

Neurobiological Foundations for the Theory of Harmony in Western Tonal Music

MARK JUDE TRAMO,^{a,b,c} PETER A. CARIANI,^{a,b} BERTRAND DELGUTTE,^{b,c}
AND LOUIS D. BRAIDA^c

^a*Department of Neurology, Harvard Medical School and
Massachusetts General Hospital, Boston, Massachusetts 02114-2696, USA*

^b*Eaton-Peabody Laboratory of Auditory Physiology,
The Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114, USA*

^c*Research Laboratory of Electronics, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139-4307, USA*

ABSTRACT: Basic principles of the theory of harmony reflect physiological and anatomical properties of the auditory nervous system and related cognitive systems. This hypothesis is motivated by observations from several different disciplines, including ethnomusicology, developmental psychology, and animal behavior. Over the past several years, we and our colleagues have been investigating the vertical dimension of harmony from the perspective of neurobiology using physiological, psychoacoustic, and neurological methods. Properties of the auditory system that govern harmony perception include (1) the capacity of peripheral auditory neurons to encode temporal regularities in acoustic fine structure and (2) the differential tuning of many neurons throughout the auditory system to a narrow range of frequencies in the audible spectrum. Biologically determined limits on these properties constrain the range of notes used in music throughout the world and the way notes are combined to form intervals and chords in popular Western music. When a harmonic interval is played, neurons throughout the auditory system that are sensitive to one or more frequencies (partials) contained in the interval respond by firing action potentials. For consonant intervals, the fine timing of auditory nerve fiber responses contains strong representations of harmonically related pitches implied by the interval (e.g., Rameau's fundamental bass) in addition to the pitches of notes actually present in the interval. Moreover, all or most of the partials can be resolved by finely tuned neurons throughout the auditory system. By contrast, dissonant intervals evoke auditory nerve fiber activity that does not contain strong representations of constituent notes or related bass notes. Furthermore, many partials are too close together to be resolved. Consequently, they interfere with one another, cause coarse fluctuations in the firing of peripheral and central auditory neurons, and give rise to perception of roughness and dissonance. The effects of auditory cortex lesions on the perception of consonance, pitch, and roughness, combined with a critical reappraisal of published psychoacoustic data on the relationship between consonance and roughness, lead us to conclude that consonance is first and foremost a function of the pitch relationships among notes. Harmony in the vertical dimension is a positive phenomenon, not just a negative phenomenon that depends on the absence of

Address for correspondence: Dr. Mark Jude Tramo, MGH EDR-405, 55 Fruit Street, Boston, MA 02114-2696. Voice: 617-726-5409; fax: 617-726-5457.
mtramo@hms.harvard.edu

roughness—a view currently held by many psychologists, musicologists, and physiologists.

KEYWORDS: Consonance; Dissonance; Harmony, musical; Intervals, musical; Perception of harmony; Psychoacoustics of harmony

INTRODUCTION

Why do some combinations of simultaneous tones sound more harmonious than others? Pythagoras's curiosity about the nature of harmony inspired some of the earliest experiments relating mathematics and physics to perceptual phenomena.¹ Approaching the problem from the perspective of neurobiology, we ask: Are there physiological and anatomical properties of the auditory system and related cognitive systems that determine the degree to which simultaneous notes sound harmonious?

We restrict our consideration of harmony to basic tenets articulated by Piston,² among others. Harmony has a vertical dimension and a horizontal dimension. The vertical dimension encompasses the relationships among simultaneous notes. By convention, *note* refers to a pitch in the musical scale, and *harmonic interval* refers to two notes sounded simultaneously (FIG. 1A–D). When a note is played on a musical instrument, its pitch corresponds to the fundamental frequency (F0) of the complex tone generated by the instrument. Some synthesizers and other types of equipment are capable of generating pure tones, in which case the pitch of the note corresponds to the frequency of the pure tone. Harmonic intervals are a type of *dyad*. Three or more notes played simultaneously make up a *chord*. Chords with three notes are called *triads*. The time window over which acoustic information is integrated in the vertical dimension spans about a hundredth of a second to a few seconds (e.g., sixteenth notes to tied whole notes at a tempo of 120 beats per min). The horizontal dimension encompasses successive tones (melodic intervals and melodic progressions) and successive harmonic intervals and chords (harmonic progressions). Certain intervals and chords are treated as consonant (e.g., fifths, major triads) and others as dissonant (e.g., minor seconds, diminished triads). Acknowledging that different psychologists have attached different perceptual attributes and meanings to the terms *consonance* and *dissonance*, we nonetheless find considerable agreement among music texts and dictionaries that consonant means harmonious, agreeable, and stable, and that dissonant means disagreeable, unpleasant, and in need of resolution.^{2–5} Psychoacoustic experiments bear out semantic overlap among the terms *consonant*, *pleasant*, *beautiful*, and *euphonious*.^{6,7}

These basic concepts apply to a wide range of musical styles enjoyed by people throughout much of the industrialized world: contemporary pop and theater (including rock, rhythm and blues, country, and Latin-American), European music from the Baroque, Classical, and Romantic eras (1600–1900), children's songs, and many forms of ritualistic music (e.g., church songs, processionals, anthems, and holiday music). The overlap in their harmonic structure incorporates commonalities in musical phonology and syntax.⁸ In our view, the widespread popularity of Western pop taps into (1) universal competence in auditory functions needed to extract the pitch of a note and to analyze the harmonic relationships among different pitches, and (2) universal competence in cognitive functions that parse acoustic information and as-

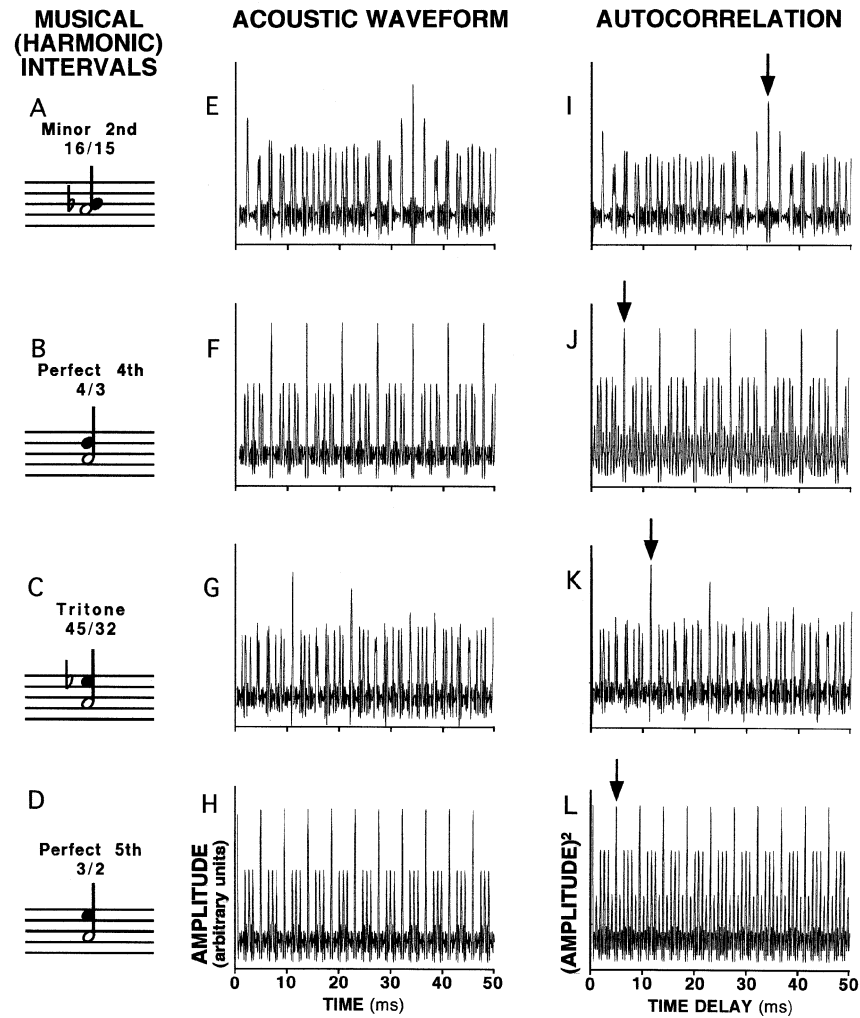


FIGURE 1. Acoustic representations of musical (harmonic) intervals in the time domain. *Left column (A–D):* Musical interval stimuli depicted in standard notation (G clef) below the name of each interval (e.g., minor second) and the F0 ratio of its notes (16/15). *Middle column (E–H):* Acoustic waveforms of each interval. *Right column (I–L):* Autocorrelations of the acoustic waveforms. Arrows indicate the peaks corresponding to the period of each interval’s “missing” F0.

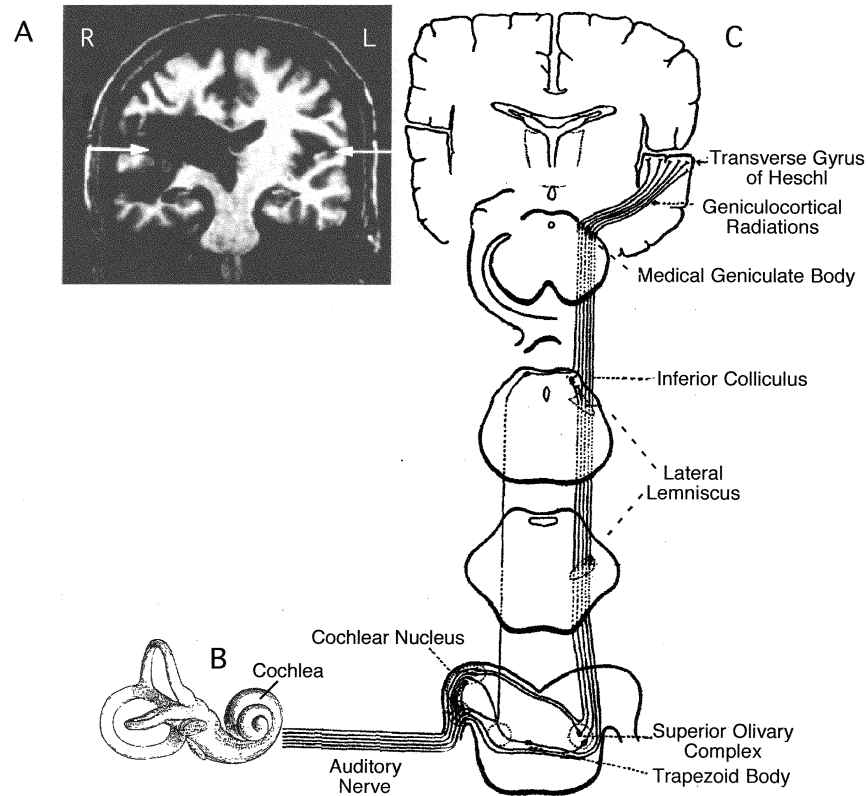


FIGURE 2. The auditory system. (A) Magnetic resonance image (coronal section) through the transverse gyri of Heschl and the superior temporal gyri of case MHS, who suffered bilateral infarction nine years earlier. *White arrows* point to the region of the transverse gyrus of Heschl in the right (R) and left hemispheres (L). The results of psychoacoustic experiments performed by MHS are shown in FIG. 7. (B) The inner ear and the auditory nerve, the obligatory pathway from the cochlea to the cochlear nuclei in the lower brain stem. (C) Schematic of the central auditory pathway showing the main relay stations and projection patterns. Adapted from Tramo *et al.*¹⁵ (A), Helmholtz¹⁶ (B), and Davis¹⁷ (C).

