Mini-Review

# Pitch Perception

**Andrew J. Oxenham**

Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455

Pitch is one of the primary auditory sensations and plays a defining role in music, speech, and auditory scene analysis. Although the main physical correlate of pitch is acoustic periodicity, or repetition rate, there are many interactions that complicate the relationship between the physical stimulus and the perception of pitch. In particular, the effects of other acoustic parameters on pitch judgments, and the complex interactions between perceptual organization and pitch, have uncovered interesting perceptual phenomena that should help to reveal the underlying neural mechanisms.

## The what and why of pitch

Pitch is one of the primary auditory sensations, along with loudness and timbre. In music, sequences of pitch define melody, and simultaneous combinations of pitch define harmony. In speech, rising and falling pitch contours help define prosody and in tone languages, such as Mandarin and Cantonese, pitch contours help define the meaning of words. In complex acoustic environments, differences in pitch can help listeners to segregate and make sense of competing sound sources.

Put simply, pitch is the perceptual correlate of the periodicity, or repetition rate, of an acoustic waveform. The most commonly considered form of pitch-evoking sound is a harmonic complex tone. This periodic waveform repeats at a rate corresponding to the fundamental frequency (F0) and can be decomposed into sinusoidal harmonics or overtones, which have frequencies at integer multiples of the F0 (Fig. 1 A, B). The relative amplitudes of the harmonics within a complex tone play an important role in determining the sound quality, or timbre, of a sound. Despite differences in timbre and loudness, two tones generally have the same pitch if they share the same F0. Although young humans with normal hearing can hear sounds with frequencies between ~20 and 20,000 Hz, only repetition rates between ~30 and 4000 Hz elicit a pitch sensation that is salient enough to carry melodic information (Attneave and Olson, 1971; Pressnitzer et al., 2001).

Scientific disputes on how we perceive the F0 arose in the mid-19th Century (Seebeck, 1841; Ohm, 1843; Helmholtz, 1885/1954), but it was only firmly established in the mid-20th Century that a tone retains the same pitch, even if all the energy at the F0 is removed or masked by noise (Schouten, 1940; Licklider, 1954). This phenomenon of the "pitch of the missing fundamental" provides an important benchmark in the search for physiological correlates of pitch (Plack et al., 2005; Griffiths and Hall, 2012; Wang and Walker, 2012). From a perceptual standpoint, it makes sense that the pitch of a sound remains constant after the lowest harmonic components are removed or masked (occluded), so that some degree of perceptual invariance of a sound source can

be maintained in a cluttered acoustic environment (McDermott and Oxenham, 2008).

## Relationship between pitch and early auditory transformations

Figure 1 A shows the time waveform of a musical tone with an F0 of 440 Hz. As shown in Figure 1 B, when a time segment of a waveform is analyzed, its power spectrum—the distribution of sound intensity across the frequency spectrum—can be extracted.

### Cochlear filtering

When the sound enters the cochlea, different frequencies within the sound selectively stimulate different regions of the cochlea. This frequency-to-place mapping, or tonotopy, is maintained throughout the auditory pathways up to at least primary auditory cortex (AI), and forms a major organizational principle of auditory neural processing. The perceptual consequences of tonotopic organization are manifold and can be measured using behavioral experiments in a variety of ways, often involving masking (Oxenham and Wojtczak, 2010). The results from such experiments are explained in terms of the frequency selectivity of the "auditory filters" (Fig. 1C). These behaviorally defined filters are thought to have their basis in cochlear filtering (Shera et al., 2002, 2010). The output of the auditory filters can be represented in terms of the long-term average, which is referred to as the "excitation pattern" (Fig. 1 D), which can be thought of as a schematic representation of mechanical activation of the cochlear partition or neural activity as a function of characteristic frequency (CF) (Glasberg and Moore, 1990). Alternatively, the outputs of the auditory filters can be considered in terms of their time waveform (Fig. 1 E). Because the filter bandwidths increase with increasing CF, regions of the cochlea tuned to the frequencies of low-numbered harmonics will respond almost exclusively to a single harmonic, whereas regions tuned to the frequencies of high-numbered harmonics will respond to several harmonics. Harmonics that are exclusively represented within single filters are referred to as "resolved," whereas harmonics that interact with others within auditory filters are referred to as "unresolved."

