# Chapter 3
# Shaping Brainstem Representation of Pitch-Relevant Information by Language Experience

**Ananthanarayan Krishnan and Jackson T. Gandour**

**Abstract** Pitch is a robust perceptual attribute that plays an important role in speech, language, and music. We present compelling evidence supporting the notion that long-term language experience enhances the neural representation of behaviorally relevant attributes of pitch in the brainstem. Pitch relevant neural activity in the brainstem is crucially dependent on specific dimensions or features of pitch contours. By focusing on specific properties of the auditory signal, irrespective of a speech or nonspeech context, it is argued that the neural representation of acoustic–phonetic features relevant to speech perception is already emerging in the brainstem and, importantly, can be shaped by experience. Such effects of language experience on sensory processing are compatible with a more integrated approach to language and the brain. Long-term language experience appears to shape an adaptive, integrated, distributed pitch-processing network. A theoretical framework for a neural network is proposed involving coordination between local, feedforward and feedback components that can account for experience-induced enhancement of pitch representations at multiple locations of the distributed pitch processing network. Feedback, feedforward connections provide selective gating of inputs to both cortical and subcortical structures to enhance neural representation of behaviorally relevant attributes of the stimulus and instantiate local mechanisms that exhibit enhanced sensitivity to behaviorally relevant pitch attributes. The focus on pitch processing in tonal languages notwithstanding, the findings presented here should be contextualized within the broader framework of language experience shaping subcortical processing.

---

A. Krishnan (✉) · J.T. Gandour
Department of Speech Language Hearing Sciences, Lyles-Porter Hall,
Purdue University, 715 Clinic Drive, West Lafayette, IN 47907-2122, USA
e-mail: rkrish@purdue.edu

J.T. Gandour
e-mail: gandour@purdue.edu

© Springer International Publishing AG 2017

N. Kraus et al. (eds.), *The Frequency-Following Response*,
Springer Handbook of Auditory Research 61,
DOI 10.1007/978-3-319-47944-6_3

**Keywords** Auditory brainstem · Domain specificity · Experience-dependent plasticity · Feedback · Feedforward · Frequency-following response · Iterated rippled noise · Mandarin Chinese · Neural phase-locking · Pitch · Pitch acceleration · Pitch salience · Pitch strength

## 3.1 Introduction

Pitch is an important information-bearing perceptual component in the processing of speech, language, and music (Oxenham 2012). Functional brain imaging studies provide strong evidence for hierarchical processing of pitch (Kumar et al. 2007), starting in subcortical structures (Griffiths et al. 2001) and continuing up through Heschl's gyrus on to the planum polare and planum temporale (Gutschalk et al. 2007). Thus, pitch provides an excellent window for studying experience-dependent effects on both cortical and brainstem components of a well-coordinated, hierarchical processing network.

A complete understanding of the neural organization of language can only be achieved by viewing these processes as a set of hierarchical computations at different levels of biological abstraction—both cortical and subcortical. These computations are applied to representations that, in turn, are shaped by experience within a specific domain, such as language. Indeed, recent empirical data show that these neural representations of pitch, at both the brainstem and cortical level, are shaped by one's experience with language and music (Besson et al. 2011; Zatorre and Baum 2012).

While it is not known how language experience shapes subcortical and cortical stages of pitch processing, it is likely that the neural processes underlying such experience-dependent plasticity at each stage along the processing hierarchy are modulated by a coordinated interplay between ascending, descending, and local neural pathways that involve both sensory and cognitive components (Chandrasekaran and Kraus 2010; Kraus and White-Schwoch 2015). That is, feedback from language-dependent cortical processes shape early sensory-level processing to generate more precise outputs that facilitate stronger cortical representations of sound that, in turn, contribute to linguistic performance.

### 3.1.1 *Neural Bases of Pitch*

The neural bases of pitch perception are still a matter of debate. One view is that the auditory system extracts pitch from complex sounds by deriving a spectral profile from frequency-specific auditory input, followed by pattern-matching mechanisms (Cohen et al. 1995). A contrasting view is that the auditory system extracts pitch from the timing of auditory nerve fiber activity irrespective of frequency organization. These temporal models are based solely on the timing information available in the interspike intervals represented in simulated (Patterson et al. 1995; Meddis and O'Mard 1997) or actual (Cariani and Delgutte 1996a, b) auditory nerve activity. They derive a pitch estimate by pooling timing information across auditory nerve fibers without regard to the frequency-to-place mapping. More recently, a hybrid pitch encoding mechanism that uses both spectral and temporal information has been proposed (Cedolin and Delgutte 2005). Thus, neural phase locking related to voice fundamental frequency ($F_0$) plays a dominant role in the encoding of low pitch associated with complex sounds. Temporal encoding schemes provide a unified and parsimonious way of explaining a diverse range of pitch phenomena for at least pitch processing at or below the level of the inferior colliculus (IC) (Meddis and O'Mard 1997).

For any neural code of pitch to be useful it should be operational, at least in some form, at all levels of the processing hierarchy. There is evidence that neurons in primary auditory cortex exhibit temporal and spectral response properties that could enable these pitch-encoding schemes (Steinschneider et al. 1998; Lu et al. 2001), but whether they form a network with pitch-selective neurons to carry out this process is not known. In subcortical auditory structures, periodicity and pitch are often represented by regular temporal patterns of action potentials that are phase locked to the sound waveform. Whereas the most commonly observed code for periodicity and pitch within cortical neurons is a modulation of spike rates as a function of $F_0$, it is possible that the wider temporal integration window at the cortical level may render the auditory cortical neurons too sluggish to provide phase-locked representations of periodicity within the pitch range (Walker et al. 2011). Thus, it is not yet clear how cortical neurons transform the autocorrelation-like temporal analysis in the brainstem to a spike-rate code to extract pitch-relevant information. One possibility is that the temporal code is transformed into a response synchrony code whereby temporally coherent activity from the subcortical stages will produce greater spike rates, yielding larger response amplitudes at the cortical level. Analyses of the statistical properties of spike rates for virtual neural units have shown that frequency tuning and spike rate characteristics of their neural units are similar to those of auditory cortical neural units (Micheyl et al. 2013). Their findings indicate that sufficient statistical information is present in the population spike rate to account for small differences in frequency (pitch) and intensity (loudness).

### 3.1.2 Hierarchical Nature of Pitch Processing

Considerable progress has been made over the past decade in our understanding of the complex series of processing stages that are required to translate speech sounds into meaning at the level of the cerebral cortex. Functional imaging evidence points to multiple, parallel, hierarchically organized processing pathways that are related to speech processing in the cerebral cortex (Hickok and Poeppel 2007; Poeppel et al. 2008). Speech processing in the cortex also emerges from differential demands on distributed brain regions shared by both verbal and nonverbal auditory processing (Price et al. 2005).

In the case of pitch, functional imaging reveals hierarchical processing in subcortical regions along the auditory pathway. Encoding of temporal regularities of pitch begins as early as the cochlear nucleus but is not completed until the auditory cortex (Griffiths et al. 2001). The IC is reported to be more sensitive to changes in temporal regularity than the cochlear nucleus. Further evidence of a hierarchy of pitch processing is found in the cerebral cortex (Patterson et al. 2002). When the pitch is varied to produce a melody, activation moves beyond primary auditory cortex with relatively more activity in the right hemisphere.

Electrophysiological recordings are crucial for investigating questions about the hierarchy of pitch processing not only cortically but subcortically as well (Griffiths et al. 2004). In addition, the focus on language and pitch in the brainstem reflects the view that a complete understanding of the processing of linguistically relevant dimensions of the auditory signal can be achieved only within a framework involving a series of computations that apply to representations at different stages of processing. Hickok and Poeppel (2004) argue that early processing stages (e.g., within the brainstem) may perform transformations on the acoustic data that are relevant to linguistic as well as nonlinguistic auditory perception. Scott (2003) similarly argues for hierarchical processing at the cortical level, allowing for the possibility of differences in the degree of processing of speech and nonspeech stimuli.

### *3.1.3 Linguistic Functions of Pitch*

Tone languages exploit phonologically contrastive pitch at the word or syllable level (Gandour 1994; Yip 2002). Such languages are common in the Far East and Southeast Asia. In Mandarin, for example, four words may comprise a minimal quadruplet, minimally distinguished by variations in pitch, but otherwise identical in terms of consonant and vowel segments. Mandarin has four lexical tones (Howie 1976): *yi*$^{1}$ “clothing” high level [T1]; *yi*$^{2}$ “aunt” high rising [T2]; *yi*$^{3}$ “chair” low falling-rising [T3]; *yi*$^{4}$ “easy” high falling [T4] (see Fig. 3.1, panel B). Such languages are to be distinguished from those in which pitch variations are usually not contrastive at the syllable or word level (e.g., English). In languages that are not tonal, however, variations in pitch may be used to signal stress and intonation patterns at post-lexical levels of representation. The crucial feature that differentiates between these two types of languages is whether or not pitch variations are contrastive in the lexicon. All languages use pitch variations for intonation, but fewer possibilities are available in tone languages because of co-occurring demands for pitch variation at the lexical level. Thus, tone languages not only provide a physiologic window to evaluate how neural representations of linguistically relevant pitch attributes emerge along the early stages of sensory processing in the hierarchy, but they may also shed light on the nature of interactions between early sensory levels and later, higher levels of cognitive processing in the human brain.

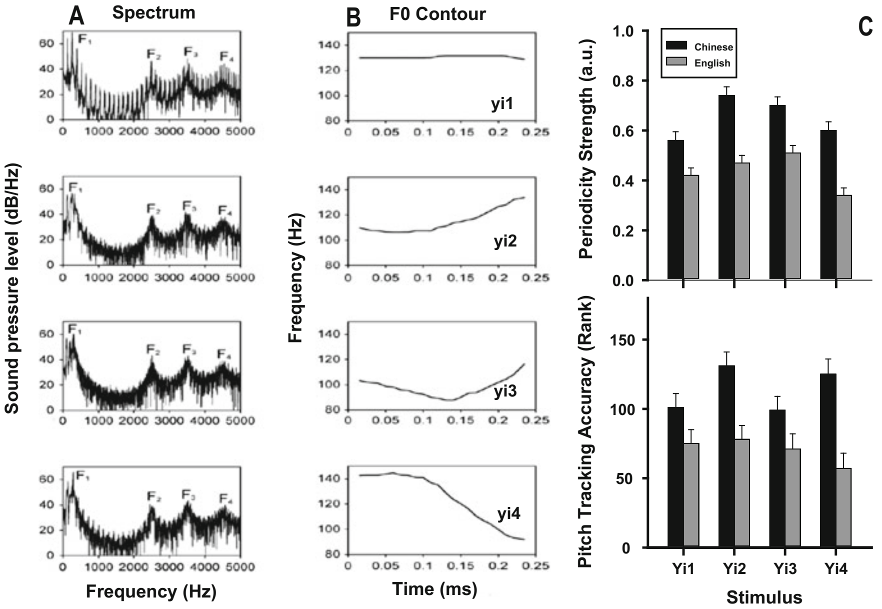


**Fig. 3.1** (**A**) Stimulus spectrum for corresponding $F_0$ contours of the four Mandarin Chinese synthetic speech stimuli shown in **B** with invariant spectra across the four tones. (**B**) $F_0$ contours for Mandarin Chinese syllables identified in bottom right of each panel. (**C**) FFR periodicity strength (*top*) and FFR pitch tracking accuracy (*bottom*) of the four Mandarin lexical tones compared for Chinese and English listeners per lexical tone. Chinese listeners show greater periodicity strength and tracking accuracy across the four stimuli compared to English listeners. (Adapted from Krishnan et al. 2005, with permission of Elsevier B.V.)