sociate perceptual attributes with emotion and meaning. Experimental results suggest that similar perceptual attributes can be associated with similar emotions and social contexts across different cultures.^{9,10} Moreover, listeners from different cultures often use similar cognitive schemata to structure the processing of pitch-sequences.^{11,12}

Much has been written in the psychology literature about the terms *harmony*, *consonance*, and *dissonance*. At present, many psychologists and musicologists subscribe to Terhardt's¹³ two-component model of *musical consonance*, which subsumes all these terms. One component is *sensory consonance*, the absence of an-

noying features, such as roughness, in both musical and nonmusical sounds. The other component, *harmony*, is based on music-specific principles that govern pitch relationships in melodic and harmonic progressions. Terhardt asserts that “sensory consonance...dominates the evaluation of single isolated chords...whereas harmony does not enter into the subject’s response” (p. 282).¹³

With respect to terminology, we adhere in this discussion to the simple distinction between the vertical and horizontal dimensions of harmony. We restrict use of the terms *consonance* and *dissonance* to the vertical dimension and keep the term *harmony* supraordinate to them. We make no assumptions about the level of auditory processing (e.g., sensory, peripheral) where the perceptual attribute of consonance takes shape. We consider it likely that a listener’s implicit (or explicit) knowledge about harmony in the horizontal dimension¹⁴ bears on harmony perception in the vertical dimension.

In this paper, we present neurophysiological, neurological, and psychoacoustic evidence to support our contentions that (1) pitch relationships among tones in the vertical dimension influence consonance perception and (2) consonance cannot be explained solely by the absence of roughness. First, we review terminology and basic psychoacoustics pertinent to our subsequent discussion of experimental results. Second, we demonstrate that the harmonic relationships of tones in musical intervals are represented in the temporal discharge patterns of auditory nerve fibers. Third, we critically reevaluate the psychoacoustic literature concerning the consonance of isolated intervals and chords, paying particular attention to (1) the relationships among interval width, roughness detection thresholds, and consonance ratings; and (2) the predictions of roughness-based computational models about relative consonance as a function of spectral energy distribution. Finally, we discuss evidence that impairments in consonance perception following auditory cortex lesions are more likely to result from deficits in pitch perception than to deficits in roughness perception. This evidence highlights the dependence of so-called low-level perceptual processing on the integrity of the auditory cortex, the highest station in the auditory nervous system (FIG. 2).

PSYCHOACOUSTICS AND NEUROPHYSIOLOGY OF HARMONY

For authoritative reviews of the psychoacoustics of harmony, we refer the reader to Krumhansl,¹⁴ Parncutt,¹⁸ and Deutsch.¹⁹ Here, only basic concepts and terminology pertinent to our subsequent discussion of psychoacoustic and neurophysiological experiments are covered.

Let us consider a modern restatement of Pythagoras’s observation: The degree to which two simultaneous notes (a harmonic interval) sound consonant is determined by the simplicity of the ratio $x:y$, where x is the F0 associated with one tone and y the F0 of the other, lower tone. In musical terms, y is the root of the interval. x and y can take on any value between about 25 Hz and about 5 kHz. This upper limit coincides with (1) the upper F0 of notes on a piccolo (~4500 Hz); (2) the upper F0 for which octave similarity can be reliably judged;^{20,21} and (3) the upper F0 of strong phase locking by auditory nerve fibers—that is, the highest frequency at which neurons can fire in time with amplitude fluctuations in the acoustic waveform.^{22,23} This convergence of facts from music, psychoacoustics, and physiology suggests that limitations in the phase-locking capacity of neurons in the auditory periphery constrain the

range of note F0s that are used in music and the way they are combined in the vertical dimension of harmony. Other authors have discussed the relationships among the temporal discharge patterns of auditory nerve fibers, fundamental pitch perception, octave equivalence, and the consonance of intervals formed by simple integer ratios.^{24–29}

By convention, notes in popular Western music are tuned to the scale of equal temperament, which chunks the F0 continuum into octaves (i.e., doublings of F0) and each octave into twelve discrete, equal, logarithmic steps. Each step within the octave is called a semitone, and the F0s of adjacent semitones differ by a factor of $2^{1/12}$, or about 6 percent. The chromatic scale is made up of all twelve tones in the octave, whereas the major and minor (diatonic) scales are made up of partially overlapping sets of seven tones in the octave. Harmonic intervals and melodic intervals are named according to the scale relationship of the upper note to the lower note. Thus when the fifth note in the major or minor scale sits atop the first note on the scale, the interval is called a *fifth* (Fig. 1D). The note named A_4 is assigned an F0 value of 440 Hz. The letter name of each note corresponds to one of the twelve notes in the octave; the number of each note indicates the octave the note is in, with increments at each occurrence of C along the scale (e.g., ... A_4 — $A\sharp_4$ — B_4 — C_5 ...). The frequency range over which pure tones are audible to humans extends from approximately E_0 (20.6 Hz) to E_{10} (21.1 kHz). For the purposes of this discussion, we will set y , the F0 of the root of a harmonic interval, equal to 440 Hz (A_4).

In music theory, the interval formed by notes that are an octave apart (e.g., A_4 and A_5) is the most consonant interval, followed by the fifth (A_4 and E_5), fourth (A_4 and D_5), major third (A_4 and $C\sharp_5$), and minor third (A_4 and C_5). In the scale of just intonation, these intervals correspond to $x:y$ ratios of 2:1, 3:2, 4:3, 5:4, and 6:5, respectively, consistent with Pythagoras's claim that the simplicity of the integer ratio correlates with perceived consonance. Combinations of some other notes on the scale between A_4 and A_5 have more complex $x:y$ ratios and sound dissonant. For example, the minor second and the tritone (also known as the augmented fourth, which, in equal-tempered tuning, is equivalent to the diminished fifth) have $x:y$ ratios of 16:15 and 45:32 (or approximately 7:5), respectively. The dependence of consonance on the simplicity of F0 ratios tolerates small deviations from perfect integer relationships. For example, because the scale of equal temperament is based on equal logarithmic steps within each octave, a major third in this scale has a ratio of 5.04:4, not 5:4. This deviation amounts to 0.8 percent. Even highly practiced listeners participating in psychoacoustic experiments under ideal listening conditions cannot reliably detect a mistuned lower harmonic embedded within a harmonic complex tone if the deviation is less than 0.9% (harmonics 1–12 at 60 dB SPL [sound pressure level] and isophase, $F_0 = 200$ Hz, duration ≤ 410 ms).³⁰ Moreover, conservatory students who excel at interval identification cannot reliably judge whether a mistuned major third composed of two harmonic complex tones has been stretched or compressed away from a perfect 5:4 ratio if the deviation is less than 1.2% (each tone with harmonics 1–20 in isophase, first harmonic at 80 dB SPL, higher harmonics at a 6 dB decrease per octave, F_0 between 260 Hz and 525 Hz, and duration = 1000 ms).³¹

All experimental studies that have used stimuli consisting of single, isolated, harmonic intervals formed by two complex tones (as would be the case if the intervals were sung or played on guitar or piano) show that listeners consistently perceive the fifth and fourth as more consonant than the minor second and tritone.^{6,32–35} This

convergence of results across study populations from different countries (USA, Germany, Japan), generations (1909–1969), and musical backgrounds, combined with results obtained in infants from the USA³⁶ and European starlings,³⁷ motivates the hypothesis that common, basic auditory mechanisms underlie perceptual categorization of harmonic intervals as consonant or dissonant. However, there is disagreement about the nature of the underlying neural mechanisms, and few physiological experiments have systematically analyzed responses to harmonic intervals at any level of the auditory nervous system.^{25,27,38–41} Still, a large body of data is available on the responses of neurons to other types of complex tones in the auditory nerve,⁴² cochlear nucleus,⁴³ inferior colliculus,^{44,45} medial geniculate nucleus,⁴⁶ and auditory cortex^{47–49} (FIG. 2; only a few of the many available papers are cited here; for review see Ehret and Romand⁵⁰).

NEURAL CODING OF PITCH RELATIONSHIPS AS A PHYSIOLOGICAL BASIS OF HARMONY

We synthesized simultaneous complex tones forming four musical intervals: the minor second (F_0 ratio = 16/15), perfect fourth (4:3), tritone (45:32), and perfect fifth (3:2, FIG. 1). Each of the two complex tones in the interval contained the first six harmonics with equal amplitude (60 dB SPL re: 20 μ Pa) and equal phase (cosine, FIG. 3). Each interval had a duration of 200 ms (a bit shorter than an eighth note at a tempo of 120 beats per minute), including 5-ms rise and fall times. These stimuli are acoustically similar to the inputs into the computational models used by Plomp and Levelt,⁵¹ Kameoka and Kuriyagawa,³⁵ and Hutchinson and Knopoff⁵² to predict the consonance of complex-tone intervals on the basis of psychoacoustic data on pure-tone intervals.