Resolved harmonics produce peaks in the excitation pattern (Fig. 1 D), and should elicit filtered waveforms that are similar to single pure tones at that frequency (Fig. 1 E), whereas unresolved

harmonics produce no distinct peaks, and elicit complex waveforms that reflect the interaction between multiple harmonics. The point of transition between resolved and unresolved harmonics is somewhat fuzzy and depends on many factors, including sound level and F0, as well as on how resolvability is defined and measured (Bernstein and Oxenham, 2003; Moore and Gockel, 2011). Nevertheless, a number of phenomena related to pitch can be explained in terms of harmonic resolvability (Flanagan and Guttman, 1960; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2006a,b).

*Relationship of auditory-nerve and brainstem responses to pitch*
Neurons in the auditory nerve are more likely to fire at one phase within the cycle of a waveform than at other phases. This precise stimulus-driven spike timing, or phase locking, is known to underlie our ability to perceive minute differences in the time of arrival of sound at the two ears, down to as little as 10 $\mu$s, which in turn helps us to localize sounds in space (Blauert, 1997). It is entirely reasonable that the same precise timing is used to help encode stimulus periodicity and hence pitch. In species such as cat and guinea pig, phase locking is known to extend to ~1–2 kHz, and then to degrade at higher frequencies (Palmer and Russell, 1986). Because of the invasive nature of the measurements, little is known about phase locking in the human auditory nerve. Nevertheless, behavioral data, showing that the ability to discriminate frequency and to recognize melodies degrades for pure tones above ~4–5 kHz (Attneave and Olson, 1971; Moore, 1973), has been interpreted as reflecting the degradation of phase locking at high frequencies, suggesting that accurate pitch perception relies on timing information in the auditory nerve. There are some indications that timing information alone may not be sufficient to provide accurate pitch: Oxenham et al. (2004) found that when the timing information from low-frequency harmonics was "transposed" to a higher-frequency cochlear location, listeners were not able to use the information to extract the F0. This result suggested that complex pitch was based either on tonotopic or "place" information, or on timing information that had to be presented to the "correct" place along the cochlea, as required by spatio-temporal models of periodicity coding (Loeb et al., 1983; Shamma and Klein, 2000; Cedolin and Delgutte, 2010; Carlyon et al., 2012).
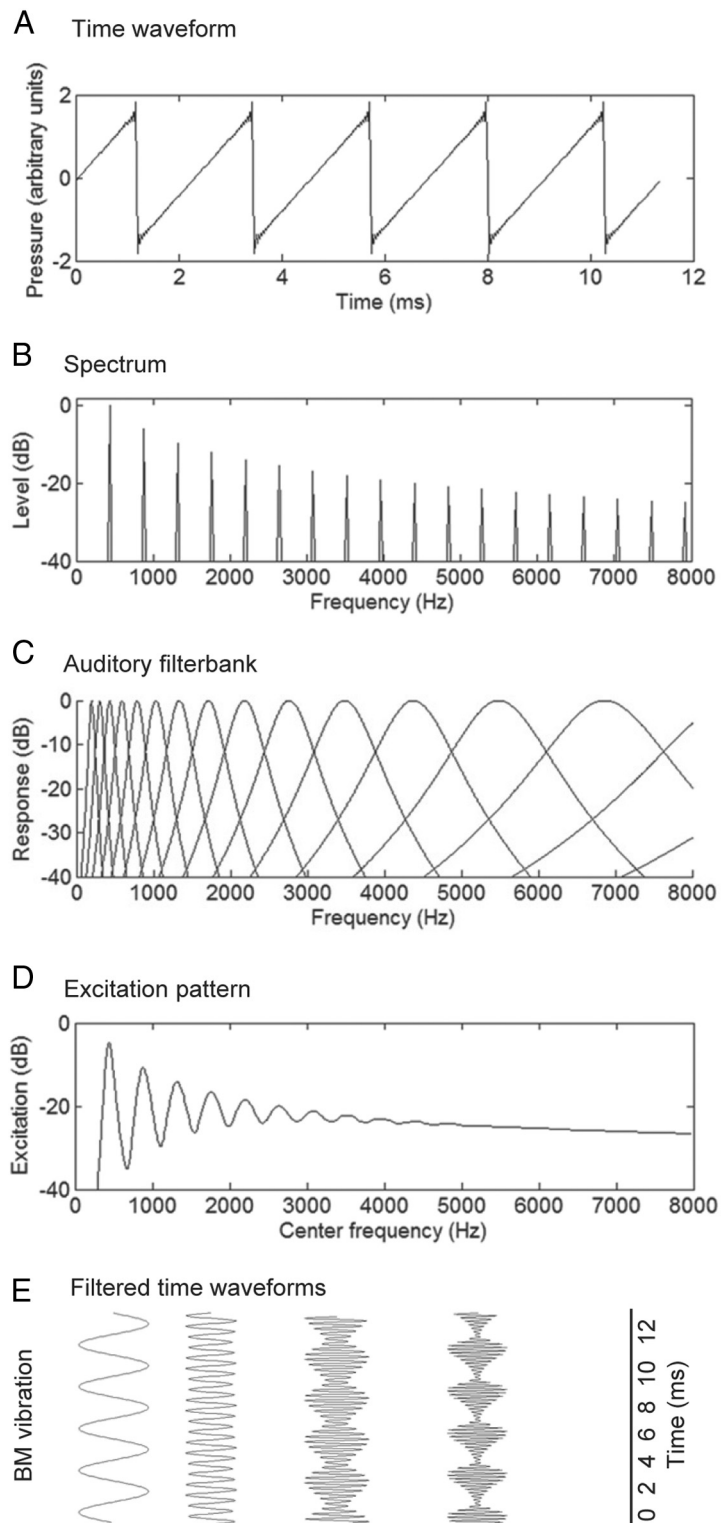