## 3.1.4 *Perceptual Dimensions of Pitch in Tone Languages*

Voice $F_0$ is the most important acoustic correlate of tones, although there may be concomitant changes in duration, intensity, and phonation of tones as well. The primacy of $F_0$ as a cue to tonal identification of citation forms has been confirmed in perception tests using both natural and synthetic speech stimuli. In Mandarin, $F_0$ contours (shape of the temporal change in pitch over the duration of the stimulus) provide the dominant cue for tone recognition (Xu 1997). Both $F_0$ height and movement provide sufficient information for high intelligibility of tones in Thai (Abramson 1962), Mandarin (Howie 1976), and Cantonese (Khouw and Ciocca 2007). Rapid $F_0$ movements are required for high intelligibility of contour tones (Abramson 1978). Identification and discrimination tasks reveal classical patterns of categorical perception (CP) in Mandarin Chinese listeners for both speech and nonspeech stimuli varying along a linear rising or falling $F_0$ continuum (Xu et al. 2006a; Peng et al. 2010). CP for pitch direction depends on a listener's experience with a tone language, as shown by the lack of similar CP effects in nonnative listeners (Peng et al. 2010).

As a multidimensional perceptual attribute, pitch relies on several acoustic dimensions (e.g., height, direction). Psychophysical evidence comes primarily from cross-language multidimensional scaling (MDS) studies of dissimilarity ratings. Based on tone languages from the Far East and West Africa, including English as a control, three dimensions are reported to underlie a common perceptual space: average $F_0$ height, direction of $F_0$ movement, and magnitude of $F_0$ slope (Gandour and Harshman 1978; Gandour 1983). Their relative importance varies depending on a listeners' familiarity with specific types of pitch patterns that occur in their native language. For example, the perceptual saliency of the contour dimension is greater for native speakers of tone languages than for speakers of English, while English listeners give greater weight to the height dimension than do tone language speakers. Such differences in perceptual saliency suggest that long-term experience enhances listeners' attention to pitch dimensions that are phonetically relevant in a particular language.

### *3.1.5 A Physiologic Window into Speech and Language*

The scalp-recorded human frequency-following response (FFR) provides a physiologic window into the early stages of subcortical processing of complex sounds. The FFR reflects sustained, phase-locked activity in a population of neural elements within the rostral brainstem, presumably the IC (Krishnan 2007). These responses can be recorded easily between scalp electrodes placed at high forehead and the seventh cervical vertebra (C7). The shorter latency of the FFR (around 6–9 ms) correlates well with activity from the IC region and is too early to reflect activity from cortical generators (Galbraith 2008). Furthermore, the nature of the auditory system makes it unlikely that the low-pass filtered phase-locked activity reflected in the FFR is of cortical origin (Akhoun et al. 2008). However, there is compelling evidence to suggest that this brainstem component is indeed subject to corticofugal modulation (Suga and Ma 2003; Banai et al. 2007).

The FFR is characterized by a waveform that follows the periodicities contained in both the envelope and the temporal fine structure of complex sounds (Fig. 3.2). The temporal and spectral characteristics of complex sounds that are preserved in the FFR can be extracted by frequency domain (spectral) and time domain autocorrelation analysis (Fig. 3.3), a measure of correlation between the original signal and temporally delayed versions of the response, yielding high correlations for periodicities harmonically related to the fundamental frequency (Krishnan et al. 2004, 2005). FFRs preserve spectrotemporal information relevant to the spectrum and pitch of steady-state (Krishnan and Plack 2011), dynamic speech (Krishnan et al. 2004, 2005), and nonspeech stimuli (Krishnan et al. 2009a, b). Importantly, the pitch-relevant information preserved in the FFR is strongly correlated with perceptual measures of pitch salience—a measure of the strength of the perceived pitch (Krishnan et al. 2010a; Bidelman and Krishnan 2011). These findings suggest

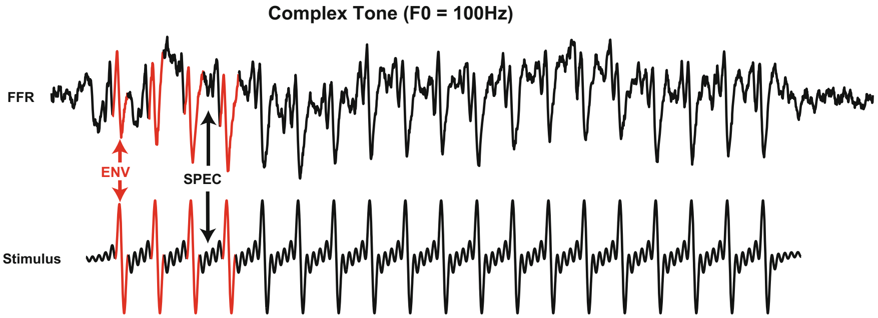


**Fig. 3.2** Frequency-following responses (FFR, *top trace*) elicited by a complex sound (*bottom trace*). The response is characterized by neural phase-locking to the envelope (*ENV*, *red* for four cycles) periodicity and the temporal fine structure (*SPEC*, *double black arrow*). Stimulus onset is shifted to the right to achieve temporal match between the stimulus and the response. *ENV*, envelope; *SPEC*, spectral. (Reprinted from Krishnan and Gandour 2014, with permission from the Acoustical Society of Australia)

acoustic features relevant to pitch are preserved in the temporal pattern of phase-locked neural activity in the brainstem.

Gockel et al. (2011) observed that FFRs recorded using frequency-shifted complex tones presented monaurally did preserve pitch-relevant information, but that this information was similar to that measured in an auditory nerve model. Also, they failed to observe any pitch-relevant information in the FFRs to three-tone harmonic stimuli presented dichotically. They concluded that there was no additional pitch-relevant processing at the level of the brainstem. Several arguments may be presented to counter this inference. First, if the temporal code for pitch available at the brainstem level also utilizes autocorrelation-like analysis to determine the global distribution of interspike intervals from the temporal pattern of neural activity across a population of neurons, it would necessarily share certain fundamental attributes of the same temporal code operating at the level of the auditory nerve. Second, it is not clear that their dichotic stimuli produced the same pitch as when all harmonics are presented to the same ear. Notwithstanding, the salience of their stimuli would be quite weak. It is possible that FFR-related neural activity is not sufficiently robust to preserve the less salient pitch for their stimuli. Finally, the inferences of Gockel et al. (2011) cannot account adequately for the experience-dependent effects reflected in the FFR that are sensitive to specific attributes of dynamic pitch contours.

Within the past decade there has been increasing interest in the role of the auditory brainstem in speech processing. In terms of speech intelligibility, FFRs show increased amplitude in response to forward speech as compared to reversed speech, indicating that familiar phonetic and prosodic properties of forward speech selectively activate brainstem neurons (Galbraith et al. 2004). Using the /da/ syllable to elicit the brainstem response, Kraus and colleagues demonstrated how FFRs separately encode source and filter characteristics of the speech signal (Kraus and Nicol 2005;

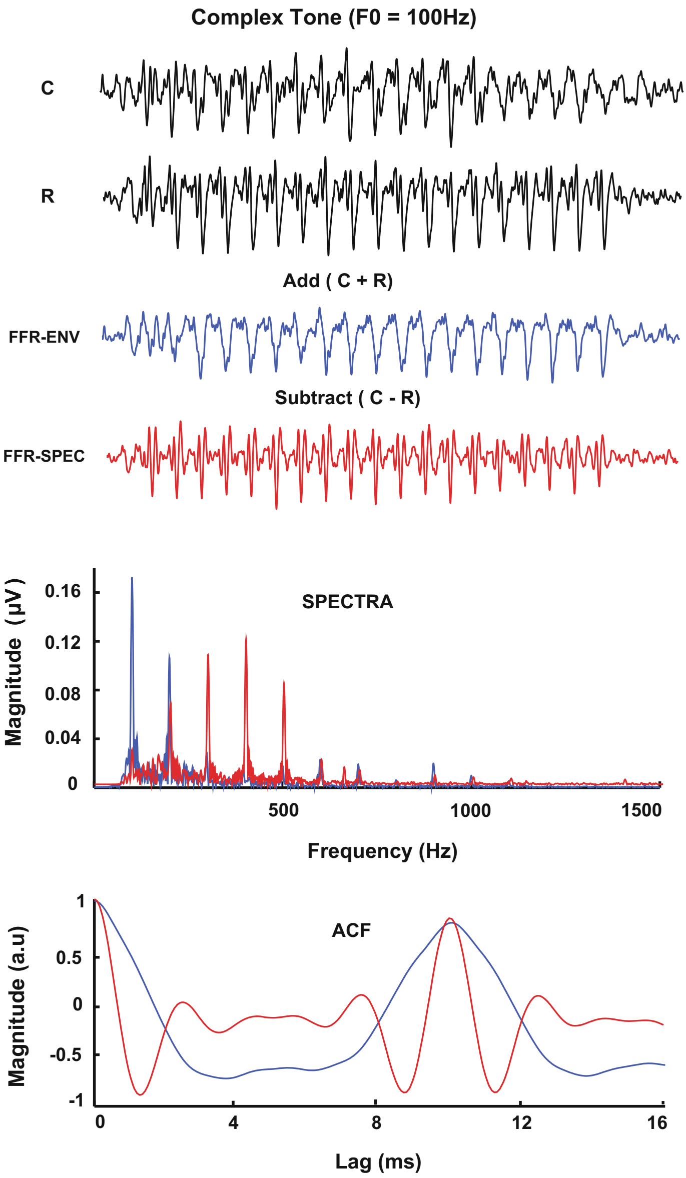


**Fig. 3.3** FFR analyses: *Top two* FFR waveforms (*black*) are responses to condensation (***C***) and rarefaction (***R***) onset stimulus polarity. Addition of these waveforms results in the FFR-ENV waveform (*blue*). It is characterized by a prominent phase-locking to the envelope periodicity of the complex tone. Subtraction of the C and R responses (***C***−***R***) results in a FFR phase locked to the temporal fine structure (spectrum) of the complex tone (FFR-SPEC, *red*). Frequency domain analysis (SPECTRA, *middle panel*) shows that the envelope phase locking has a larger peak at $F_0$ (*blue*) as expected; whereas the temporal fine-structure phase locking shows multiple robust peaks at the harmonics of the complex tone. Using autocorrelation, the temporal analyses (ACF, *bottom panel*) show that there is a major peak at the fundamental periodicity of the complex tone for both responses. (Reprinted from Krishnan and Gandour 2014, with permission from the Acoustical Society of Australia)

Chandrasekaran and Kraus 2010) and how brainstem timing predicts cerebral asymmetry for speech (Abrams et al. 2006).