FIGURE 1 illustrates two time-domain representations of our stimuli: the acoustic waveform, which plots sound pressure amplitude as a function of time (FIG. 1E–H); and the autocorrelation of the waveform (FIG. 1I–L). In the acoustic waveform of the most consonant interval, the perfect fifth, we see a regular pattern of major and minor peaks (FIG. 1H). The pattern with one major peak and three minor peaks repeats every 4.55 ms ($1/x = F_0 = 220$ Hz). This periodicity corresponds to the missing F_0 of a harmonic series containing energy at the second harmonic (440 Hz, A_4) and third harmonic (660 Hz, E_5), the F_0 s of the notes actually present in the stimulus. Rameau's concept of the "basse fondamentale" (fundamental bass) in his *Treatise on Harmony*⁵³ is related to the missing F_0 of a harmonic interval.

Autocorrelation functions provide another representation of temporal regularities and irregularities embedded in acoustic waveforms (FIG. 1I–L). Autocorrelation functions are computed by multiplying the waveform with a delayed copy of itself and integrating over time. A large value at a given delay indicates the presence of a dominant periodicity in the waveform whose period equals the delay. Like pitch percepts, but unlike acoustic waveforms, autocorrelation functions are stable despite changes in the relative phases of frequency components. In the autocorrelation functions plotted in FIGURE 1I–L, the periodicity at the upper limit of the x axis (50 ms) corresponds to 20 Hz, the lowest audible frequency.

In the autocorrelation function of the perfect fifth (FIG. 1L), the first major peak again corresponds to the missing F_0 , A_3 (220 Hz). The second major peak occurs at

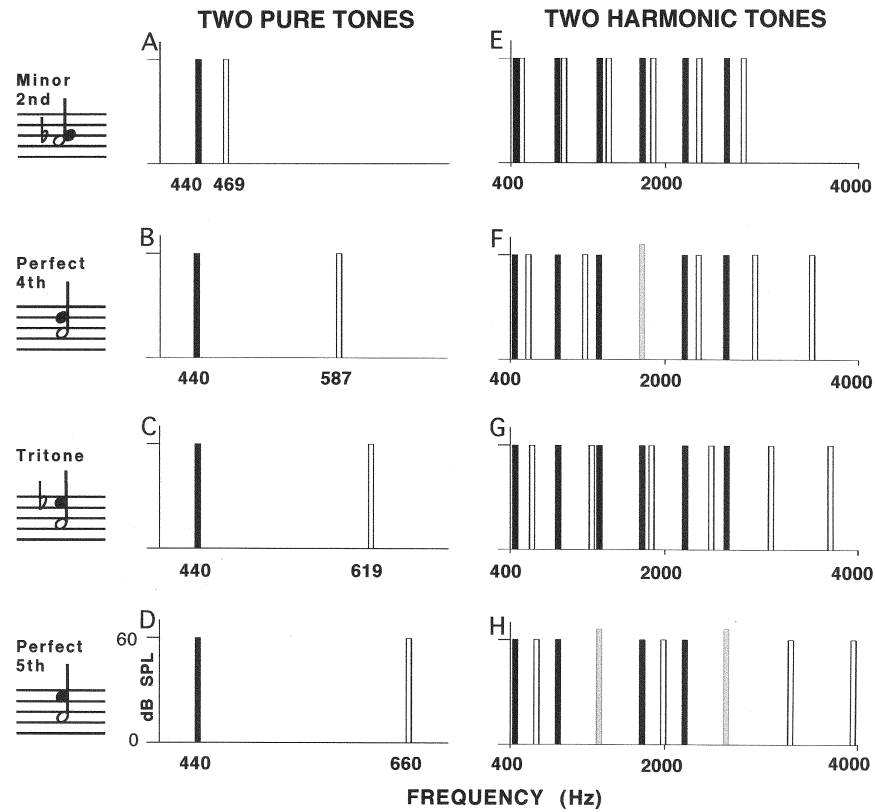


FIGURE 3. Line amplitude spectra of the four musical intervals shown in FIG. 1. *Black* indicates the root and its harmonics, *white* the interval and its harmonics, and *gray* the frequencies at which harmonics of the root and interval overlap. (A–D) A minor second, perfect fourth, tritone, and perfect fifth composed of two pure tones. (E–H) The same intervals composed of two harmonic complex tones containing the first six harmonics at equal amplitude. These are the spectral domain representations corresponding to the time domain representations in FIG. 1E–L.

9.09 ms, which corresponds to A_2 (110 Hz), the bass note an octave below. In fact, all the major peaks up to 50 ms correspond to the fundamental bass and its subharmonics (undertones) at A_2 , D_2 , A_1 , and on down to A_0 , the lowest note on the piano ($F_0 = 27.5$ Hz).

In between the major peaks is a set of three, evenly spaced, minor peaks. The first of these minor peaks occurs at 1.51 ms, which corresponds to E_5 (660 Hz), the upper note of the interval. The second minor peak occurs at 2.27 ms, which corresponds to A_4 (440 Hz), the root of the interval. The third minor peak occurs at 3.03 ms, which corresponds to E_4 (330 Hz), the octave below the fifth and the fifth above the fundamental bass at A_3 .

Temporal regularities are also seen in the waveform and autocorrelation of the perfect fourth, the other consonant interval in our stimulus set (FIG. 1F and J). Here the major peaks are at 6.82 ms (D_3 , $F_0 = 146.7$ Hz) and 13.6 ms (D_2 , $F_0 = 73.3$ Hz). Thus, in addition to a representation of the fourth, there is a representation of its inversion as a fifth with the implied root at D_3 and the fundamental bass at D_2 . The autocorrelation function of the fourth is a bit more complicated than that of the fifth, as there are two more peaks between each pair of major peaks. In the first set of minor peaks, the following notes are represented: A_4 (the root), D_4 (the interval), A_3 (the octave below the root), and G_3 (the fifth below D_4 , and the fourth of an interval rooted at D_3). Thus, we find representations of notes that function as fourths and fifths in the major and (all) minor scales of A and D.

In summary, the temporal fine structure of the perfect fifth and fourth contains representations of the two notes constituting the interval, plus harmonically related bass notes that are implied by the interval. In music, these bass notes support the deep structure of harmony. Parncutt¹⁸ demonstrated experimentally that listeners associate major triads with pitches that are harmonically related to note F_0 s, including the fundamental bass, plus the pitches of note F_0 s actually in the stimulus. These pitches cannot be accounted for simply on the basis of combination tones (for review, see Wightman and Green⁵⁴). Houtsma and Goldstein⁵⁵ showed that musicians can use missing F_0 pitches to identify melodic intervals (major and minor seconds and thirds), even when two upper harmonics are presented separately (dichotically) to each ear.

For the dissonant intervals in our stimulus set, the minor second and tritone, we find no such temporal regularity in the acoustic waveform and autocorrelation function. For the minor second (FIG. 1I), the largest peak in the autocorrelation function occurs at 34.1 ms, which corresponds to a frequency of 29.3 Hz, and it decays rapidly into the background. This periodicity lies outside the range associated with strong periodicity pitch percepts (for review, see Moore⁵⁶). In addition, we find multiple peaks between zero and the maximum peak. The largest of these is the first peak at 2.20 ms, which corresponds to mean of the note F_0 s, 455 Hz. This pitch does not correspond to any of the notes in any scale that has A_4 in it. In short, there is no strong representation of any pitch below the note pitches in the interval, and the dominant pitch is off the scale. Both of these factors contribute to the dissonance of the minor second.

Likewise, the autocorrelation function of the tritone (FIG. 1K) does not show a simple, regular pattern of peaks. The largest peak occurs at 11.4 ms, which corresponds to $F_0 = 88$ Hz, and it, too, decays into the background. This periodicity corresponds to a near coincidence between the fifth subharmonic of the root (440 Hz divided by 5) and the seventh subharmonic of the tritone (618.7 Hz divided by 7). It also lies close to F_2 , which is related to the fundamental bass of an F dominant-seventh chord in its first inversion. Thus the autocorrelation function of the tritone implies a chord that is, in music theory, less consonant than the chords implied by the autocorrelation functions of the fifth and fourth.

To summarize, for the consonant intervals (the fifth and fourth), the pattern of major and minor peaks in the autocorrelation is perfectly periodic, with a period related to the fundamental bass. This pattern is obtained because these stimuli have a unique, clearly defined fundamental period. By contrast, for dissonant intervals (the minor second and tritone), no true periodicity is seen in the autocorrelation function.

While some peaks occasionally stand out at specific delays, indicating a pseudoperiod, either there are no consistent peaks at multiples of this pseudoperiod, or the amplitudes of these peaks decay rapidly with increasing multiples of the pseudoperiod.

These observations suggest that the consonance of harmonic intervals reflects regularities in their temporal fine structure in the range of tenths to tens of milliseconds. Do neurons in the auditory system represent this information using a time code? Galileo,⁵⁷ who wrote about consonance while he was under house arrest for his work on the solar system, may have been the first to postulate that temporal coding in the auditory periphery was the physiological basis for consonance:

Agreeable consonances are pairs of tones which strike the ear with a certain regularity; this regularity consists in the fact that the pulses delivered by the two tones, in the same interval of time, shall be commensurable in number, so as not to keep the ear drum in perpetual torment, bending in two different directions in order to yield to the ever-discordant impulses....The unpleasant sensation produced by [dissonances] arises, I think, from the discordant vibrations of two different tones which strike the ear out of time. Especially harsh is the dissonance between notes whose frequencies are incommensurable;...this yields a dissonance similar to the augmented fourth or diminished fifth [*tritone o semidiapente*].” (Galileo,⁵⁷ 1638, pp. 103–104.)

To investigate the neural coding of consonance in the auditory periphery, we analyzed the responses of over 100 cat auditory nerve fibers to the minor second, perfect fourth, tritone, and perfect fifth. Auditory nerve fibers are the central axons of spiral ganglion cells that synapse on cochlear nucleus neurons in the brain stem (FIG. 2; for review, see Pickles⁵⁸). In humans, each auditory nerve contains about 30,000 auditory nerve fibers. Spiral ganglion cells also have peripheral axons that synapse on sensory receptors in the cochlea—the inner hair cells that ride atop the basilar membrane. Virtually all information about sound is transmitted from the ear to the brain via trains of action potentials fired by auditory nerve fibers.