**Figure 1.** Representations of a harmonic complex tone with a F0 of 440 Hz. *A*, Time waveform. *B*, Power spectrum of the same waveform. *C*, Auditory filter bank, representing the filtering that occurs in the cochlea. *D*, Excitation pattern, or the time-averaged output of the auditory filters. *E*, Sample time waveforms at the output of the filter bank, including filters centered at the F0 and the fourth harmonic, illustrating resolved harmonics, and filters centered at the eighth and 12th harmonic of the complex, illustrating harmonics that are less well resolved and show amplitude modulations at a rate corresponding to the F0.

How is pitch extracted? One way might be via the temporal patterns generated by unresolved harmonics in the auditory periphery (Schouten et al., 1962). These provide a direct estimate of the F0, assuming that the brain can measure the timing between

peaks of neural activity corresponding to peaks in the time waveform. In contrast, extracting pitch from the resolved harmonics may require coding of the individual frequencies and then combining that information to estimate the F0 (Schroeder, 1968; Goldstein, 1973). At face value, therefore, the unresolved harmonics provide the easier route to code F0. On the other hand, phase distortions produced by room acoustics and reverberation can change the waveform of complex tones, which affects the representation of unresolved harmonics, but not resolved harmonics, making unresolved harmonics more susceptible to interference (Qin and Oxenham, 2005; Sayles and Winter, 2008). In fact, many behavioral studies have shown that low-numbered resolved harmonics elicit a much more salient, robust, and accurate pitch than do high-numbered unresolved harmonics (Houtsma and Smurzynski, 1990; Carlyon, 1996; Bernstein and Oxenham, 2003; Micheyl et al., 2010).

Pitch-matching and melody discrimination experiments have recently revealed that the pitch of the missing fundamental can be extracted even when all the harmonics present are well above 5 kHz (Oxenham et al., 2011). Because the pitch was extracted from the individual harmonics, and not from the temporal envelope produced by unresolved harmonics (Kaernbach and Bering, 2001), the results suggest either that temporal information is not necessary for complex pitch perception (i.e., harmonics can be represented via a place code), or that phase locking in the auditory nerve extends to much higher frequencies than is generally believed (see also Heinz et al., 2001; Recio-Spinoso et al., 2005; Moore and Sęk, 2009).