In terms of segmental features of speech, FFRs preserve spectral peaks corresponding to the first two formants (Fig. 3.4) of both steady-state vowels (Krishnan 1999, 2002) and time-variant consonants (Krishnan and Parkinson 2000; Plyler and Ananthanarayan 2001). Though FFRs are known to preserve pitch-relevant information about complex sounds that produce time-invariant pitch (Greenberg et al. 1987), the question arises as to how the brainstem handles suprasegmental features of speech that are characterized by time-variant pitch.

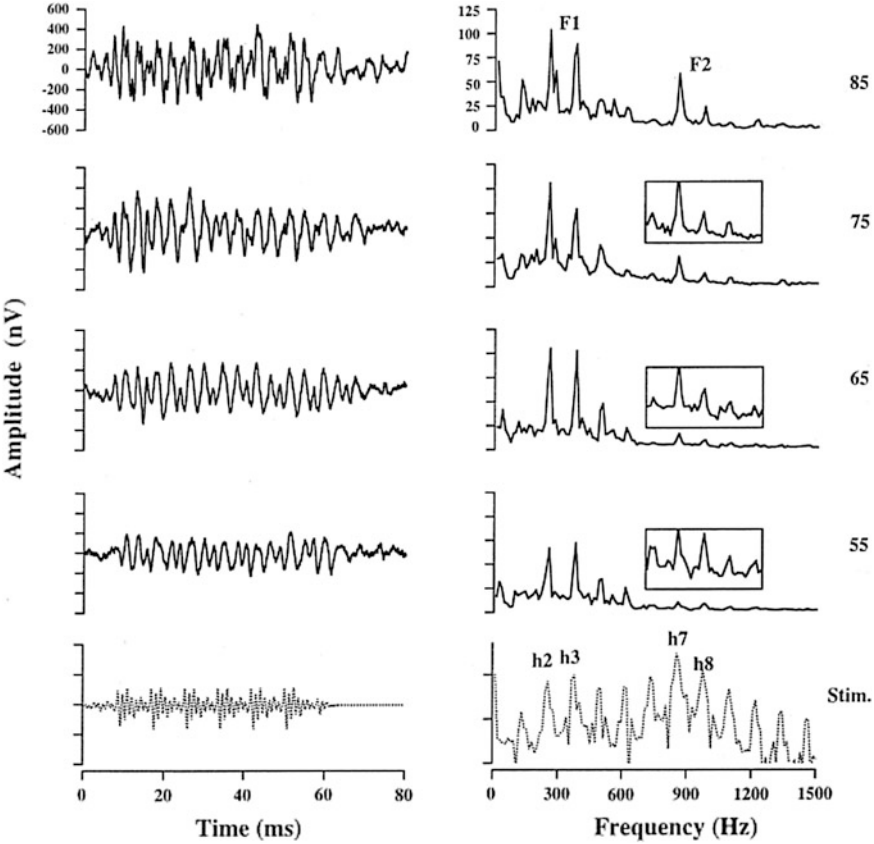


**Fig. 3.4** Grand averaged FFR waveforms (*left*) and spectra (*right*) are plotted as a function of stimulus level (55–85 dB normal hearing level) for the English vowel /u/. The stimulus waveform (*left bottom*) and its spectrum, with $F_1$ (h2, h3) and $F_2$ (h7, h8) harmonics identified on the figure (*right bottom*). The amplified inset in the FFR spectral data clearly shows the $F_2$ harmonic peaks h7 and h8. Note the different amplitude scale for the stimulus spectrum. (Reprinted from Krishnan 2002, with permission of Elsevier B.V.)

Tone languages are especially advantageous for isolating the effects of encoding voice pitch at the level of the auditory brainstem (see Sect. 3.1.3). In Mandarin, for example, all four tones exhibit voice $F_0$ trajectories and harmonics that lie within the range of easily recordable FFRs (below 2 kHz). The relatively long duration of citation forms of lexical tones (200–350 ms) necessitates use of slower stimulus repetition rates. This, in turn, enables recording of robust FFRs with little or no neural adaptation.

### *3.1.6 Language-Dependent Processing in the Brainstem*

Historically, the brainstem has not been considered to be a part of the brain worthy of interest when it comes to its contribution to speech/language processing. The conventional wisdom is that “processing operations conducted in the relay nuclei of the brainstem and thalamus are general to all sounds, and speech-specific operations probably do not begin until the signal reaches the cerebral cortex” (p. 100 in Scott and Johnsrude 2003). Though it is agreed that operations specific to speech perception are likely circumscribed to the cortex, experience-dependent modulation of pitch-relevant neural activity in the brainstem suggests that early sensory processing involves more than a simple transmission of pitch information from the ear to the cerebral cortex.

## 3.2 Language Experience Shapes Pitch-Relevant Information in the Brainstem

Long-term language experience enhances the neural representation of linguistically relevant pitch in the human brainstem, which is well before evoked neural activity relevant to pitch is detected in the auditory cortex. Indeed, neural representation of pitch-relevant attributes, as reflected in the FFR, may emerge as early as 6–10 ms after stimulus onset (Krishnan and Gandour 2009). In contrast, the pitch-related neural activity in the auditory cortex emerges at about 140–170 ms after stimulus onset (Krumbholz et al. 2003; Griffiths and Hall 2012).

### *3.2.1 Experience-Dependent Enhancement of Neural Representation of Pitch-Relevant Information*

In the cerebral cortex, the neural substrates of pitch perception in the processing of lexical tones are shaped by language experience (Zatorre and Gandour 2008; Gandour and Krishnan 2014). Based on evidence from positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies of pitch processing in Mandarin and Thai, it appears that pitch processing engages the left hemisphere only when the pitch patterns are of linguistic relevance (Wong 2002; Wong et al. 2004). These experiments all employed discrimination tasks, and thus likely reflect temporally aggregated neural events at relatively late attention-modulated stages of auditory processing. As reflected by the mismatch negativity (MMN: an early cortical event-related response associated with auditory discrimination), language experience similarly influences the early cortical automatic processing of linguistically relevant pitch contours (Chandrasekaran et al. 2007a, b); moreover, lexical tones are lateralized to the right hemisphere in contrast to the left lateralized consonants (Luo et al. 2008). The Luo et al. data suggest that hemispheric laterality effects are a result of specialized neural computations that apply to representations at different stages of auditory processing. The leftward asymmetry observed in a discrimination task likely reflects neural computations that occur downstream from preattentive auditory processing.

In animals, it is already well-established that experience-dependent neural plasticity is not limited to the cerebral cortex. Response properties and frequency maps in the IC of bats undergo change after auditory conditioning or focal electrical stimulation of the auditory cortex (Suga 1990, 1994). Auditory experience of altered interaural cues for localization in young owls leads to frequency-dependent changes in interaural time difference (ITD) tuning and frequency tuning of IC neurons (Gold and Knudsen 2000).

Other recent data also support experience-dependent neural plasticity at the level of the IC in humans. The latency of wave V in hearing-impaired listeners who use amplification is shorter than in those who do not (Philibert et al. 2005). Using the FFR, neural phase-locked activity improves after auditory training in children with learning impairments (Russo et al. 2005); pitch tracking accuracy of Mandarin tones is more accurate in nonnative musicians than nonmusicians (Wong et al. 2007); experience with sounds composed of acoustic elements relevant to speech leads to developmental changes in brainstem responses (Johnson et al. 2008); and pitch tracking accuracy improves in native English-speaking adults after undergoing short-term training on using Mandarin tones in word identification (Song et al. 2008b). Also relevant is the consequence of a disruption in the normal interaction between local processes and the corticofugal modulation of subcortical function, which contributes to plasticity. The deficits in brainstem encoding in children with a variety of language-based learning problems could very well reflect such a disruption in the ability of the corticofugal system to fine tune subcortical processes (Russo et al. 2008; Song et al. 2008a).

Preattentive stages of pitch processing in the brainstem can be influenced by language experience. A cross-language study was conducted to determine whether native speakers' long-term exposure and experience using pitch patterns in a tonal language has an influence on FFR response properties (Krishnan et al. 2005). FFRs were elicited by prototypical, curvilinear $F_0$ contours modeled after the four Mandarin lexical tones in a speech context (Fig. 3.1, panels A, B). They were presented to native speakers of Mandarin and nontone language speakers of

English. If driven by acoustic properties regardless of language experience, FFRs would be expected to be homogeneous across listeners. Results showed that both pitch strength (Fig. 3.1, panel C *top*) and pitch tracking accuracy (Fig. 3.1, panel C *bottom*), as measured by rank-transformed cross-correlation between stimulus and response pitch tracks, were greater for the Chinese group than for the English across all four Mandarin tones (Fig. 3.1, panel C *bottom*). Based on these findings, it appears that experience-driven adaptive neural mechanisms are involved subcortically that sharpen response properties of neurons tuned for processing linguistic pitch contours of a particular language. That is, language-dependent plasticity enhances or primes temporal intervals that carry linguistically relevant features of pitch contours.

From the perspective of auditory neuroethology, this adjustment in processing pitch contours of Mandarin tones is comparable to neural mechanisms that are developed for processing behaviorally relevant sounds in other nonprimate and nonhuman primate animals (Suga et al. 2003). Auditory processing is not limited to a simple representation of acoustic features of speech stimuli. Indeed, language-dependent operations may begin before the signal reaches the cerebral cortex.

While this chapter focuses primarily on the experience-dependent shaping of pitch-relevant information as it relates to tonal languages, it is important to view the information presented here in the broader context of how language experience shapes subcortical processing in general. For example, compelling evidence suggests that bilingual experience enhances subcortical representation of pitch-relevant information presented in speech sounds (Krizman et al. 2012) and/or neural response consistency (Krizman et al. 2014). Furthermore, these authors show a strong correlation between these improvements in subcortical representations and both attentional control and language proficiency, suggesting that these outcomes are related to experience-dependent strengthening of attentional control.