When a minor second or some other interval is sounded, an auditory nerve fiber will increase the number of action potentials it fires only if it is sensitive to the frequencies present in the interval (FIG. 4A). The time between consecutive action potentials in the train is called an *interspike interval* (ISI), and a plot of the number of times each ISI occurs in the spike train is called an *ISI histogram* (FIG. 4B–E). We measured all the ISIs between all possible pairs of spikes (FIG. 4A, ISI₁, ISI₂...ISI_N) with a precision of approximately one microsecond. The corresponding plot is called an *all-order ISI histogram*, which is equivalent to the autocorrelation of the spike train. The spike train of each fiber in the auditory nerve can be analyzed in this way, and the resultant ISI histograms can be combined to show the ISI distribution in the entire population of auditory nerve fibers. Single-unit physiology experiments^{59–61} and computational models⁶² have shown that the first among the major peaks in the all-order ISI histogram computed from the entire auditory nerve fiber ensemble (the population ISI distribution) matches the fundamental period of complex tones and thus their periodicity pitch. This is essentially the time-domain equivalent of Terhardt’s⁶³ spectrally based subharmonic sieve for virtual pitch extraction.

FIGURE 4B–E illustrates the population ISI distributions embedded in the spike trains fired by over 50 auditory nerve fibers in response to the minor second, fourth, tritone, and fifth. In the response to the fifth (FIG. 3E), we see major peaks corresponding to the fundamental bass (A₃, 4.55 ms) and its subharmonics, just as we did in the acoustic waveform (FIG. 1H) and the autocorrelation of the waveform (FIG.

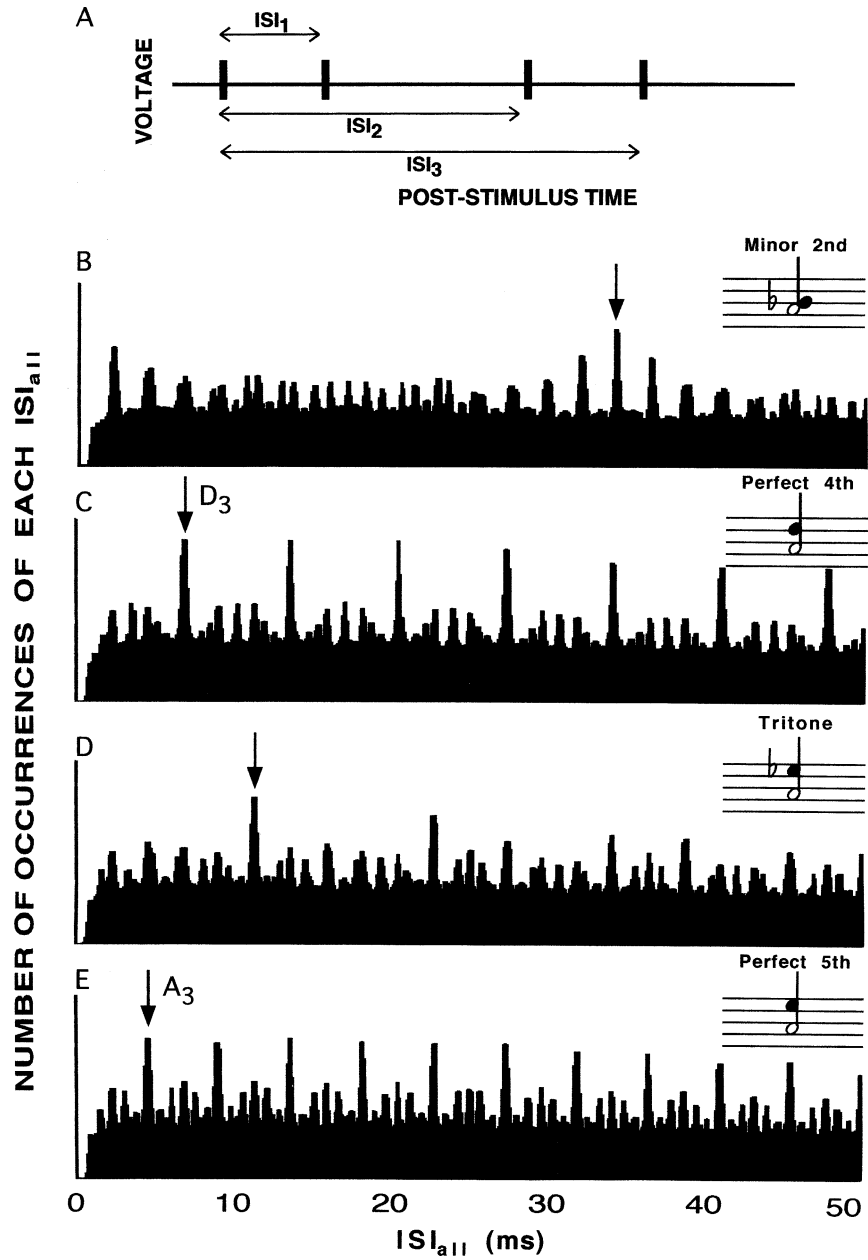


FIGURE 4. Interspike interval (ISI) distributions embedded in the responses of axons throughout the auditory nerve during stimulation with four musical intervals. (A) Schematic of a train of action potentials (or "spikes," vertical lines) fired by an auditory nerve fiber when a musical interval is played. Double arrows demarcate some of the ISIs between pairs

1L). Indeed, for all four intervals, the autocorrelation histograms of neural responses (FIG. 4B–E) mirror the fine structure of acoustic information in the time domain (FIG. 1).

Thus the peaks in the population ISI distribution evoked by consonant intervals reflect the pitches of each note, the fundamental bass, and other harmonically related pitches in the bass register. By contrast, the dissonant intervals (the minor second and tritone) are associated with population ISI distributions that are irregular. These contain little or no representation of pitches corresponding to notes in the interval, the fundamental bass, and related bass notes.

To obtain a physiological measure of the strength of the fundamental pitch of each interval relative to other pitches, we measured the number of intervals under the peak in the all-order ISI distribution corresponding to the missing F0 (arrows in FIG. 3B–E; bin width = 300 μ s). We then divided that value by the value of y in each x bin from $x = 0$ ms – 50 milliseconds. We found a high correlation ($r = 0.96$) between our physiological measure of fundamental pitch strength and previous psychoacoustic measures of the “clearness” of musical intervals composed of two complex tones.³⁵

In summary, the all-order ISI distribution embedded in auditory nerve fiber firing patterns contains representations of the pitch relationships among note F0s that influence the perception of musical intervals as consonant or dissonant. The neural coding mechanisms that provide representations of these pitch relationships form part of the neurobiological foundation for the theory of harmony in its vertical dimension.

NEURAL CODING OF ROUGHNESS AS THE PHYSIOLOGICAL BASIS FOR HARMONY PERCEPTION

Whereas pitch-based accounts treat consonance as a positive perceptual phenomenon associated with the presence of highly structured temporal information, roughness-based accounts treat consonance as a negative phenomenon associated with the absence of annoying perceptual attributes. Terhardt’s notion that the consonance of isolated intervals and chords depends on the absence of roughness¹³ echoes one of the main points in Helmholtz’s^{16,64} monumental work, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*:

As long as several simple tones of a sufficiently different pitch enter the ear together, the sensation due to each remains undisturbed in the ear, probably because entirely different bundles of [auditory] nerve fibers are affected. But tones of the same, or of nearly the same pitch, which therefore affect the same nerve fibers, do not produce a sensation which is the sum of the two they would have separately excited, but new and peculiar phenomena arise which we term interference...and beats¹⁶ (p. 160)...Rapidly beating tones are jarring and rough...the sensible impression is also unpleasant¹⁶

of spikes in the spike train. ISI_1 refers to the first-order ISI, ISI_2 and ISI_3 to higher-order ISIs. All possible ISIs are in “all-order” ISI histograms. (B–E) All-order ISI histograms showing ISIs embedded in the responses of 50 auditory nerve fibers to musical intervals composed of two complex tones (FIG. 1E–H and FIG. 3E–H). Arrows mark peaks in the ISI pattern that correspond to the missing F0 of the interval. If the peak corresponds to a note on the scale, the name of the note is given.

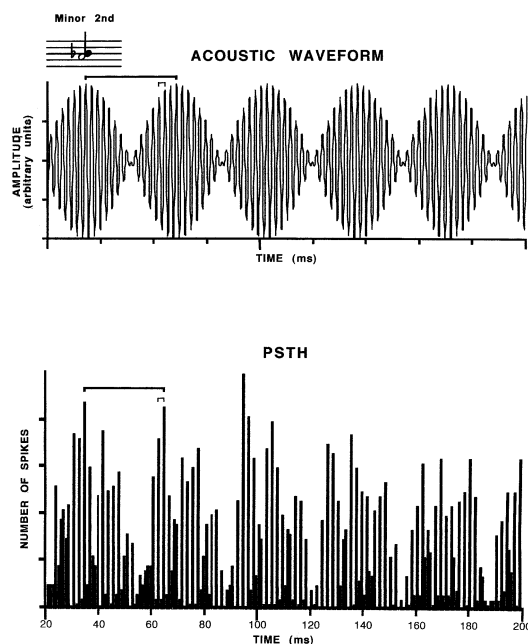


FIGURE 5. (Top) Acoustic waveform of a minor second composed of two pure tones with the root at A₄. Thick bars show the period of envelope fluctuations that render the minor second rough ($P = 1/\Delta F = 34.1$ ms). Thin bars show the period of fluctuations under the envelope that corresponds to the mean frequency of the tones and the pitch of the interval ($P = 2.20$ ms). (Bottom) Poststimulus time histogram (PSTH) showing the number of spikes fired by a single auditory nerve fiber during the steady state portion of its response to the minor second. Note that the global and local fluctuations in firing rate mirror those seen in the acoustic waveform of the minor second. This fiber was sensitive to frequencies at both the root and the interval at 60 dB SPL. Bin width = 1 ms. Number of stimulus repetitions = 100.