The frequency following response (FFR) is a measure of phase-locked brainstem activity that can be recorded from the scalp (Skoe and Kraus, 2010). It has been used as a measure of pitch encoding accuracy, and several intriguing findings have been reported recently, including stronger FFR in musicians than in people without musical training (Wong et al., 2007), stronger FFR in people with experience in tonal languages (such as Mandarin) than without (Krishnan et al., 2005), increased FFR amplitudes after training in a speech-related task (de Boer and Thornton, 2008; Tzounopoulos and Kraus, 2009), and correlations between FFR amplitudes and the ability to learn new pitch contours in a linguistic context (Chandrasekaran et al., 2012). However, we are only beginning to understand the relationship between the FFR and pitch perception (Gockel et al., 2011).

## Interactions between periodicity and other acoustic variables on pitch

Although pitch is often treated as being orthogonal to other perceptual dimensions, such as loudness and timbre, some interactions occur. For instance, small effects of stimulus intensity on pitch have been reported (Verschuure and van Meeteren, 1975). More commonly, many people find it difficult to ignore changes in brightness (produced by changes in the spectral content of a stimulus) when making pitch judgments (produced by changes in the F0) (Moore and Glasberg, 1990). In fact, a recent study reported that even many musically trained listeners find it difficult to detect small pitch differences between sounds with very different timbres (Borchert et al., 2011).

The fact that listeners find it difficult to make pitch judgments in the face of large timbral differences raises some question as to whether we should expect to find a cortical neural representation of pitch that is highly localized and invariant to changes in other dimensions. As discussed later in these mini-reviews (Wang and Walker, 2012), both localized pitch-specific invariant represen-

tations (Bendor and Wang, 2005), and more distributed population codes (Walker et al., 2011) have been proposed.

Other forms of interaction can aid comparisons between sound sequences. For instance, contours—the pattern of rising and falling pitch in melodies—have traditionally been considered specific to pitch, whereas recent findings suggest that listeners are not only able to perceive contours in dimensions other than pitch (i.e., loudness and timbre) but are also able to compare contours across perceptual dimensions, suggesting a common underlying representation (McDermott et al., 2008).

### Future directions

There are many indications that perceptual organization and object formation are affected by harmonicity and pitch, but also that pitch can be influenced by perceptual organization (Darwin, 2005). Thus, the search for neural correlates of pitch in cortical regions seems entirely reasonable and consistent with behavioral data. In addition, paradigms in which pitch can be altered by changes in perceptual grouping, in the absence of changes in stimulus periodicity, could provide a fruitful approach to dissociating neural correlates of pitch from those related strictly to physical stimulus properties.

People with hearing loss, and especially those with cochlear implants, often suffer from a deficit in pitch perception abilities (McDermott, 2004; Oxenham, 2008). A better understanding of the neural transformations involved in pitch extraction should help in designing more effective neural and acoustic prostheses.

Although the tonotopic representation of frequency is established early, it is relative, rather than absolute, pitch that is most salient and important for human acoustic communication in speech and music. Although some efforts to study the neural correlates of pitch relations and contours have been made (Patel and Balaban, 2000; Warren et al., 2003), this remains a field ripe for further study.

## References

Attneave F, Olson RK (1971) Pitch as a medium: a new approach to psychophysical scaling. Am J Psychol 84:147–166. CrossRef Medline

Bendor D, Wang X (2005) The neuronal representation of pitch in primate auditory cortex. Nature 436:1161–1165. CrossRef Medline

Bernstein JG, Oxenham AJ (2003) Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? J Acoust Soc Am 113:3323–3334. CrossRef Medline

Bernstein JG, Oxenham AJ (2006a) The relationship between frequency selectivity and pitch discrimination: effects of stimulus level. J Acoust Soc Am 120:3916–3928. CrossRef Medline

Bernstein JG, Oxenham AJ (2006b) The relationship between frequency selectivity and pitch discrimination: sensorineural hearing loss. J Acoust Soc Am 120:3929–3945. CrossRef Medline

Blauert J (1997) Spatial hearing: the psychophysics of human sound localization. Cambridge, MA: MIT.