### 3.2.2 Feature Specificity of Experience-Dependent Effects in the Brainstem

If brainstem pitch processing is shaped by long-term language experience, what specific $F_0$ properties or features of the pitch stimuli, static or dynamic, are relevant? To what extent can a stimulus deviate from natural speech exemplars before exceeding the upper or the lower limit of linguistic sensitivity of brainstem neurons? For instance, linear ramps do not occur in natural speech because of physiological constraints of the speech production apparatus. Linear $F_0$ ramps (90–140 Hz, rising; 140–90 Hz, falling) were generated, similar to Mandarin T2 (rising) and T4 (falling) in direction but dissimilar in trajectories observed in natural speech (Fig. 3.5, panels A and B). By examining FFRs elicited by linear approximations of Mandarin T2 and T4 (Xu et al. 2006b), it was possible to assess the tolerance limits for priming

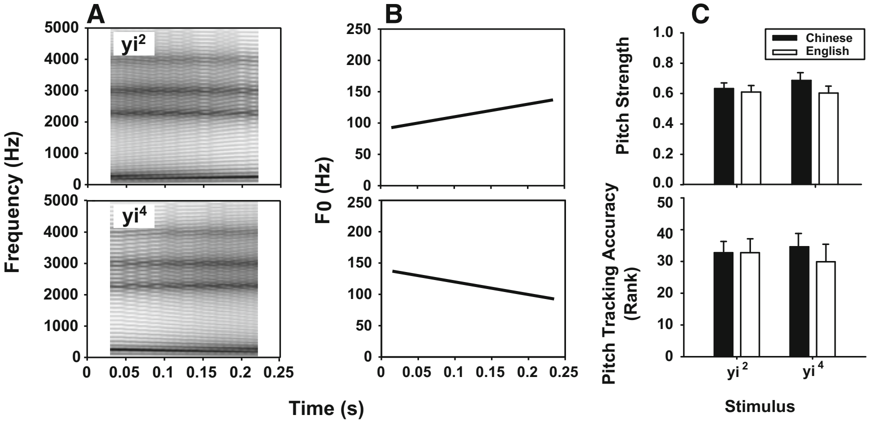


**Fig. 3.5** (**A**) Stimulus spectrograms, (**B**) $F_0$ contours of Mandarin Chinese synthetic speech stimuli (*yi*$^2$ "aunt", rising linear ramp; *yi*$^4$ "easy", falling linear ramp), FFR pitch strength (**C** *top*), and FFR pitch tracking accuracy (**C** *bottom*). FFR pitch tracking accuracy revealed no significant main effects for either language group (*Chinese, English*) or pitch direction (*rising, falling*). (Adapted from Xu et al. 2006a, b, with permission of Lippincott Williams & Wilkins)

linguistically relevant features of the auditory signal involved in pitch extraction at the level of the brainstem. Results show no differences in FFR pitch strength or pitch tracking accuracy (Fig. 3.5, panel C top and bottom, respectively) between language groups (Chinese, English) or pitch direction (rising, falling).

It appears that no language-dependent effects are observed in response to linear rising or falling $F_0$ ramps because they are not part of native Chinese listeners' experience. Even though the $F_0$ ramps are dynamic, linear approximations of T2 and T4, they are constant in acceleration and deceleration, respectively. The fact that Mandarin and English FFRs are homogeneous in response to linear trajectories suggests that representations of pitch-relevant information in the brainstem are acutely sensitive to dynamic, curvilinear changes in trajectory throughout the duration of a pitch contour. In the auditory brainstem, neural mechanisms respond to specific dimensions of pitch contours to which native speakers have been exposed. Language dependent neuroplasticity occurs only when salient dimensions of pitch relevant to speech perception are present in the auditory signal.

Further support for feature specificity comes from FFRs recorded from Chinese and English participants in response to iterated rippled noise (IRN) homologs of pitch contours. The IRN stimuli preserve the perception of pitch, but do not have waveform periodicity or highly modulated stimulus envelopes that are characteristic of speech stimuli. An IRN stimulus is generated using a broadband noise, which is delayed and added to itself repeatedly and, therefore, does not have a prominent modulated envelope (Patterson et al. 1996; Yost 1996). The perceived pitch corresponds to the reciprocal of the delay, and the pitch salience increases with the number of iterations of the delay-and-add process. Increases in temporal regularity of steady-state IRN stimuli lead to better temporally locked neural activity in auditory structures from the cochlear nucleus to cortex (Griffiths et al. 1998; Shofner 1999). Importantly, a novel generalization of the IRN algorithm makes it possible to generate time-variant, dynamic curvilinear pitch contours representative of those that occur in natural speech (Swaminathan et al. 2008). The stimulus waveform (Fig. 3.6A), spectrogram (Fig. 3.6B), pitch contour (Fig. 3.6C), and the autocorrelation function (Fig. 3.6D) are illustrated for an IRN stimulus with a T2 pitch contour. Note the increase in temporal regularity with increase in the number of iteration steps.

The IRN homologs of a prototypical T2 were presented in contrast to three $F_0$ variants (two linear, one curvilinear) that do not occur in the Mandarin tonal space (Krishnan et al. 2009a). Of the two linear variants, one represented a linear ascending ramp and the other represented a tri-linear approximation of T2, preserving the major points of inflection in addition to onset and offset. The curvilinear variant was an inverted version of T2. No group differences in pitch strength were observed for any of these variants. The absence of language group effects in response to curvilinear and linear variants of T2 emphasizes that language-dependent neuroplasticity at the level of the brainstem extends only to those pitch patterns that actually occur in the Mandarin tonal space.

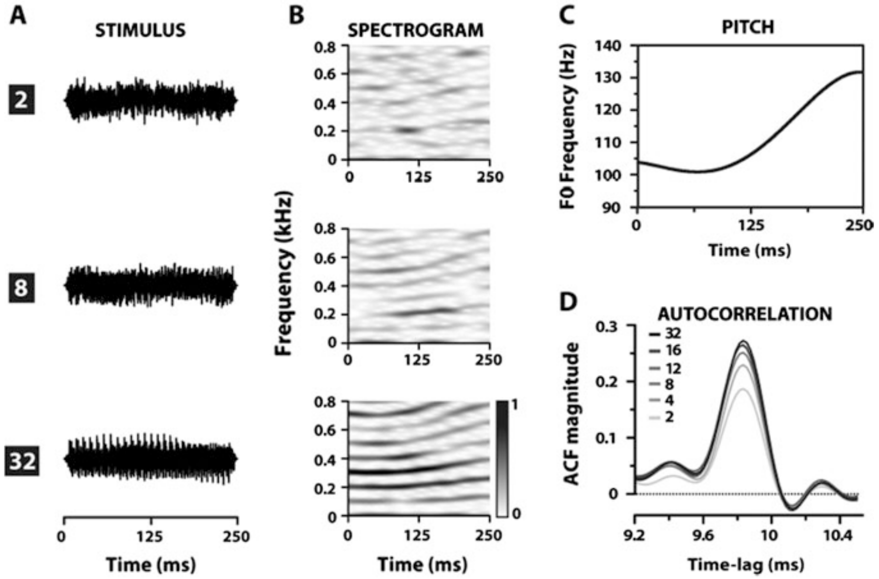


**Fig. 3.6** (**A**) Iterated ripple noise (IRN) stimulus waveforms, (**B**) spectrograms, (**C**) $F_0$ contour and autocorrelation functions (**D**) plotted as a function of iteration steps (*2, 8, 32*). Note the increase in temporal regularity, clearer spectral bands, and increase in the peak magnitude of the autocorrelation function with increase in iteration steps. (Adapted from Krishnan et al. 2010a, with permission from Elsevier, B.V.)

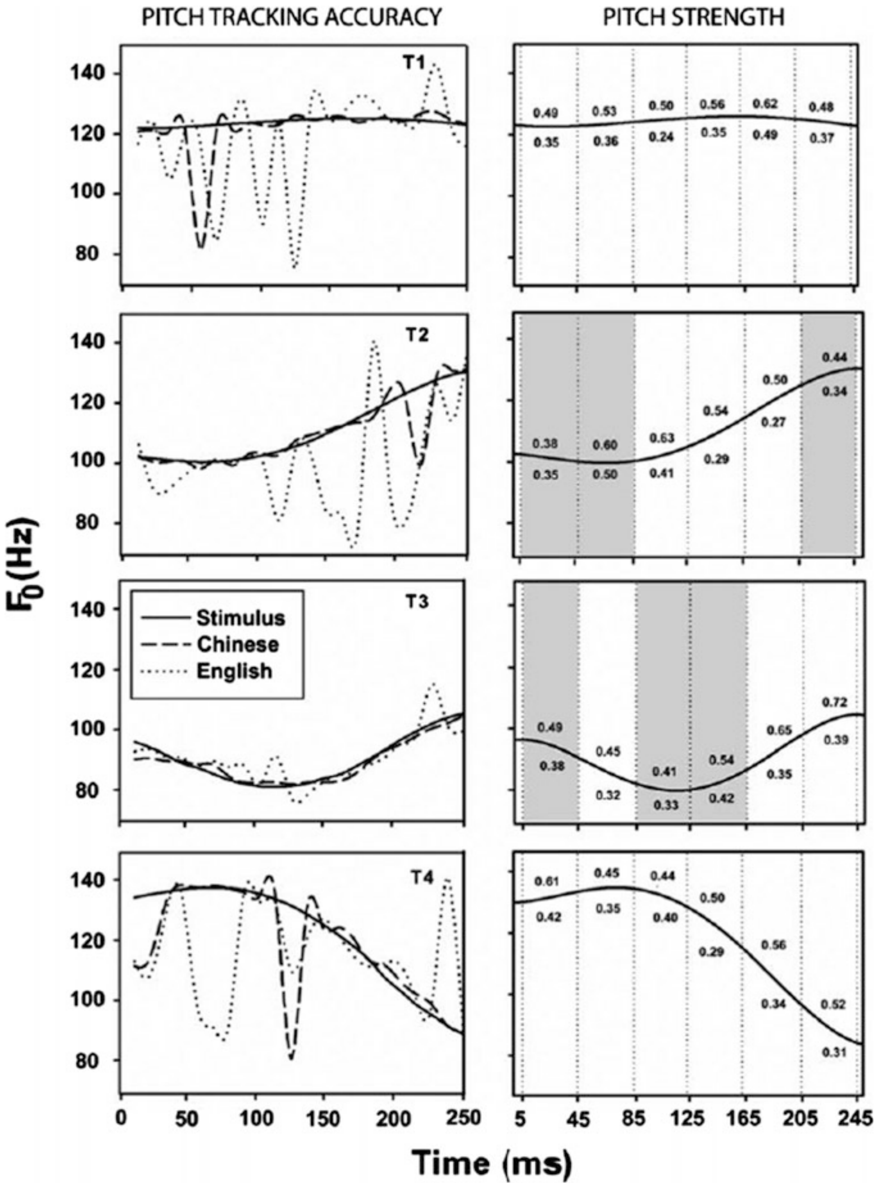
### *3.2.3 Domain Specificity of the Experience-Dependent Effects in the Brainstem*

#### 3.2.3.1 Speech Versus Nonspeech

To address the question of domain specificity of experience-dependent effects on pitch processing in the brainstem, FFRs were recorded from native speakers of Mandarin and English speakers using IRN (nonspeech) homologs of the four Mandarin tones (Krishnan et al. 2009b). The Chinese group exhibited smoother pitch tracking than the English group in three (T2–4) out of the four tones (Fig. 3.7, left column). FFR pitch strength of 40-ms segments revealed that the Chinese group exhibited more robust pitch representation of those segments containing rapidly changing pitch movements across all four tones (Fig. 3.7, right column). These findings suggest that neural mechanisms underlying pitch representation are shaped by particular dimensions of the auditory stream rather than speech per se.

