H., & Hackett, T. A. (1999). 'What' and 'where' processing in auditory cortex. *Nature Neuroscience*, 2, 1045–1047.
- Keuroghlian, A. S., & Knudsen, E. I. (2007). Adaptive auditory plasticity in developing and adult animals. *Progress in Neurobiology*, 82, 109–121.
- King, C., Nicol, T., McGee, T., & Kraus, N. (1999). Thalamic asymmetry is related to acoustic signal complexity. *Neuroscience Letters*, 267, 89–92.
- King, C., Warrier, C. M., Hayes, E., & Kraus, N. (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neuroscience Letters*, 319, 111–115.

- Kral, A., & Eggermont, J. (2007). 'What's to lose and what's to learn: Development under auditory deprivation, cochlear implants and limits of cortical plasticity. *Brain Research Reviews*, 56, 259–269.
- Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends in Neurosciences*, 28, 176–181.
- Krishnan, A. (1999). Human frequency-following responses to two-tone approximations of steady-state vowels. *Audiology and Neurotology*, 4, 95–103.
- Krishnan, A. (2002). Human frequency-following responses: Representation of steady-state synthetic vowels. *Hearing Research*, 166, 192–201.
- Krishnan, A., Swaminathan, J., & Gandour, J. T. (2008). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience*, 21, 1092–1105.
- Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Research. Cognitive Brain Research*, 25, 161–168.
- Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. A. (2004). Human frequency-following response: Representation of pitch contours in Chinese tones. *Hearing Research*, 189, 1–12.
- Langner, G. (1992). Periodicity coding in the auditory system. *Hearing Research*, 60, 115–142.
- Laver, J. (1994). *Principles of phonetics*. Cambridge, UK: Cambridge University Press.
- Liu, L.-F., Palmer, A. R., & Wallace, M. N. (2006). Phase-locked responses to pure tones in the inferior colliculus. *Journal of Neurophysiology*, 95, 1926–1935.
- Lomber, S. G., & Malhotra, S. (2008). Double dissociation of 'what' and 'where' processing in auditory cortex. *Nature Neuroscience*, 11, 609–616.
- Lu, T., Liang, L., & Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nature Neuroscience*, 4, 1131–1138.
- Luo, F., Wang, Q., Kashani, A., & Yan, J. (2008). Corticofugal modulation of initial sound processing in the brain. *Journal of Neuroscience*, 28, 11615–11621.
- Ma, X., & Suga, N. (2008). Corticofugal modulation of the paradoxical latency shifts of inferior collicular neurons. *Journal of Neurophysiology*, 100, 1127–1134.
- Marsh, J. T., Brown, W. S., & Smith, J. C. (1974). Differential brainstem pathways for the conduction of auditory frequency-following responses. *Electroencephalography and Clinical Neurophysiology*, 36, 415–424.
- Marsh, J. T., Worden, F. G., & Smith, J. C. (1970). Auditory frequency-following response: Neural or artifact? *Science*, 169, 1222–1223.
- Moushegian, G., Rupert, A. L., & Stillman, R. D. (1973). Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalography and Clinical Neurophysiology*, 35, 665–667.
- Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences, USA*, 104, 15894–15898.
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38, 1–21.
- Nahum, M., Nelken, I., & Ahissar, M. (2008). Low-level information and high-level perception: The case of speech in noise. *PLoS Biology*, 6, e126.
- Nakamoto, K., Jones, S., & Palmer, A. (2008). Descending projections from auditory cortex modulate sensitivity in the midbrain to cues for spatial position. *Journal of Neurophysiology*, 99, 2347–2356.
- Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Current Opinion in Neurobiology*, 14, 474–480.
- Palmer, A., & Russell, I. (1986). Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hearing Research*, 24, 1–15.
- Poepfel, D., & Hickok, G. (2004). Towards a new functional anatomy of language. *Cognition*, 92, 1–12.
- Poepfel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363, 1071–1086.
- Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*, 2, 1131–1136.
- Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology*, 115, 2021–2030.
- Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research*, 156, 95–103.
- Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., et al. (2008). Deficient brainstem encoding of pitch in children with Autism Spectrum Disorders. *Clinical Neurophysiology*, 119, 1720–1731.
- 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.
- Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26, 100–107.
- Scott, S. K., & Wise, R. J. S. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, 92, 13–45.
- Smith, J. C., Marsh, J. T., & Brown, W. S. (1975). Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources. *Electroencephalography and Clinical Neurophysiology*, 39, 465–472.
- Sohmer, H., Pratt, H., & Kinarti, R. (1977). Sources of frequency following responses (FFR) in man. *Electroencephalography and Clinical Neurophysiology*, 42, 656–664.
- Song, J. H., Skoe, E., Wong, P. C., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience*, 10, 1892–1902.
- Steinschneider, M., Arezzo, J., & Vaughan, H. G. Jr. (1980). Phase-locked cortical responses to a human speech sound and low-frequency tones in the monkey. *Brain Research*, 198, 75–84.
- Steinschneider, M., Fishman, Y. I., & Arezzo, J. C. (2008). Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. *Cerebral Cortex*, 18, 610–625.
- Stillman, R. D., Crow, G., & Moushegian, G. (1978). Components of the frequency-following potential in man. *Electroencephalography and Clinical Neurophysiology*, 44, 438–446.
- Strait, D. L., Skoe, E., Kraus, N., & Ashley, R. (2009). Musical experience and neural efficiency: Effects of training on subcortical processing of vocal expressions of emotion. *European Journal of Neuroscience*, 29, 661–668.
- Suga, N. (2008). Role of corticofugal feedback in hearing. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 194, 169–183.
- Suga, N., O'Neill, W., Kujirai, K., & Manabe, T. (1983). Specificity of combination-sensitive neurons for processing of complex biosonar signals in auditory cortex of the mustached bat. *Journal of Neurophysiology*, 49, 1573–1626.
- Suga, N., Xiao, Z., Ma, X., & Ji, W. (2002). Plasticity and corticofugal modulation for hearing in adult animals. *Neuron*, 36, 9–18.
- Sun, X., Xia, Q., Lai, C. H., Shum, D. K., Chan, Y. S., & He, J. (2007). Corticofugal modulation of acoustically induced Fos expression in the rat auditory pathway. *Journal of Comparative Neurology*, 501, 509–525.
- Sussman, E., Steinschneider, M., Gumenyuk, V., Grushko, J., & Lawson, K. (2008). The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hearing Research*, 236, 61–79.
- Suzuki, T., & Hirabayashi, M. (1987). Age-related morphological changes in auditory middle-latency response. *Audiology*, 26, 312–320.
- Swaminathan, J., Krishnan, A., & Gandour, J. T. (2008). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *NeuroReport*, 19, 1163–1167.
- Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research. Brain Research Reviews*, 43, 231–246.
- Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, 6, 391–398.
- Ulfendahl, M. (1997). Mechanical responses of the mammalian cochlea. *Progress in Neurobiology*, 53, 331–380.