Borchert EM, Micheyl C, Oxenham AJ (2011) Perceptual grouping affects pitch judgments across time and frequency. J Exp Psychol Hum Percept Perform 37:257–269. CrossRef Medline

Carlyon RP (1996) Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. J Acoust Soc Am 99:517–524. CrossRef Medline

Carlyon RP, Long CJ, Micheyl C (2012) Across-channel timing differences as a potential code for the frequency of pure tones. J Assoc Res Otolaryngol 13:159–171. CrossRef Medline

Cedolin L, Delgutte B (2010) Spatiotemporal representation of the pitch of harmonic complex tones in the auditory nerve. J Neurosci 30:12712–12724. CrossRef Medline

Chandrasekaran B, Kraus N, Wong PC (2012) Human inferior colliculus activity relates to individual differences in spoken language learning. J Neurophysiol 107:1325–1336. CrossRef Medline

Darwin CJ (2005) Pitch and auditory grouping. In: Pitch: neural coding and

perception (Plack CJ, Oxenham AJ, Fay R, Popper AN, eds), pp 278–305. New York: Springer.

de Boer J, Thornton AR (2008) Neural correlates of perceptual learning in the auditory brainstem: efferent activity predicts and reflects improvement at a speech-in-noise discrimination task. J Neurosci 28:4929–4937. CrossRef Medline

Flanagan JL, Guttman N (1960) On the pitch of periodic pulses. J Acoust Soc Am 32:1308. CrossRef

Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. Hear Res 47:103–138. CrossRef Medline

Gockel HE, Carlyon RP, Mehta A, Plack CJ (2011) The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. J Assoc Res Otolaryngol 12:767–782. CrossRef Medline

Goldstein JL (1973) An optimum processor theory for the central formation of the pitch of complex tones. J Acoust Soc Am 54:1496–1516. CrossRef Medline

Griffiths TD, Hall DA (2012) Mapping pitch representation in neural ensembles with fMRI. J Neurosci 32:13343–13347. CrossRef

Heinz MG, Colburn HS, Carney LH (2001) Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve. Neural Comput 13:2273–2316. CrossRef Medline

Helmholtz HLF (1885/1954) On the sensations of tone. New York: Dover.

Houtsma AJM, Smurzynski J (1990) Pitch identification and discrimination for complex tones with many harmonics. J Acoust Soc Am 87:304–310. CrossRef

Kaernbach C, Bering C (2001) Exploring the temporal mechanism involved in the pitch of unresolved harmonics. J Acoust Soc Am 110:1039–1048. CrossRef Medline

Krishnan A, Xu Y, Gandour J, Cariani P (2005) Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res Cogn Brain Res 25:161–168. CrossRef Medline

Licklider JCR (1954) "Periodicity" pitch and "place" pitch. J Acoust Soc Am 26:945.

Loeb GE, White MW, Merzenich MM (1983) Spatial cross correlation: a proposed mechanism for acoustic pitch perception. Biol Cybern 47:149–163. CrossRef Medline

McDermott HJ (2004) Music perception with cochlear implants: a review. Trends Amplif 8:49–82. CrossRef Medline

McDermott JH, Oxenham AJ (2008) Spectral completion of partially masked sounds. Proc Natl Acad Sci U S A 105:5939–5944. CrossRef Medline

McDermott JH, Lehr AJ, Oxenham AJ (2008) Is relative pitch specific to pitch? Psychol Sci 19:1263–1271. CrossRef Medline

Micheyl C, Keebler MV, Oxenham AJ (2010) Pitch perception for mixtures of spectrally overlapping harmonic complex tones. J Acoust Soc Am 128:257–269. CrossRef Medline

Moore BCJ (1973) Frequency difference limens for short-duration tones. J Acoust Soc Am 54:610–619. CrossRef Medline

Moore BCJ, Glasberg BR (1990) Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. J Acoust Soc Am 87:2163–2177. CrossRef Medline

Moore BCJ, Gockel HE (2011) Resolvability of components in complex tones and implications for theories of pitch perception. Hear Res 276:88–97. CrossRef Medline

Moore BCJ, Sêk A (2009) Sensitivity of the human auditory system to temporal fine structure at high frequencies. J Acoust Soc Am 125:3186–3193. CrossRef Medline

Ohm GS (1843) Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen [On the definition of tones, including a theory of sirens and similar tone-producing apparatuses]. Ann Phys Chem 59:513–565.