A discriminant analysis was used to determine the extent to which individual subjects can be classified into their respective language groups based on a weighted linear combination of their pitch strength of three 40-ms temporal intervals that were maximally differentiated in terms of slope (flat, rising, falling). About 83% of the subjects were correctly classified into their respective language groups. The average discriminant $z$ scores of the Chinese group were larger than those of the English. Univariate tests of pitch strength confirmed that more dynamic changes in pitch (rising, falling) had a greater influence on the FFR responses of the Chinese group compared to the English, whereas less dynamic changes in pitch (flat) did not yield a language group effect. Pitch strength of the rising $F_0$ trajectory was the most important variable in discriminating listeners by language affiliation. Both psychoacoustic (Collins and Cullen 1978; Schouten 1985) and physiologic studies (Shore et al. 1987; Krishnan and Parkinson 2000) indicate better sensitivity for rising versus falling tones. Multidimensional scaling analyses showed that the perceptual dimension related to direction of pitch change is spatially distributed primarily in terms of rising versus non-rising $F_0$ movements (Gandour and Harshman 1978; Gandour 1983). This response asymmetry in FFRs presumably reflects greater neural synchrony (Shore and Nuttall 1985) and more coherent temporal response patterns to rising than to falling tones (Shore et al. 1987).

Such experimental findings support the view that at early stages of brain processing, particular features or dimensions of pitch patterns—regardless of the stimulus context in which they are embedded—shape neural mechanisms underlying speech perception. The role of the brainstem may be to facilitate cortical level processing of pitch-relevant information by optimally capturing those dimensions of the auditory signal that are of linguistic relevance. By focusing on tonal sections instead of the whole tone, it is possible to assess whether language-dependent effects are better conceptualized as applying to sections that exhibit certain acoustic features irrespective of tonal category.



The question arises whether neural representation of pitch-relevant information in the brainstem is equally robust for speech and nonspeech stimuli. FFRs were recorded from Chinese and English listeners in response to four Mandarin tones presented in speech and nonspeech contexts (Swaminathan et al. 2008). Dynamic IRN stimuli preserve fine-grained measures of pitch representation at the level of the brainstem. However, given the relatively less robust temporal

◀ **Fig. 3.7** Pitch tracking accuracy of Iterated ripple noise (IRN) homologs of Mandarin tones (*left*) and pitch strength of tonal sections (*right*) derived from the grand averaged FFR waveforms of Chinese and English subjects. The four Mandarin tonal categories are represented as *T1*, *T2*, *T3*, and *T4*. Left panels show that the FFR-derived $F_0$ contours of the Chinese group (*dashed line*) more closely approximate those of the original IRN stimuli (*solid line*) when compared to the English group (*dotted line*). Right panels show that the pitch strength of the Chinese group (*value above the solid line*) is greater than that of the English group (*value below the solid line*). *Vertical dotted lines* demarcate six 40-ms sections within each $F_0$ contour: 5–45, 45–85, 85–125, 125–165, 165–205, and 205–245. Sections that yielded significantly larger pitch strength for the Chinese group relative to English are *unshaded*; those that did not are shaded in *gray*. (Reprinted from Krishnan et al. 2009b, with permission of MIT Press)

periodicity in the IRN waveform, pitch strength was observed to be greater for speech than nonspeech stimuli across language groups. Regardless of context, pitch strength of the Chinese group was greater than that of the English. Interestingly, group differences in pitch strength were not uniform throughout the duration of FFR responses. The Chinese group exhibited relatively more robust pitch representation of rapidly changing pitch segments. These findings support the view that at the level of the brainstem, neural mechanisms underlying pitch representation are shaped by particular features of the auditory stream rather than speech per se.

#### 3.2.3.2 Speech Versus Music

Neural encoding of pitch in the auditory brainstem is shaped by long-term experience with language or music, implying that early sensory processing is subject to experience-dependent neural plasticity. The comparisons between the language and music domains reveal overall enhancement in brainstem FFRs elicited by either musical or linguistic pitch patterns in musicians and tone language speakers alike (Bidelman et al. 2011b, c). Thus, long-term pitch experience seems to improve the brain’s ability to represent pitch-relevant information regardless of the domain of expertise. However, subtle differences in these sensory representations suggest a domain-specific sensitivity to acoustic features that are part of the experience in each domain. Musicians, for example, show enhanced responses when pitch patterns intersect discrete notes along the musical scale; tone language speakers, on the other hand, during rapidly changing portions of tonal contours (Bidelman et al. 2011a, c). Such cue weighting is consistent with the relative importance of these perceptual dimensions in their respective domains. These findings collectively suggest that both language and musical experience provide some mutual benefit to the neural representation of pitch-relevant information, but also they suggest that specific features of the acoustic signal are highlighted in subcortical responses depending on their perceptual salience and function within a listener’s domain of expertise.

### *3.2.4 Experience-Dependent Effects are More Resilient to Signal Degradation*

Most human communication occurs against a background of noise. It is important that the auditory system has a mechanism(s) in place to encode behaviorally relevant acoustic features of pitch that may be degraded in the presence of noise. By using IRN homologs of pitch contours associated with lexical tones, it is possible to systematically vary their degree of temporal regularity and, as a consequence, their pitch salience. The question then arises whether pitch representation in the brainstem is less vulnerable to systematic degradation in the temporal regularity of an IRN stimulus that represents a native pitch contour (Krishnan et al. 2010b). In response to IRN homologs of Mandarin T2 varying in pitch salience along a six-step continuum ranging from low to high, FFR pitch tracking accuracy is higher in the Chinese group relative to the English except for the three lowest steps along the continuum (Fig. 3.8, left panel). Also, FFR pitch strength is greater in the Chinese group, even in severely degraded stimuli, for sections of the response that exhibit rapid changes in pitch. Exponential time constants reveal that pitch strength growth emerges 2–3 times faster in Chinese than in English listeners as a function

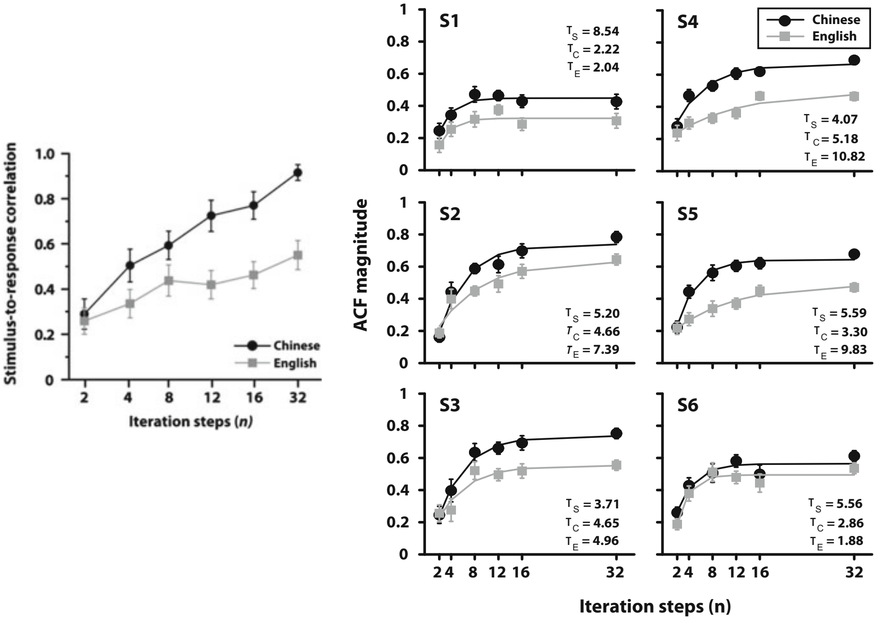


**Fig. 3.8** Comparisons of FFR pitch tracking accuracy (*left panel*), and segmental periodicity strength (*right panel*) in response to iterated ripple noise (IRN) homologues of Mandarin Tone 2 (*T*2) as a function of iteration steps. Both pitch-tracking accuracy and periodicity strength (especially the rapidly changing segments 4 and 5 as reflected by the time constants) increase more rapidly with increasing pitch salience for the Chinese listeners. (Adapted from Krishnan et al. 2010b, with permission from Elsevier, B.V.)

of increasing temporal regularity of the stimulus (Fig. 3.8, right panels). These findings altogether suggest that experience-dependent brainstem mechanisms for pitch are especially sensitive to those dimensions of tonal contours that provide cues of high perceptual saliency in degraded as well as normal listening conditions.