(p. 168). Consonance is a continuous, dissonance an intermittent tone sensation. The nature of dissonance is simply based on very fast beats. These are rough and annoying to the auditory nerve.⁶⁴ (Helmholtz 1863,⁶⁴ 1885.¹⁶)

Because frequency selectivity throughout the auditory nervous system is finite, simultaneous pure tones that are separated by small frequency differences (ΔF s), such as a minor second (FIG. 3A), cannot be separated or “filtered out” from one another. Consequently, their waveforms are effectively summed, and the pitch of the tone combination matches their mean frequency.⁶⁵ The envelope of the summed waveform contains periodic amplitude fluctuations whose frequency equals ΔF (FIG. 5, top). If these envelope fluctuations fall in the range of 20–200 Hz (the precise values depend on the frequencies of the two tones, FIG. 6A), interruptions in continuous tone sensation are perceptible. These interruptions make the tone combination sound “rough,” analogous to the interruptions one feels on the fingertips when touching coarse sandpaper. At smaller frequency differences, and thus slower amplitude modulations, one perceives a single, continuous tone that is slowly fluctuating in loudness, or “beating.” Auditory nerve fibers^{38,39} (FIG. 5, bottom), inferior colliculus neurons,^{40,41} and

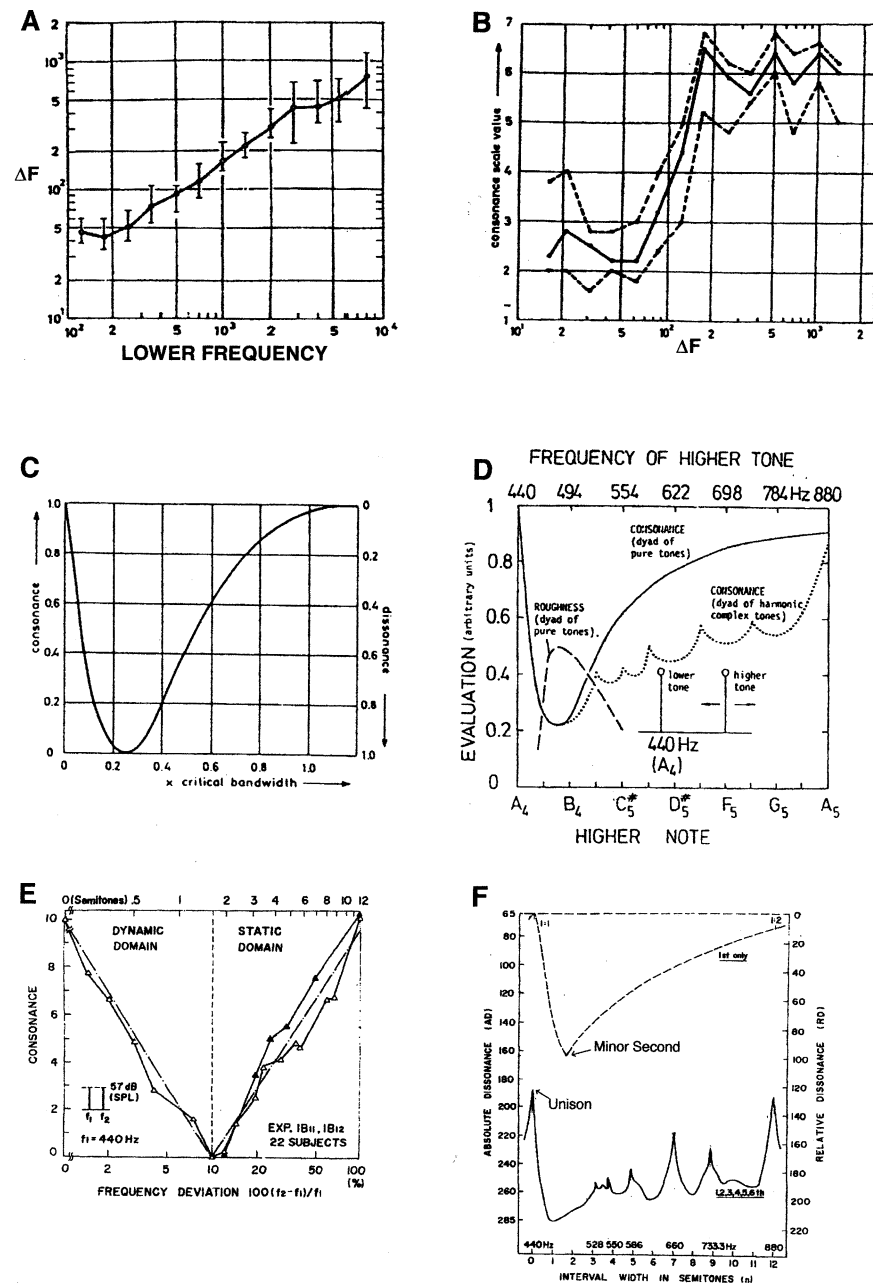


FIGURE 6. See following page for caption.

populations of primary auditory cortex neurons⁴⁸ can fire in synchrony with amplitude fluctuations in the ΔF range associated with perception of roughness and beats.

The concept of *critical bandwidth* refers to the limits of ΔF over which frequency selectivity operates in the auditory system. Critical bandwidth has been estimated psychoacoustically in several ways that have yielded somewhat different results depending on the method (for reviews, see Greenwood⁶⁶ and Moore⁵⁶). One estimate of critical bandwidth is based on the ΔF above which roughness disappears.

When musical intervals are composed of two complex tones (FIG. 3E–H), the partials may interfere with one another and produce amplitude fluctuations at the corresponding ΔF . There is more interference between adjacent partials in the minor second and tritone (FIG. 3E and G) than in the fourth and fifth (FIG. 4F and H). Several computational models of consonance^{16,35,51,52} assume that (1) the roughness generated by all the partials in the interval are added together (presumably by a central processor in the auditory brain stem or cortex), and (2) this total roughness determines the degree to which the interval is perceived as consonant.

FIGURE 6A shows Plomp and Steeneken's data on the relationships among the ΔF between two pure tones, the frequency of the lower tone (or root), and just-noticeable roughness (line).⁶⁷ The data at $x = 500$ Hz would thus apply to the case of a harmonic interval composed of two pure tones with the root at B_4 ($x = 494$ Hz). The fifth of B_4 , $F\sharp_5$, has a frequency of 741 Hz, so the ΔF between the root and the fifth is 247 Hz. According to FIGURE 6A, at $x = 500$ Hz, roughness disappears for ΔF values above 90 Hz–125 Hz. Therefore, the fifth should not be associated with roughness. Experimental studies agree that an isolated fifth composed of pure tones in this frequency register sounds “consonant”^{7,51} or “pleasant.”^{6,33} The pure-tone fifth is thus

FIGURE 6. (A) Just-noticeable roughness (*line*) as a function of the frequency difference between two pure tones (ΔF) and the lower frequency of the tones. Tones were presented monaurally at 60 phons. The *bars* show the interquartile range of 20 subjects (musical background not given). (Adapted from Plomp and Steeneken,⁶⁷ FIG. 1, p. 883.) (B) Plomp and Levelt's data on consonance ratings as a function of ΔF for simultaneous pure tones with a mean frequency of 1000 Hz. Tones were presented in free field at 65 dB SPL. The *solid line* shows the mean consonance ratings of 10 subjects (musical background not given). *Dashed lines* show the interquartile range. (Adapted from Plomp and Levelt,⁵¹ FIG. 6, p. 554.) (C) Plomp and Levelt's idealized plot of the relationship between consonance and critical bandwidth. The y axis is in units of consonance (*left*) and dissonance (*right*). (From Plomp and Levelt⁵¹ [FIG. 10, p. 556]. Reproduced by permission.) (D) Terhardt's idealized plot showing the relationship of consonance and roughness to interval width (root at A_4 , 440 Hz). The *solid line* shows consonance vs. ΔF when the interval is composed of two pure tones. The *dotted line* shows consonance vs. ΔF_0 when the interval is composed of two complex tones, each containing several lower harmonics. The *dashed line* shows the roughness of two pure tones as a function of ΔF . (From Terhardt¹³ [FIG. 1, p. 281]. Reproduced by permission.) (E) Kameoka and Kuriyagawa's data on consonance as a function of ΔF for simultaneous pure tones with the root at A_4 and intensities of 57 dB SPL. The *dashed line* is the mean performance of 22 audio engineers who performed the task twice (*solid lines with black and white triangles*). (From Kameoka and Kuriyagawa⁷² [FIG. 1, p. 1452]. Reproduced by permission.) (F) Kameoka and Kuriyagawa's plot of dissonance as a function of interval width. The *dashed line at the top* shows the function when the interval is composed of two pure tones. The *solid line at the bottom* gives the calculated dissonance of an interval composed of two complex tones, each containing the first six harmonics at 57 dB SPL. (Adapted from Kameoka and Kuriyagawa,³⁵ FIG. 8, p. 1465.)

associated with the absence of roughness and with strong pitches associated with temporal regularities in its acoustic waveform and autocorrelation (similar to those described in the preceding section for a fifth composed of two harmonic complex tones; FIG. 1H and L). The same set of observations applies to a perfect fourth composed of two pure tones.

Now consider the case of a minor second composed of two pure tones with the root at B₄ (494 Hz). Here, ΔF (between B₄ and C₅) is 33 Hz. This falls well within the range of noticeable roughness (FIG. 6A). In fact, it lies near the ΔF associated with maximal roughness^{67–69} (not shown). Experimental studies agree that a pure-tone minor second (and other tone combinations close to it) sounds “dissonant”^{7,51} or “unpleasant.”^{6,33}

The case of a tritone composed of two pure tones provides an interesting test of the roughness hypothesis. The ΔF between a tritone at F₅ and a root at B₄ is 201 Hz, well above the ΔF for just-noticeable roughness. Thus the fourth, tritone, and fifth are all above the roughness range. Does that mean they all have the same consonance?

FIGURE 6B shows Plomp and Levelt’s data on consonance ratings as a function of ΔF for two pure tones whose mean frequency is 1000 Hz.⁵¹ Because the y axis is an ordinal scale, not an interval or ratio scale, it is inappropriate to assume that equal distances reflect equal differences in consonance. It follows that $4 > y > 4$ is not to be taken as the categorical boundary for dissonance and consonance, respectively. In addition, because Plomp and Levelt intentionally avoided using standard intervals like the fourth and tritone (they were concerned that interval recognition would influence consonance ratings), it is difficult to estimate where on the curve these intervals would fall. These caveats aside, it is clear that all pure-tone combinations with ΔF s above approximately 150 Hz are consonant. This would apply to the fourth, tritone, and fifth with their roots in the vicinity of A₅. Tone combinations with $\Delta F_0 = 20$ to 80 Hz are dissonant; this would apply to a minor second with the root near A₅. Superficially, it would appear that we have a convergence between the disappearance of roughness at $\Delta F_0 = 150$ –250 Hz (FIG. 6A, interquartile range for a lower frequency of 1 kHz) and a steep increase in consonance ratings at $\Delta F_0 > 80$ Hz (at and above a minor third, FIG. 6B). However, the ΔF associated with the highest mean consonance rating ($\Delta F \sim 180$ Hz) is within the range of noticeable roughness for many of Plomp and Steeneken’s subjects.⁶⁷

Beyond about $\Delta F = 180$ Hz, mean consonance ratings vary by only one rating point or less, but they are not perfectly flat (FIG. 6B). One can discern alternating peaks and valleys out to about $\Delta F = 1200$ Hz. We estimated where the minor second, perfect fourth, tritone, and perfect fifth might fall on the interpolated lines drawn by Plomp and Levelt⁵¹ (FIG. 6B), and we estimated the frequency ratios and intervals that correspond to the peaks and valleys beyond $\Delta F = 1000$ Hz. The first peak is near 6:5, which would correspond to a minor third. The second peak falls close to 5:3, which would correspond to a major sixth (or inverted minor third). The ratio 5:3 could also be thought of as the fifth and third harmonics of a harmonic series corresponding to the third and first notes of a major triad in its second inversion. The third peak is at or close to a ratio of 3:1, which corresponds to the interval of a twelfth, that is, a root and a fifth in the octave above. High consonance ratings for the twelfth are also found in Plomp and Levelt’s data for two pure tones with a mean frequency of 500 Hz⁵¹ (not shown). The first valley between the first two peaks is near the tritone, and the second valley appears to be a mistuned octave, with a frequency ratio near 2.03:1.