- Wallace, M., Shackleton, T., Anderson, L., & Palmer, A. (2005). Representation of the purr call in the guinea pig primary auditory cortex. *Hearing Research, 204*, 115–126.
- Wallace, M. N., Shackleton, T. M., & Palmer, A. R. (2002). Phase-locked responses to pure tones in the primary auditory cortex. *Hearing Research, 172*, 160–171.
- Wang, X., Lu, T., Bendor, D., & Bartlett, E. (2008). Neural coding of temporal information in auditory thalamus and cortex. *Neuroscience, 154*, 294–303.
- Wever, E. G., & Bray, C. W. (1930). The nature of acoustic response. *Journal of Experimental Psychology, 13*, 373–387.
- 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. *Biological Psychology, 67*, 299–317.
- Winter, I., & Palmer, A. (1990). Responses of single units in the anteroventral cochlear nucleus of the guinea pig. *Hearing Research, 44*, 161–178.
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*, 420–422.
- Worden, F., & Marsh, J. (1968). Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalography and Clinical Neurophysiology, 25*, 42–52.
- Xu, Y., Krishnan, A., & Gandour, J. T. (2006). Specificity of experience-dependent pitch representation in the brainstem. *NeuroReport, 17*, 1601–1605.
- Yip, M. J. (2002). *Tone*. Cambridge, UK: Cambridge University Press.
- Zhang, Y., & Suga, N. (2005). Corticofugal feedback for collicular plasticity evoked by electric stimulation of the inferior colliculus. *Journal of Neurophysiology, 94*, 2676–2682.

(RECEIVED December 8, 2008; ACCEPTED March 25, 2009)