Another way to degrade pitch-relevant information in the stimulus is to systematically increase the rate of pitch change in the dynamic portions of the stimulus. The question then arises whether language-related expertise in pitch encoding of linguistically relevant stimuli can transfer to pitch encoding of stimuli that are characterized by acceleration rates that do not occur in natural speech. Four click-train homologs of Mandarin T2 (Fig. 3.9, left panels) with maximum rates of

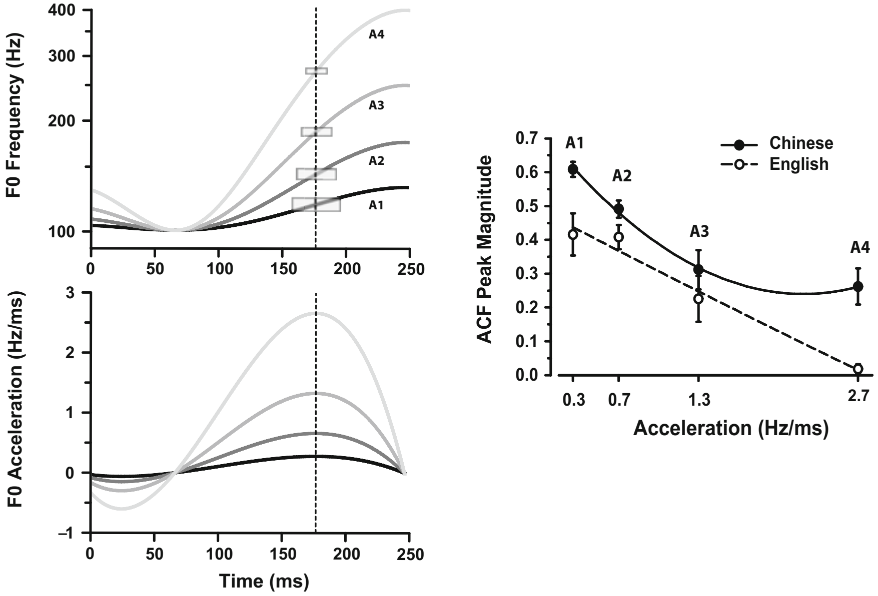


**Fig. 3.9** Click train stimuli that are differentiated by varying degrees of rising acceleration. The $F_0$ contours of all four stimuli are displayed on a logarithmic scale spanning two octaves: 100.8 Hz, the minimum stimulus frequency, to 400 Hz (*left panel, top*). These stimuli represent a continuum of rates of acceleration from Mandarin Tone 2 (*A1*) rate in natural speech to an $F_0$ rate that falls well beyond the normal voice range (*A4*). The *vertical dotted line* at 177 ms defines the center of the analysis window for each stimulus (*left panel, bottom*). The location of this line is calculated from the maximum pitch acceleration. Group comparisons of pitch strength are shown within the region of maximum acceleration of the stimulus derived from FFR responses to click-train stimuli as a function of pitch acceleration (*right panel*). FFR pitch strength of the Chinese group—as measured by the magnitude of the normalized autocorrelation peak—is greater than that of the English in response to Mandarin Tone 2 (*A1*, 0.3 Hz/ms) as well as to a pitch pattern that does not occur in natural speech (*A4*, 2.7 Hz/ms). In the English group, pitch strength shows a steady, steep decline across the continuum, approaching zero at its opposite end (i.e., the absence of a phase-locked response). In the Chinese group, on the other hand, pitch strength exhibits a more gradual decline, but never approaches zero. Instead, pitch strength begins to level off once a pitch pattern moves clearly beyond the normal voice range (*A3*, 1.3 Hz/ms). (Adapted from Krishnan et al. 2010c, with permission of Elsevier B.V.)

pitch acceleration ranging from low (0.3 Hz/ms; Mandarin Tone 2) to high (2.7 Hz/ms; 2 octaves) were presented to Chinese and English listeners (Krishnan et al. 2010c). Regardless of language group, neural periodicity strength is greater in response to acceleration rates within or proximal to natural speech relative to those beyond its range (Fig. 3.9, right panel). Though both groups show decreasing pitch strength with increasing acceleration rates, pitch representations of the Chinese group are more resistant to degradation. These findings indicate that perceptually salient pitch cues associated with lexical tone influence brainstem pitch extraction not only in the speech domain but also in auditory signals that clearly fall outside the range of dynamic pitch to which a native listener is exposed.

While the focus of this review is on the influence of long-term language experience on the neural representation of pitch-relevant information in the auditory brainstem, there is growing empirical evidence suggesting that it also shapes pitch mechanisms at early sensory levels of processing in the auditory cortex, and that the hemispheric preference for processing pitch information may vary depending on the relative linguistic importance of specific temporal attributes of dynamic pitch (Krishnan et al. 2014a, b). As in the brainstem, cortical responses sharpen properties of neural elements to enable optimal representation of temporal attributes of native pitch contours.

## 3.3 Hierarchical Processing Underlies Experience-Dependent Pitch Processing

Human functional magnetic resonance imaging and lesion studies have suggested that an area in the vicinity of lateral Heschl's gyrus (HG) is specialized for pitch representation (Griffiths and Hall 2012). This region also appears to be important for computations relevant to the extraction of pitch in complex sounds (Zatorre and Belin 2001; Schönwiesner et al. 2005). However, growing evidence shows pitch-related neural activity in both primary auditory cortex as well as in the adjacent more lateral non-primary areas of HG. These studies suggest that pitch-relevant information is available in multiple areas of the auditory cortex: functional magnetic resonance imaging (Griffiths et al. 2010; Puschmann et al. 2010), direct cortical recordings (Patterson et al. 2002; Penagos et al. 2004), and magnetoencephalography (Krumbholz et al. 2003; Gutschalk et al. 2004).

How do these areas interact in a coordinated manner during pitch processing? To address this question, a predictive coding model of perception has been applied to depth-electrode recordings of pitch-relevant neural activity along HG (Rao and Ballard 1999; Kumar and Schönwiesner 2012). Essentially, the model of predictive coding of pitch posits that cortical areas contributing to pitch are organized hierarchically. The effective connection strengths between and within levels are continually adjusted in a recursive manner to optimize pitch representation at the higher level. Operationally, higher-level areas in the hierarchy contributing to pitch (lateral HG) use stored information of pitch to make a pitch prediction. This prediction is passed to the lower areas in the processing hierarchy (medial and middle HG) via top down connection(s) (Fig. 3.10, blue arrows, #*1*). The lower areas of the hierarchy then compute a prediction error (difference between the higher level prediction and the lower level representation), which is passed to the higher level via bottom-up connections (Fig. 3.10, red arrows, #*2*). The strength of these connections is continually adjusted in a recursive manner in order to minimize predictive error and to optimize representation at the higher level. The lateral connections (same level in the hierarchy) between middle and medial HG (Fig. 3.10, green bidirectional arrows, #*3*) are also subject to modulation and presumably play a role in reducing redundancy and making representations more efficient.

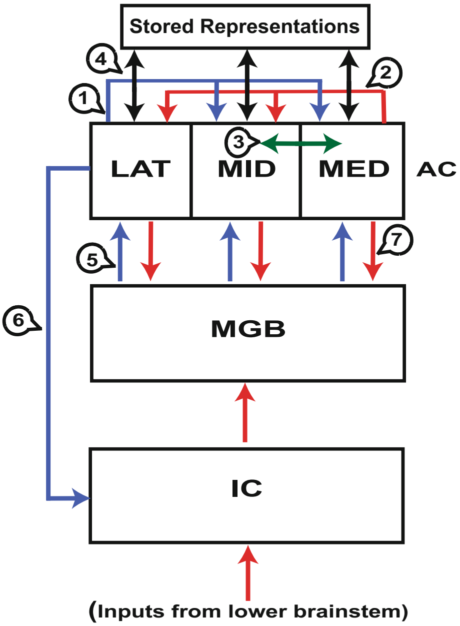


**Fig. 3.10** Block diagram of the proposed predictive coding model for hierarchical processing of pitch at both subcortical and cortical levels. Operationally, higher-level areas in the hierarchy contributing to pitch (lateral HG) use stored representations of pitch to make a pitch prediction. This prediction is passed to the lower areas in the processing hierarchy (medial and middle HG) via top-down connection(s) (*blue arrows, #1*). The lower areas then compute a prediction error (difference between the higher level prediction and the lower level representation), which is passed to the higher level via bottom-up connections (*red arrows, #2*). The lateral connections (same level in the hierarchy) between middle and medial HG (*green bidirectional arrows, #3*) are also subject to modulation and presumably play a role in reducing redundancy and making representations more efficient. Inputs from subcortical (bottom-up) structures (*#5*) that are themselves subject to experience-dependent plasticity are presumably mediated by top-down connections (*#6*, *#7*). It is likely that these top-down connections in the hierarchy provide feedback to adjust the effective time scales of processing at each stage to optimally control the temporal dynamics of pitch processing. Language dependent changes at the early sensory level of processing in the auditory cortex may reflect interplay between sensory and cognitive processing (*black bidirectional arrows*, *#4*)

Dynamic causal modeling essentially tries to determine how the activity of one brain area changes the dynamics and/or responses of other areas. Using Bayesian model comparisons to determine the configuration(s) that best explain the data, Kumar and Schönwiesner (2012) showed that the lateral part of HG is at a higher level in the hierarchy compared to middle and medial HG, with the latter two at the same level. This is in agreement with evidence from depth-electrode recordings along HG. It confirms that middle and medial electrode contacts are indeed in the primary auditory cortex, whereas the lateral contacts are in non-primary auditory cortex (Brugge et al. 2009). Consistent with the predictions of the model, they also show that strength of connectivity varies with pitch salience such that the strength of the top-down connection from lateral HG to medial and middle HG increases with pitch salience, whereas the strength of the bottom-up connection from middle HG to lateral HG decreases. This distributed view of pitch processing, however, is not necessarily at odds with a single specialized pitch center. It is likely that lateral HG has more pitch-specific mechanisms and, therefore, plays a relatively greater role in pitch perception.

In this case, the predictive coding of the pitch model provides a framework to explain the language-dependent (cognitive) and language-universal (sensory) effects on pitch-related neural activity in the brainstem and auditory cortex. Changes at different stages of processing attributable wholly to acoustic properties of the stimulus implicate the recursive process (initial pitch prediction, error generation, error correction) in the representation of pitch. At this fundamental level of pitch processing, the hierarchical flow of processing and the connectivity strengths along the HG are essentially the same regardless of one's language background. For changes that are dependent on language experience, the initial pitch prediction at the level of the lateral HG is more precise for Chinese because of their access to stored information about lexical tones with a smaller error term. Consequently, the top-down connections from lateral HG to medial and middle HG and to the brainstem are stronger than the bottom-up connections from the medial and middle HG to the lateral HG and from the brainstem. The opposite would be true for English because of their less precise initial prediction. In addition, the recursive process itself would be expected to take a longer time for English relative to Chinese in determining pitch. Language experience, therefore, alters the nature of the interaction between functional components of the distributed network by modulating connection strengths.

It is clear that pitch processing in the auditory cortex is influenced by inputs from subcortical structures (Fig. 3.10, *#5*) that are themselves subject to experience-dependent plasticity, presumably mediated by top-down connections (Fig. 3.10, *#6*, *#7*). It is likely that these top-down connections in the hierarchy provide feedback to adjust the effective time scales of processing at each stage to optimally control the temporal dynamics of pitch processing (Balaguer-Ballester et al. 2009). Language-dependent changes at the sensory level of processing in the auditory cortex may reflect interplay between sensory and cognitive processing

(Fig. 3.10, black bidirectional arrows, #*4*). This model represents a unified, physiologically plausible, theoretical framework that includes both cortical and subcortical components in the hierarchical processing of pitch.