FIGURE 6C shows Plomp and Levelt's idealized plot of the relationship between consonance and critical bandwidth⁵¹ (the latter is defined here by loudness summation⁷⁰). Note that the curve reaches an asymptote near the end of the x axis, at about one critical bandwidth. Thus a critical band account of consonance as the absence of roughness *cannot* apply to pure-tone intervals that are wider than a minor third or so.

Yet Terhardt's¹³ idealized plot (FIG. 6D) of the relationships among roughness, consonance, and pure-tone ΔF shows a monotonic increase in perceived consonance all the way out to the octave, well beyond the ΔF at which (1) roughness disappears (FIG. 6A and D), (2) consonance ratings plateau (FIG. 6B), and (3) loudness summation and masking effects are observed (FIG. 6C; for review see Yost⁷¹). The representation of the psychoacoustic literature summarized in FIGURE 6D appears to draw upon Kameoka and Kuriyagawa's data showing increases in consonance well beyond the ΔF associated with disappearance of roughness⁷² (FIG. 6E). Although it is generally accepted that Kameoka and Kuriyagawa's work supports the idea that consonance is a function of roughness and critical bandwidth, comparisons of FIGURE 6A, C, and E reveal that their data actually argue against it, at least for musical interval widths greater than the ΔF for just-noticeable roughness.

The disagreement may arise from two sources. First, Kameoka and Kuriyagawa's Japanese audio engineers were instructed to judge tones for *sunda* (which they translate as "clearness" in English) and *nigotta* ("turbidity").⁷² Consequently, these listeners may have been rating different perceptual attributes than Plomp and Levelt's Dutch subjects, who were instructed to judge how "consonant" [or *mooi* ("beautiful") or *welluidend* ("euphonious")] the intervals sounded.⁵¹ Second, Kameoka and Kuriyagawa used an incomplete paired comparison paradigm—incomplete because only three or four adjacent intervals were paired for comparisons of relative consonance,⁷² a much more restricted ΔF range than the one Plomp and colleagues used in their one-interval consonance rating paradigm.^{7,51} When Kameoka and Kuriyagawa tried the method of magnitude estimation, presumably using all possible pairings, the task turned out to be "rather difficult" for "naive subjects"⁷² (p. 1453), and they dropped it in favor of incomplete pairings. Comparing only adjacent intervals may have biased subjects to focus on differences they would not have otherwise attended to if all intervals had been paired with one another. The pattern of results suggests that pitch height, rather than absence of roughness, influenced consonance judgments beyond a minor third or so. Since the authors used these data to calculate the consonance of musical intervals formed by two complex tones (FIG. 6F), they may have confounded roughness and pitch height in their predictions.

We reviewed previous studies that used isolated minor seconds, fourths, tritones, and fifths composed of two pure (or nearly pure) tones as experimental stimuli. Kaestner,⁶ who used a *Stimmgabelklängen* to generate tones that were "poor in overtones," found that subjects judged the fourth to be slightly more "pleasant" than the tritone. Malmberg,³² who used tuning forks, found a more marked preference for the fourth over the tritone for judgments of "blending," "purity," and "smoothness." Pratt,⁷³ who used a Stern variator that may have produced weak overtones, found that the fourth was judged to be more "pleasant," "smoother," and more "unitary" than the tritone. Brues,⁷⁴ using a Stern variator that produced weak energy at the first overtone, found the fourth, tritone, and fifth were similar with respect to "unitariness." Guthrie and Morrill⁷⁵ used a Stern variator that produced "very faint traces of

the third partial” and reported that the fourth was judged to be more “pleasant” than the tritone and of equal “consonance.” Guernsey,³³ who used tuning forks, reported that nonmusicians, amateur musicians, and professional musicians found the fourth more “pleasant” and “smooth” than the tritone. Schellenberg and Trehub’s recent experiments with nine-month-old babies are also relevant here.⁷⁶ When the upper pure tone of a repeating harmonic interval was flattened by one-fourth of a semitone, infants could detect the change if the interval was a fourth but not if it was a tritone. Their findings indicate that fourths provided a more stable background against which changes in tuning could be detected. All in all, these results indicate that the fourth, even when it is composed of two pure tones, is often perceived as more consonant than the tritone.

Another challenge for roughness-based accounts of consonance arises when we compare the consonance of pure-tone intervals and the consonance of complex-tone intervals. In FIGURE 6D, Terhardt¹³ plots consonance ratings for pure-tone and complex-tone intervals on the same scale. In fact, Kameoka and Kuriyagawa’s psychoacoustic data and calculations put them on different scales³⁵ (FIG. 6F). Likewise, Plomp and Levelt use a dissonance scale from one to zero for their pure-tone data (FIG. 6C) and a dissonance scale from six to zero for their complex-tone calculations (not shown).⁵¹ Kameoka and Kuriyagawa’s calculations predict that a minor second composed of two pure tones is more consonant than the unison of two complex tones with the first six harmonics at isoamplitude³⁵ (FIG. 6F). Intuitively, this notion is untenable; however, direct comparisons between pure-tone intervals and complex-tone intervals have not been reported in the literature. We synthesized a pure-tone minor second and the unison of two complex tones (with the acoustic parameters specified by Kameoka and Kuriyagawa³⁵) and asked several of our students to judge which of these two stimuli sounded more “consonant.” These and Huron’s (personal communication) informal observations raise the possibility that a combination of pitch height effects and loudness (shrillness), rather than or in addition to roughness, accounts for Kameoka and Kuriyagawa’s predictions. For example, in the case of unison at A₄ (FIG. 6F), spectral energy extends all the way up to 2200 Hz (fifth harmonic) and 2640 Hz (sixth harmonic), so there are high-frequency components that are greater in sensation level than the note F₀s. At the same time, ΔF (440 Hz) is higher than the highest ΔF associated with just-noticeable roughness when the root is at 2000 Hz (FIG. 6A, interquartile range for $\Delta F \sim 250\text{--}400$ Hz⁶⁷).

In summary, the neural coding mechanisms that provide representations of roughness form part of the neurobiological foundation for the theory of harmony in its vertical dimension. However, our reappraisal of the psychoacoustic literature leads us to conclude that the dependence of consonance on the absence of roughness is overstated. We believe pitch relationships, as well as roughness, influence the perception of intervals and chords as consonant or dissonant in the vertical dimension.

EFFECTS OF AUDITORY CORTEX LESIONS

Another approach to assessing the relative contributions of pitch and roughness to consonance perception might be to determine whether impairments in consonance perception caused by brain lesions are associated with deficits in one, the other, or both.

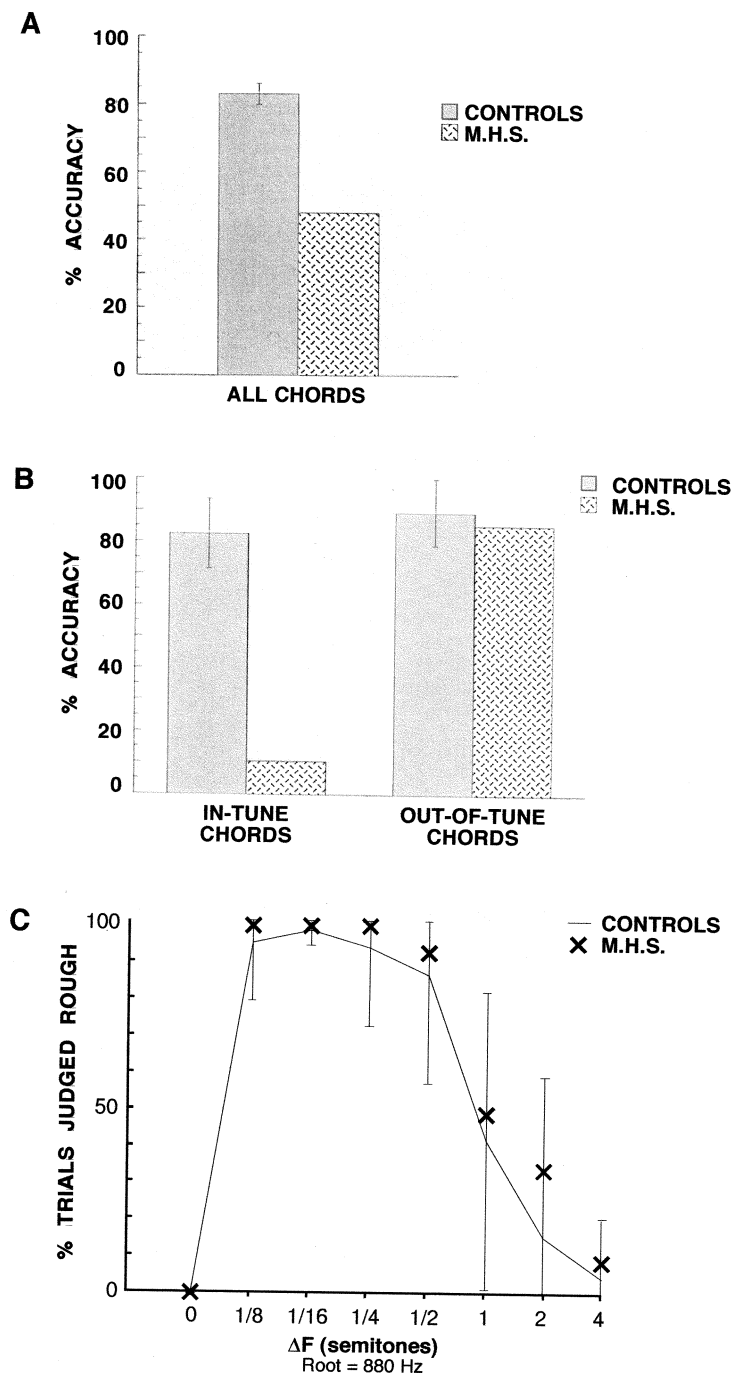


FIGURE 7. See following page for caption.