Oxenham AJ (2008) Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. Trends Amplif 12:316–331. CrossRef Medline

Oxenham AJ, Wojtczak M (2010) Frequency selectivity and masking. In:

Oxford handbook of auditory science: perception (Plack CJ, ed). Oxford: Oxford UP.

Oxenham AJ, Bernstein JG, Penagos H (2004) Correct tonotopic representation is necessary for complex pitch perception. Proc Natl Acad Sci U S A 101:1421–1425. CrossRef Medline

Oxenham AJ, Micheyl C, Keebler MV, Loper A, Santurette S (2011) Pitch perception beyond the traditional existence region of pitch. Proc Natl Acad Sci U S A 108:7629–7634

Palmer AR, Russell IJ (1986) Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. Hear Res 24:1–15. CrossRef Medline

Patel AD, Balaban E (2000) Temporal patterns of human cortical activity reflect tone sequence structure. Nature 404:80–84. CrossRef Medline

Plack CJ, Oxenham AJ, Fay R, Popper AN, eds (2005) Pitch: neural coding and perception. New York: Springer.

Pressnitzer D, Patterson RD, Krumbholz K (2001) The lower limit of melodic pitch. J Acoust Soc Am 109:2074–2084. CrossRef Medline

Qin MK, Oxenham AJ (2005) Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. Ear Hear 26:451–460. CrossRef Medline

Recio-Spinoso A, Temchin AN, van Dijk P, Fan YH, Ruggero MA (2005) Wiener-kernel analysis of responses to noise of chinchilla auditory-nerve fibers. J Neurophysiol 93:3615–3634. CrossRef Medline

Sayles M, Winter IM (2008) Reverberation challenges the temporal representation of the pitch of complex sounds. Neuron 58:789–801. CrossRef Medline

Schouten JF (1940) The residue and the mechanism of hearing. Proc Kon Akad Wetenschap 43:991–999.

Schouten JF, Ritsma RJ, Cardozo BL (1962) Pitch of the residue. J Acoust Soc Am 34:1418–1424. CrossRef

Schroeder MR (1968) Period histogram and product spectrum: new methods for fundamental-frequency measurement. J Acoust Soc Am 43:829–834. CrossRef Medline

Seebeck A (1841) Beobachtungen über einige Bedingungen der Entstehung von Tönen [Observations on some conditions for the formation of tones]. Ann Phys Chem 53:417–436.

Shackleton TM, Carlyon RP (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. J Acoust Soc Am 95:3529–3540. CrossRef Medline

Shamma S, Klein D (2000) The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. J Acoust Soc Am 107:2631–2644. CrossRef Medline

Shera CA, Guinan JJ Jr, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc Natl Acad Sci U S A 99:3318–3323. CrossRef Medline

Shera CA, Guinan JJ Jr, Oxenham AJ (2010) Otoacoustic estimation of cochlear tuning: validation in the chinchilla. J Assoc Res Otolaryngol 11:343–365. CrossRef Medline

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

Tzounopoulos T, Kraus N (2009) Learning to encode timing: mechanisms of plasticity in the auditory brainstem. Neuron 62:463–469. CrossRef Medline

Verschuure J, van Meeteren AA (1975) The effect of intensity on pitch. Acustica 32:33–44.

Walker KM, Bizley JK, King AJ, Schnupp JW (2011) Multiplexed and robust representations of sound features in auditory cortex. J Neurosci 31:14565–14576. CrossRef Medline

Wang X, Walker KM (2012) Neural mechanisms for the abstraction and use of pitch information in auditory cortex. J Neurosci 32:13339–13342. CrossRef

Warren JD, Uppenkamp S, Patterson RD, Griffiths TD (2003) Separating pitch chroma and pitch height in the human brain. Proc Natl Acad Sci U S A 100:10038–10042. CrossRef Medline

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