## 3.4 Summary and Directions for Future Research

The evidence presented herein suggests that neural activity relevant to pitch in the brainstem is crucially dependent on specific dimensions or features of pitch contours. By focusing on specific properties of the auditory signal, irrespective of a speech or nonspeech context, it is argued that the neural representation of acoustic-phonetic features relevant to speech perception is already emerging in the brainstem (≈8–10 ms after stimulus onset) and, importantly, can be shaped by experience. These sensory level auditory processes are tuned differentially to those features depending upon their linguistic relevance. Such effects of language experience on lower level sensory processing are compatible with a more integrated approach to language and the brain (Hickok and Poeppel 2004; Zatorre and Gandour 2008). The focus on pitch processing in tonal languages notwithstanding, these findings should be contextualized within the broader framework of language experience shaping subcortical processing.

While language experience shapes pitch processing at both subcortical and cortical levels, neural representations are transformed and fundamentally different at each biological level of abstraction. The representation of pitch-relevant information in the brainstem is more fine-grained spectrotemporally as it reflects sustained neural phase locking to pitch-relevant periodicities contained in the dynamic stimulus. In contrast, the cortical representation is coarser. That is, the cortical pitch-relevant neural activity reflects primarily a series of distinct transient temporal neural events marking only certain temporal attributes of the pitch contour. These differences notwithstanding, long-term language experience appears to shape adaptive, hierarchical pitch processing. Top-down connections provide selective gating of inputs to both cortical and subcortical structures to enhance neural representation of behaviorally relevant attributes of the stimulus and instantiate local mechanisms that exhibit enhanced representation of behaviorally relevant pitch attributes. The ability to record brainstem and cortical pitch-relevant responses concurrently may also provide a new window to evaluate the online interplay between feedforward and feedback components in the processing of pitch-relevant information at the level of the brainstem and the auditory cortex. While it is not known how language experience shapes subcortical and cortical stages of pitch processing, it is likely that the neural processes underlying such experience-dependent plasticity at each stage along the processing hierarchy are modulated by a coordinated interplay between ascending, descending, and local neural pathways that involve both sensory and cognitive components. The challenge is to develop experiments that systematically manipulate pitch attributes in order to optimally evaluate the relationship between representation of pitch-relevant information at the brainstem and cortical levels.

The results of these experiments are essential to further our understanding of the nature of interplay between cortical and subcortical functional components and the interactions between sensory and cognitive processes influencing pitch representation in what appears to be an integrated distributed processing network. Complementary studies using magnetoencephalography will be crucial to determine the anatomical sources of these components in an effort to shed more light on specific cortical generators contributing to pitch processing and how experience may shape these processes.

**Acknowledgments** The research was supported by NIH 5R01DC008549-06 (A.K.).

**Compliance with Ethics Requirements** Ananthanarayan Krishnan and Jackson T. Gandour declared that they had no conflicts of interest.

## References

Abrams, D. A., Nicol, T., Zecker, S. G., & Kraus, N. (2006). Auditory brainstem timing predicts cerebral asymmetry for speech. *The Journal of Neuroscience, 26*(43), 11131–11137. Doi:10.1523/JNEUROSCI.2744-06.2006. (26/43/11131 [pii]).

Abramson, A. S. (1962). *The vowels and tones of standard Thai: Acoustical measurements and experiments.* Research Center in Anthropology, Folklore, and Linguistics, Pub. 20. Bloomington, IN: Indiana University.

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech, 21*, 319–325.

Akhoun, I., Gallégo, S., Moulin, A., Ménard, M., 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.

Balaguer-Ballester, E., Clark, N. R., Coath, M., Krumbholz, K., & Denham, S. L. (2009). Understanding pitch perception as a hierarchical process with top-down modulation. *PLoS Computational Biology, 5*(3), e1000301. Doi:10.1371/journal.pcbi.1000301

Banai, K., Abrams, D., & Kraus, N. (2007). Sensory-based learning disability: Insights from brainstem processing of speech sounds. *International Journal of Audiology, 46*(9), 524–532. Doi:10.1080/14992020701383035. (781872134 [pii]).

Besson, M., Chobert, J., & Marie, C. (2011). Language and music in the musician brain. *Language and Linguistics Compass, 5*(9), 617–634. Doi:10.1111/j.1749-818x.2011.00302

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011a). Musicians demonstrate experience-dependent brainstem enhancement of musical scale features within continuously gliding pitch. *Neuroscience Letters, 503*(3), 203–207. Doi:10.1016/j.neulet.2011.08.036

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011b). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition, 77*(1), 1–10. Doi:10.1016/j.bandc.2011.07.006

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011c). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience, 23*(2), 425–434. Doi:10.1162/jocn.2009.21362

Bidelman, G. M., & Krishnan, A. (2011). Brainstem correlates of behavioral and compositional preferences of musical harmony. *NeuroReport, 22*(5), 212–216. Doi:10.1097/WNR.0b013e328344a689

Brugge, J. F., Nourski, K. V., Oya, H., Reale, R. A., et al. (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *Journal of Neurophysiology, 102*(4), 2358–2374. Doi:10.1152/jn.91346.2008

Cariani, P. A., & Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *Journal of Neurophysiology, 76*(3), 1717–1734.

Cariani, P. A., & Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology, 76*(3), 1698–1716.

Cedolin, L., & Delgutte, B. (2005). Pitch of complex tones: Rate-place and interspike interval representations in the auditory nerve. *Journal of Neurophysiology, 94*(1), 347–362. Doi:10.1152/jn.01114.2004. (01114.2004 [pii]).

Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007a). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience, 25*(3–4), 195–210.

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology, 47*(2), 236–246. Doi:10.1111/j.1469-8986.2009.00928.x. (PSYP928 [pii]).

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007b). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128*(1), 148–156. Doi:10.1016/j.brainres.2006.10.064. (S0006-8993(06)03179-9 [pii]).

Cohen, M. A., Grossberg, S., & Wyse, L. L. (1995). A spectral network model of pitch perception. *The Journal of the Acoustical Society of America, 98*(2), 862–879.

Collins, M. J., & Cullen, J. K., Jr. (1978). Temporal integration of tone glides. *The Journal of the Acoustical Society of America, 63*(2), 469–473.

Galbraith, G. C. (2008). Deficient brainstem encoding in autism. *Clinical Neurophysiology, 119*(8), 1697–1700. Doi:10.1016/j.clinph.2008.04.012

Galbraith, G. C., Amaya, E. M., de Rivera, J. M., Donan, N. M., et al. (2004). Brain stem evoked response to forward and reversed speech in humans. *NeuroReport, 15*(13), 2057–2060.

Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics, 11*, 149–175.

Gandour, J. T. (1994). Phonetics of tone. In R. Asher & J. Simpson (Eds.), *The encyclopedia of language & linguistics* (Vol. 6, pp. 3116–3123). New York: Pergamon Press.

Gandour, J. T., & Harshman, R. A. (1978). Cross language differences in tone perception: A multidimensional scaling investigation. *Language and Speech, 21*(1), 1–33.

Gandour, J. T., & Krishnan, A. (2014). Neural bases of lexical tone. In H. Winskel & P. Padakannaya (Eds.), *Handbook of South and Southeast Asian psycholinguistics* (pp. 339–349). Cambridge, UK: Cambridge University Press.

Gockel, H. E., Carlyon, R. P., Mehta, A., & Plack, C. J. (2011). The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. *Journal of the Association for Research in Otolaryngology, 12*(6), 767–782. Doi:10.1007/s10162-011-0284-1

Gold, J. I., & Knudsen, E. I. (2000). A site of auditory experience-dependent plasticity in the neural representation of auditory space in the barn owl's inferior colliculus. *Journal of Neuroscience, 20*(9), 3469–3486.

Greenberg, S., Marsh, J. T., Brown, W. S., & Smith, J. C. (1987). Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hearing Research, 25*(2–3), 91–114.

Griffiths, T. D., Buchel, C., Frackowiak, R. S., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience, 1*(5), 422–427.

Griffiths, T. D., & Hall, D. A. (2012). Mapping pitch representation in neural ensembles with fMRI. *The Journal of Neuroscience, 32*(39), 13343–13347. Doi:10.1523/JNEUROSCI.3813-12.2012

Griffiths, T. D., Kumar, S., Sedley, W., Nourski, K. V., et al. (2010). Direct recordings of pitch responses from human auditory cortex. *Current Biology, 20*(12), 1128–1132. Doi:10.1016/j.cub.2010.04.044

Griffiths, T. D., Uppenkamp, S., Johnsrude, I., Josephs, O., & Patterson, R. D. (2001). Encoding of the temporal regularity of sound in the human brainstem. *Nature Neuroscience, 4*(6), 633–637.

Griffiths, T. D., Warren, J. D., Scott, S. K., Nelken, I., & King, A. J. (2004). Cortical processing of complex sound: A way forward? *Trends in Neurosciences, 27*(4), 181–185. Doi:10.1016/j.tins.2004.02.005. (S0166223604000566 [pii]).

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., & Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *Neuroimage, 22*(2), 755–766. Doi:10.1016/j.neuroimage.2004.01.025. (S1053811904000680 [pii]).

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., & Rupp, A. (2007). The effect of temporal context on the sustained pitch response in human auditory cortex. *Cerebral Cortex, 17*(3), 552–561. Doi:10.1093/cercor/bhj180

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92*(1–2), 67–99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393–402. Doi:10.1038/nrn2113. (nrn2113 [pii]).

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. New York: Cambridge University Press.

Johnson, K. L., Nicol, T., Zecker, S. G., & Kraus, N. (2008). Developmental plasticity in the human auditory brainstem. *The Journal of Neuroscience, 28*(15), 4000–4007. Doi:10.1523/JNEUROSCI.0012-08.2008

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics, 35*(1), 104–117. Doi:10.1016/j.wocn.2005.10.003

Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends in Neurosciences, 28*(4), 176–181.

Kraus, N., & White-Schwoch, T. (2015). Unraveling the biology of auditory learning: A cognitive-sensorimotor-reward framework. *Trends in Cognitive Sciences, 19*(11), 642–654. Doi:10.1016/j.tics.2015.08.017

Krishnan, A. (1999). Human frequency-following responses to two-tone approximations of steady-state vowels. *Audiology and Neuro-Otology, 4*(2), 95–103.

Krishnan, A. (2002). Human frequency-following responses: Representation of steady-state synthetic vowels. *Hearing Research, 166*(1–2), 192–201.