Consonance perception has been reported to be severely impaired following bilateral lesions of the auditory cortex.¹⁵ In an experiment employing a one-interval, two-alternative, forced-choice paradigm, two types of stimuli were presented: a major triad, and a triad whose fifth was flattened by a fraction of a semitone. In each trial of the experiment, a young stroke patient, MHS, was asked if a single, isolated chord sounded “in tune” or “out of tune.” His response accuracy was 56%, better than chance ($p < 0.05$), but more than two standard deviations below the mean of thirteen normal controls (FIG. 7A). Magnetic resonance imaging revealed that MHS’s infarcts involved the primary auditory cortex in both hemispheres, all or almost all of the auditory association cortex in the right hemisphere, and about 20% of the posterior auditory association cortex in the left hemisphere (FIG. 2A). His pure-tone audiograms were within normal limits. Speech perception was impaired.

We subsequently compared MHS’s performance on in-tune trials versus out-of-tune trials. If roughness perception were impaired, then MHS might make more errors in in-tune trials than out-of-tune trials. If frequency selectivity were coarsened and pitch perception impaired, then MHS might make more errors on in-tune trials. Consistent with the latter possibility, we found a marked response bias for out-of-tune judgments (FIG. 7B). Two possible interpretations follow: (1) MHS was having difficulty extracting the pitches of chord frequency components and analyzing their harmonic relationships; and/or (2) he heard more roughness in the chords than normals because his effective critical bandwidths were wider.

To assess whether MHS was having difficulty with frequency discrimination, we examined his performance on the Pitch Discrimination subtest of the Seashore Measures of Musical Talents.^{77,78} This test uses the method of constant stimuli to measure one’s ability to judge whether the second of two pure tones is higher or lower in pitch than the first tone. The ΔF between the tones gets smaller over successive blocks of trials. The tones were centered at 500 Hz, 600 ms in duration, and 600 ms apart, with an intensity of 35–40 dB above sensation level. Overall, MHS scored in the 15th percentile. His error pattern was again revealing. He performed poorly in the last third of the test, where the ΔF s between the tones were smallest. We subsequently measured pure-tone frequency difference thresholds for pitch discrimination using an adaptive procedure and a two-interval, two-alternative, forced-choice paradigm.⁷⁹ Whereas normal controls and patient controls had frequency difference thresholds corresponding to Weber fractions ($\Delta F/\text{mean frequency}$) of around 1%, MHS’s Weber fractions were over 10 percent. In short, his ability to judge the direction of a pitch change was markedly impaired. A similar deficit has since been reported in patients with surgical lesions of the right primary auditory cortex and right

FIGURE 7. Perception of consonance (**A** and **B**) and roughness (**C**) in a patient, MHS, with bilateral lesions of auditory cortex (FIG. 2A). (**A**) Data from Tramo *et al.*¹⁵ showing response accuracy on a task that required MHS to determine whether a major triad was “in tune” or “out of tune.” (**B**) MHS’s error rate as a function of stimulus condition. (**C**) Percentage of trials in which two simultaneous pure tones were judged to be fluctuating and rough (as opposed to steady and smooth) as a function of the ΔF between the tones. The mean (*line*) and standard deviation (*bars*) show the combined performance of 10 normal subjects and 8 neurological patients who did not have bilateral lesions of auditory cortex. Three of the patients had chronic unilateral lesions of superior temporal cortex (left or right), and one had a lesion of the right inferior colliculus.

anterior auditory association cortex.⁸⁰ We also found that perception of the missing F0 of harmonic complex tones was impaired,⁸¹ again consistent with observations in most right temporal lobectomy patients with partial or complete excisions involving the primary auditory cortex.⁸² Taken together, these findings are consistent with the hypothesis that MHS's impaired consonance perception was related to deficits in pitch processing.

Was this increase in ΔF thresholds for pure-tone pitch discrimination associated with higher ΔF values for disappearance of roughness? We assessed MHS's ability to judge two simultaneous pure tones as "steady and smooth" versus "fluctuating and rough" using a one-interval, two-alternative, forced-choice paradigm and the method of constant stimuli. The lower tone was fixed at either 220 Hz (A_3) or 880 Hz (A_5), and the upper tone was above the root by a variable number of semitones: 0, 1/16, 1/8, 1/4, 1/2, 1 (a minor second), 2 (a major second), or 4 (a major third). FIGURE 7C shows the results when the root was at 880 Hz. When the tones were between 1/16 to 1/2 semitone apart, MHS, like controls, judged the combination as rough on >80% of trials. When the tones were zero, two, or four semitones apart, MHS and controls judged the combination to be rough on less than 20% of trials. At one and two semitones apart, MHS's performance fell near the mean of controls, but there is too much variability in the normal data to meaningfully assess MHS's performance. Still, these observations mitigate the possibility that consonance perception was impaired because he heard more roughness in chords than normals.

In summary, MHS's bias to hear major triads as mistuned appears to be associated with impairments in pitch perception but not roughness perception. Consistent with our physiological data and review of the psychoacoustic literature, this pattern of lesion effects indicates that pitch relationships influence harmony perception in the vertical dimension.

CONCLUSIONS

Basic physiological and anatomical properties of auditory and cognitive systems determine why some combinations of simultaneous tones sound more harmonious than others. Distinctive acoustic features of consonant and dissonant intervals are translated into distinctive patterns of neural activity. A faithful representation of temporal regularities in the acoustic structure of consonant intervals exists in the population interspike interval (ISI) distribution of auditory nerve fibers. The most common ISIs in the distribution correspond not only to the pitches of note F0s actually present in the consonant intervals, but also to the pitches of harmonically related notes in the bass register, such as the fundamental bass. By contrast, for dissonant intervals, the most common ISIs in the distribution do not correspond to one of the note F0s, nor do they correspond to harmonically related notes. The relative strength of the missing F0 in the population ISI distribution predicts the relative consonance of the minor second, perfect fourth, tritone, and fifth. Limits on the temporal precision and frequency selectivity of neurons throughout the auditory system constrain the range of note F0s we can hear as strong pitches and how they are combined into intervals and chords. Implicit knowledge about the hierarchical relationships of pitches in a given tonal system is likely to exert cognitive influences on the degree

to which intervals and chords sound consonant or dissonant, even when they are heard in isolation.

Representations of roughness exist in temporal patterns of neural activity at several levels of the auditory system. For the minor second, fourth, tritone, and fifth, the amount of 20- to 200-Hz temporal fluctuations in the firing patterns of auditory nerve fibers inversely correlates with perceived consonance. These representations of roughness are multiplexed with pitch representations in the spike trains of auditory nerve fibers. These two neural time codes operate over different time regimes. The fine timing of action potential firing with precision in the submillisecond range carries information about fundamental pitch. Periodic fluctuations in discharge rate with precision in the range of milliseconds to tens of milliseconds carry information about roughness.

Bilateral lesions of primary auditory cortex and auditory association cortex can lead to severe impairments in consonance perception, with a bias to judge well-tuned chords as out of tune. In our patient, MHS, impaired consonance perception was associated with severely impaired pitch perception, but roughness perception appeared to be normal or near normal.

We interpret our findings and the results of previous psychoacoustic experiments as evidence in favor of the hypothesis that harmony in the vertical dimension, like harmony in the horizontal dimension, is principally a function of the pitch relationships among tones, with roughness playing a secondary role. In light of these observations, and in view of the likelihood that cognitive representations of pitch hierarchies influence harmony perception in the vertical dimension, we urge that the terms *sensory consonance* and *sensory dissonance* be reconsidered.

ACKNOWLEDGMENTS

This work was supported by NIH DC03382 (M.J. Tramo), DC03054 (P.A. Cariani), DC02258 (B. Delgutte), and DC00117 (L.D. Braida); and the McDonnell-Pew Program in Cognitive Neuroscience (M.J. Tramo). We gratefully acknowledge the instructive comments of David Cohen and Kay Shelemay of the Harvard Music Department; the technical support of Leslie Liberman of EPL, and David Lum, Andrew Grant, and Danielle Lemay of RLE; the graphics and manuscript assistance of Adda Kridler, Janelle Mallett, and Monique James; and, especially, our collaborators in the experimental studies cited here: Jamshed Bharucha, Martin McKinney, Michael Gazzaniga, Frank Musiek, Robert Zatorre, Carla Kovacs, Gaurav Shah, Eiling Yee. Special thanks go to Mark Smith and his family.

REFERENCES

1. COHEN, M.R. & I.E. DRABKIN. 1948. A Source Book in Greek Science. McGraw-Hill Book Company, Inc. New York.
2. PISTON, W. 1941/1987. Harmony, 5th edit. Revised and Expanded by Mark DeVoto. W. Norton & Company, Inc. New York.
3. APEL, WILLI, Ed. 1972. Harvard Dictionary of Music, 2d edit. Belknap Press of Harvard University Press. Cambridge, MA.
4. RANDEL, D.M. 1986. The New Harvard Dictionary of Music, 3d edit. Belknap Press of Harvard University Press. Cambridge, MA.