Krishnan, A. (2007). Human frequency following response. In R. F. Burkard, M. Don, & J. J. Eggermont (Eds.), *Auditory evoked potentials: Basic principles and clinical application* (pp. 313–335). Baltimore: Lippincott Williams & Wilkins.

Krishnan, A., Bidelman, G. M., & Gandour, J. T. (2010a). Neural representation of pitch salience in the human brainstem revealed by psychophysical and electrophysiological indices. *Hearing Research, 268*(1–2), 60–66. Doi:10.1016/j.heares.2010.04.016

Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically relevant pitch patterns. *Brain and Language, 110*(3), 135–148. Doi:10.1016/j.bandl.2009.03.005. (S0093-934X(09)00042-X [pii]).

Krishnan, A., & Gandour, J. T. (2014). Language experience shapes processing of pitch relevant information in the human brainstem and auditory cortex: Electrophysiological evidence. *Acoustics Australia, 42*(3), 187–199

Krishnan, A., Gandour, J. T., Ananthakrishnan, S., & Vijayaraghavan, V. (2014a). Cortical pitch response components index stimulus onset/offset and dynamic features of pitch contours. *Neuropsychologia, 59*, 1–12. Doi:10.1016/j.neuropsychologia.2014.04.006

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010b). Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. *Brain Research, 1313*, 124–133. Doi:10.1016/j.brainres.2009.11.061

Krishnan, A., Gandour, J. T., Bidelman, G. M., & Swaminathan, J. (2009a). Experience-dependent neural representation of dynamic pitch in the brainstem. *NeuroReport, 20*(4), 408–413. Doi:10.1097/WNR.0b013e3283263000

Krishnan, A., Gandour, J. T., Smalt, C. J., & Bidelman, G. M. (2010c). Language-dependent pitch encoding advantage in the brainstem is not limited to acceleration rates that occur in natural speech. *Brain and Language, 114*(3), 193–198. Doi:10.1016/j.bandl.2010.05.004

Krishnan, A., Gandour, J. T., & Suresh, C. H. (2014b). Cortical pitch response components show differential sensitivity to native and nonnative pitch contours. *Brain and Language, 138*, 51–60. Doi:10.1016/j.bandl.2014.09.005

Krishnan, A., & Parkinson, J. (2000). Human frequency-following response: Representation of tonal sweeps. *Audiology and Neuro-Otology, 5*(6), 312–321.

Krishnan, A., & Plack, C. J. (2011). Neural encoding in the human brainstem relevant to the pitch of complex tones. *Hearing Research, 275*(1–2), 110–119. Doi:10.1016/j.heares.2010.12.008

Krishnan, A., Swaminathan, J., & Gandour, J. T. (2009b). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience, 21*(6), 1092–1105. Doi:10.1162/jocn.2009.21077

Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Research. Cognitive Brain Research, 25*(1), 161–168. Doi:10.1016/j.cogbrainres.2005.05.004. (S0926-6410(05)00123-0 [pii]).

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*(1–2), 1–12. Doi:10.1016/S0378-595(03)00402-7. (S037859503004027 [pii]).

Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences of the USA, 109*(20), 7877–7881. Doi:10.1073/pnas.1201575109

Krizman, J., Skoe, E., Marian, V., & Kraus, N. (2014). Bilingualism increases neural response consistency and attentional control: Evidence for sensory and cognitive coupling. *Brain and Language, 128*(1), 34–40. Doi:10.1016/j.bandl.2013.11.006

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lutkenhoner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex, 13*(7), 765–772.

Kumar, S., & Schönwiesner, M. (2012). Mapping human pitch representation in a distributed system using depth-electrode recordings and modeling. *The Journal of Neuroscience, 32*(39), 13348–13351. Doi:10.1523/JNEUROSCI.3812-12.2012

Kumar, S., Stephan, K. E., Warren, J. D., Friston, K. J., & Griffiths, T. D. (2007). Hierarchical processing of auditory objects in humans. *PLoS Computational Biology, 3*(6), e100. Doi:10.1371/journal.pcbi.0030100

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*(11), 1131–1138. Doi:10.1038/nn737

Luo, F., Wang, Q., Kashani, A., & Yan, J. (2008). Corticofugal modulation of initial sound processing in the brain. *The Journal of Neuroscience, 28*(45), 11615–11621. Doi:10.1523/JNEUROSCI.3972-08.2008. (28/45/11615[pii]).

Meddis, R., & O'Mard, L. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America, 102*(3), 1811–1820.

Micheyl, C., Schrater, P. R., & Oxenham, A. J. (2013). Auditory frequency and intensity discrimination explained using a cortical population rate code. *PLoS Computational Biology, 9*(11), e1003336. Doi:10.1371/journal.pcbi.1003336

Oxenham, A. J. (2012). Pitch perception. *The Journal of Neuroscience, 32*(39), 13335–13338. Doi:10.1523/JNEUROSCI.3815-12.2012

Patterson, R. D., Allerhand, M. H., & Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America, 98*(4), 1890–1894.

Patterson, R. D., Handel, S., Yost, W. A., & Datta, A. J. (1996). The relative strength of the tone and noise components in iterated ripple noise. *The Journal of the Acoustical Society of America, 100*(5), 3286–3294.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron, 36*(4), 767–776.

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *The Journal of Neuroscience, 24*(30), 6810–6815. Doi:10.1523/JNEUROSCI.0383-04.2004. (24/30/6810 [pii]).

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., et al. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics, 38*, 616–624.

Philibert, B., Collet, L., Vesson, J. F., & Veuillet, E. (2005). The auditory acclimatization effect in sensorineural hearing-impaired listeners: Evidence for functional plasticity. *Hearing Research, 205*(1–2), 131–142.

Plyler, P. N., & Ananthanarayan, A. K. (2001). Human frequency-following responses: Representation of second formant transitions in normal-hearing and hearing-impaired listeners. *Journal of the American Academy of Audiology, 12*(10), 523–533.

Poeppel, 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 B: Biological Sciences, 363*(1493), 1071–1086. Doi:10.1098/rstb.2007.2160. (TM42571U1117682 [pii]).

Price, C., Thierry, G., & Griffiths, T. (2005). Speech-specific auditory processing: Where is it? *Trends in Cognitive Sciences, 9*(6), 271–276.

Puschmann, S., Uppenkamp, S., Kollmeier, B., & Thiel, C. M. (2010). Dichotic pitch activates pitch processing centre in Heschl's gyrus. *Neuroimage, 49*(2), 1641–1649. Doi:10.1016/j.neuroimage.2009.09.045

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79–87. Doi:10.1038/4580

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*(1), 95–103.

Russo, N. M., Skoe, E., Trommer, B., Nicol, T., et al. (2008). Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clinical Neurophysiology, 119*(8), 1720–1731. Doi:10.1016/j.clinph.2008.01.108

Schönwiesner, M., Rubsamen, R., & von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience, 22*(6), 1521–1528. Doi:10.1111/j.1460-9568.2005.04315.x

Schouten, M. E. (1985). Identification and discrimination of sweep tones. *Perception and Psychophysics, 37*(4), 369–376.

Scott, S. K. (2003). How might we conceptualize speech perception? The view from neurobiology. *Journal of Phonetics, 31*, 417–422.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences, 26*(2), 100–107.

Shofner, W. P. (1999). Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: Analysis of neural autocorrelograms. *Journal of Neurophysiology, 81*(6), 2662–2674.

Shore, S. E., Clopton, B. M., & Au, Y. N. (1987). Unit responses in ventral cochlear nucleus reflect cochlear coding of rapid frequency sweeps. *Journal of the Acoustical Society of America, 82*(2), 471–478.

Shore, S. E., & Nuttall, A. L. (1985). High-synchrony cochlear compound action potentials evoked by rising frequency-swept tone bursts. *The Journal of the Acoustical Society of America, 78*(4), 1286–1295.

Song, J. H., Banai, K., & Kraus, N. (2008a). Brainstem timing deficits in children with learning impairment may result from corticofugal origins. *Audiology and Neuro-Otology, 13*(5), 335–344. Doi:10.1159/000132689

Song, J. H., Skoe, E., Wong, P. C. M., & Kraus, N. (2008b). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience, 20*(10), 1892–1902.

Steinschneider, M., Reser, D. H., Fishman, Y. I., Schroeder, C. E., & Arezzo, J. C. (1998). Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception. *The Journal of the Acoustical Society of America, 104*(5), 2935–2955.

Suga, N. (1990). Biosonar and neural computation in bats. *Scientific American, 262*(6), 60–68.

Suga, N. (1994). Processing of auditory information carried by complex species specific sounds. In M. S. Gazzaniga & E. Bizzi (Eds.), *The cognitive neurosciences* (pp. 295–318). Cambridge, MA: MIT Press.

Suga, N., & Ma, X. (2003). Multiparametric corticofugal modulation and plasticity in the auditory system. *Nature Reviews Neuroscience, 4*(10), 783–794.

Suga, N., Ma, X., Gao, E., Sakai, M., & Chowdhury, S. A. (2003). Descending system and plasticity for auditory signal processing: Neuroethological data for speech scientists. *Speech Communication, 41*(1), 189–200.

Swaminathan, J., Krishnan, A., & Gandour, J. T. (2008). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *NeuroReport, 19*(11), 1163–1167. Doi:10.1097/WNR.0b013e3283088d31. (00001756-200807160-00017 [pii]).

Walker, K. M., Bizley, J. K., King, A. J., & Schnupp, J. W. (2011). Cortical encoding of pitch: Recent results and open questions. *Hearing Research, 271*(1–2), 74–87. Doi:10.1016/j.heares.2010.04.015

Wong, P. C. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin, 59*(2), 83–95.

Wong, P. C., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *The Journal of Neuroscience, 24*(41), 9153–9160.

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*(4), 420–422. Doi:10.1038/nn1872. (nn1872 [pii]).

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*, 61–83.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006a). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America, 120*(2), 1063–1074.

Xu, Y., Krishnan, A., & Gandour, J. T. (2006b). Specificity of experience-dependent pitch representation in the brainstem. *NeuroReport, 17*(15), 1601–1605. Doi:10.1097/01.wnr.0000236865.31705.3a. (00001756-200610230-00008 [pii]).

Yip, M. (2002). *Tone*. New York: Cambridge University Press.

Yost, W. A. (1996). Pitch of iterated rippled noise. *The Journal of the Acoustical Society of America, 100*(1), 511–518.

Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biology, 10*(7), e1001372. Doi:10.1371/journal.pbio.100137

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*(10), 946–953.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363*(1493), 1087–1104. Doi:10.1098/rstb.2007.2161. (J412P80575385013 [pii]).