5. TYRRELL, J. & S. SADIE, Eds. 2001. *The New Grove Dictionary of Music and Musicians*, 2d edit. Grove's Dictionaries, Inc. New York.
6. KAESTNER, G. 1909. Untersuchungen über den Gefühlseindruck unanalysierter Zweiklänge. *Psychol. Studien*. **4**: 473–504.
7. VAN DE GEER, J.P., W.J.M. LEVELT & R. PLOMP. 1962. The connotation of musical consonance. *Acta Psychologica* **20**: 308–319.
8. BERNSTEIN, L. 1976. *The Unanswered Question*. Harvard University Press. Cambridge, MA.
9. GUNDLACH, R.H. 1935. Factors determining the characterization of musical phrases. *Am. J. Psychol.* **47**: 624–643.
10. BALKWILL, L. & W.F. THOMPSON. 1999. A cross-cultural investigation of the perception of emotion in music: psycho physical and cultural cues. *Music Percept.* **17**: 43–64.
11. CASTELLANO, M.A., J.J. BHARUCHA & C.L. KRUMHANSL. 1984. Tonal hierarchies in the music of north India. *J. Exp. Psychol. Gen.* **113**: 394–412.
12. KESSLER, E.J., C. HANSEN & R.N. SHEPARD. 1984. Tonal schemata in the perception of music in Bali and in the West. *Music Percept.* **2**: 131–165.
13. TERHARDT, E. 1984. The concept of musical consonance: a link between music and psychoacoustics. *Music Percept.* **1**: 276–295.
14. KRUMHANSL, C.L. 1990. *Cognitive Foundations of Musical Pitch*. Oxford University Press. New York.
15. TRAMO, M.J., J.J. BHARUCHA & F.E. MUSIEK. 1990. Music perception and cognition following bilateral lesions of auditory cortex. *J. Cognit. Neurosci.* **2**: 195–212.
16. HELMHOLTZ, H. 1877/1954. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Translated by A.J. Ellis, 1885. Dover Publications, Inc. New York.
17. STEVENS, S.S. & H. DAVIS. 1938/1983. *Hearing: Its Psychology and Physiology*. American Institute of Physics, Inc. New York.
18. PARNCUTT, R. 1989. *Harmony: A Psychoacoustical Approach*. Springer-Verlag. Berlin.
19. DEUTSCH, D., Ed. 1999. *The Psychology of Music*, 2nd edit. Academic Press. San Diego, CA.
20. DEMANY, L. & C. SEMAL. 1990. Harmonic and melodic octave templates. *J. Acoust. Soc. Am.* **88**: 2126–2135.
21. DEMANY, L., C. SEMAL & R.P. CARLYON. 1991. On the perceptual limits of octave harmony and their origin. *J. Acoust. Soc. Am.* **90**: 3019–3027.
22. KIANG, N.Y.-S., T. WATANABE, E.C. THOMAS & L.F. CLARK. 1965. *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve*. M.I.T. Press. Cambridge, MA.
23. ROSE, J.E., J.F. BRUGGE, D.J. ANDERSON & J.E. HIND. 1967. Phase-locked response to low frequency tones in single auditory nerve fibers of the squirrel monkey. *J. Neurophysiol.* **30**: 769–793.
24. BOOMSLITTER, P. & W. CREEL. 1961. The long pattern hypothesis in harmony and hearing. *J. Music Theory* **5**: 2–31.
25. BRUGGE J.F., D.J. ANDERSON, J.E. HIND & J.E. ROSE. 1969. Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. *J. Neurophysiol.* **32**: 386–401.
26. MOORE, B.C.J. 1980. Neural interspike intervals and pitch. *Audiology* **19**: 363–365.
27. ROSE, J.E. 1980. Neural correlates of some psychoacoustic experiences. *In* *Neural Mechanisms in Behavior*. D. McFadden, Ed.: 1–37. Springer-Verlag. New York.
28. OHGUSHI, K. 1983. The origin of tonality and a possible explanation of the octave enlargement phenomenon. *J. Acoust. Soc. Am.* **73**: 1695–1697.
29. PATTERSON, R.D. 1986. Spiral detection of periodicity and the spiral form of musical scales. *Psychol. Music* **14**: 44–61.
30. MOORE, B.C.J., R.W. PETERS & B.R. GLASBERG. 1985. Thresholds for the detection of inharmonicity in complex tones. *J. Acoust. Soc. Am.* **77**: 1861–1867.
31. VOS, J. 1982. The perception of pure and mistuned musical fifths and major thirds: thresholds for discrimination, beats, and identification. *Percept. & Psychophys.* **32**: 297–313.
32. MALMBERG, C.F. 1918. The perception of consonance and dissonance. *Psychol. Monogr.* **25**: 93–133.
33. GUERNSEY, M. 1928. The role of consonance and dissonance in music. *Am. J. Psychol.* **40**: 173–204.

34. BUTLER, J.W. & P.G. DASTON. 1968. Musical consonance as musical preference: a cross-cultural study. *J. Gen. Psychol.* **79**: 129–142.
35. KAMEOKA, A. & M. KURIYAGAWA. 1969. Consonance theory part II: consonance of complex tones and its calculation method. *J. Acoust. Soc. Am.* **45**: 1460–1469.
36. ZENTNER, M.R. & J. KAGAN. 1998. Infants' perception of consonance and dissonance in music. *Infant Behav. & Dev.* **21**: 483–492.
37. HULSE, S., D.J. BERNARD & R.F. BRAATEN. 1995. Auditory discrimination of chord-based spectral structures by European starlings (*Sturnus vulgaris*). *J. Exp. Psychol. Gen.* **124**: 409–423.
38. TRAMO, M.J., P. CARIANI & B. DELGUTTE. 1992. Representation of tonal consonance and dissonance in the temporal firing patterns of auditory nerve fibers. *Soc. Neurosci. Abstr.* **18**: 382.
39. TRAMO, M.J., M.C. MCKINNEY, P.A. CARIANI & B. DELGUTTE. 2000. Physiology of tonal consonance and dissonance. *Assoc. Res. Otolaryngol. Abstr.* **23**: 275–276.
40. MCKINNEY, M.F., M.J. TRAMO & B. DELGUTTE. 2001. Neural correlates of musical dissonance in the inferior colliculus. In *Physiological and Psychophysical Bases of Auditory Function*. A.J.M. Houtsma *et al.*, Eds.: 71–77. Shaker Publishing. Maastricht, the Netherlands.
41. MCKINNEY, M.F., M.J. TRAMO & B. DELGUTTE. 2001. Neural correlates of the dissonance of musical intervals in the inferior colliculus. *Assoc. Res. Otolaryngol. Abstr.* **24**: 54–55.
42. JAVEL, E. 1980. Coding of AM tones in the chinchilla auditory nerve: implications for the pitch of complex tones. *J. Acoust. Soc. Am.* **68**: 133–146.
43. RHODE, W.S. 1995. Interspike intervals as a correlate of periodicity pitch in cat cochlear nucleus. *J. Acoust. Soc. Am.* **95**: 2414–2429.
44. SCHREINER, C.E. & G. LANGNER. 1988. Coding of temporal patterns in the central auditory nervous system. In *Auditory Function Neurobiological Bases of Hearing*. G.M. Edelman *et al.*, Eds: 337–361. John Wiley & Sons. New York.
45. KRISHNA, B.S. & M.N. SEMPLE. 2000. Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J. Neurophysiol.* **84**: 255–273.
46. PREUSS, A. & P. MULLER-PREUSS. 1990. Processing of amplitude modulated sounds in the medial geniculate body of squirrel monkeys. *Exp. Brain Res.* **79**: 207–211.
47. BIESER, A. & P. MULLER-PREUSS. 1996. Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. *Exp. Brain Res.* **108**: 273–284.
48. FISHMAN, Y.I., D.H. RESER, J.C. AREZZO & M. STEINSCHNEIDER. 2000. Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness. *J. Acoust. Soc. Am.* **108**: 235–246.
49. FISHMAN, Y.I., D.H. RESER, J.C. AREZZO & M. STEINSCHNEIDER. 2000. Complex tone processing in primary auditory cortex of the awake monkey. II. Pitch versus critical band representation. *J. Acoust. Soc. Am.* **108**: 247–262.
50. EHRET, G. & R. ROMAND, Eds. 1997. *The Central Auditory System*. Oxford University Press. New York.
51. PLOMP, R. & W.J.M. LEVELT. 1965. Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* **38**: 548–560.
52. HUTCHINSON, W. & L. KNOPOFF. 1978. The acoustic component of Western consonance. *Interface* **7**: 1–29.
53. RAMEAU, J.-P. 1722/1971. *Treatise on Harmony*. Translated by P. Gossett. Dover Publications, Inc. New York.
54. WIGHTMAN, F.L. & D.M. GREEN. 1974. The perception of pitch. *Am. Sci.* **62**: 208–215.
55. HOUTSMA, A.J.M. & J.L. GOLDSTEIN. 1971. The central origin of the pitch of complex tones: evidence from musical interval recognition. *J. Acoust. Soc. Am.* **51**: 520–529.
56. MOORE, B.C.J. 1997. *An Introduction to the Psychology of Hearing*. Academic Press. San Diego, CA.
57. GALILEO GALILEI. 1638/1954. *Dialogues Concerning Two New Sciences*. Translated by H. Crew & A. de Salvio, 1914. Dover Publications. New York.
58. PICKLES, J.O. 1988. *An Introduction to the Physiology of Hearing*. Academic Press, Harcourt Brace Jovanovich, Publishers. London.

59. CARIANI, P.A. & B. DELGUTTE. 1996. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* **76**: 1698–1716.
60. CARIANI, P.A. & B. DELGUTTE. 1996. Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch. *J. Neurophysiol.* **76**: 1717–1734.
61. CARIANI, P. 1999. Temporal coding of periodicity pitch in the auditory system: an overview. *Neural Plast.* **6**: 147–172.
62. MEDDIS, R. & M.J. HEWITT. 1991. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II. Phase sensitivity. *J. Acoust. Soc. Am.* **89**: 2883–2894.
63. TERHARDT, E. 1974. Pitch, consonance, and harmony. *J. Acoust. Soc. Am.* **55**: 1061–1069.
64. HELMHOLTZ, H.L.F. 1863/1913. *Die Lehre von der Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, 6th edit. F. Vieweg. Braunschweig. Quote translated by E. Terhardt,¹³ pp. 283–284.
65. DAI, H. 1993. On the pitch of two-tone complexes. *J. Acoust. Soc. Am.* **94**: 730–734.
66. GREENWOOD, D.D. 1991. Critical bandwidth and consonance in relation to cochlear frequency-position coordinates. *Hear. Res.* **54**: 164–208.
67. PLOMP, R. & H.J.M. STEENEKEN. 1968. Interference between two simple tones. *J. Acoust. Soc. Am.* **43**: 883–884.
68. MAYER, A.M. 1874. Researches in acoustics. *Am. J. Sci. Arts*, 3rd Ser. **8**: 241–255.
69. CROSS, C.R. & H.M. GOODWIN. 1891. Some considerations regarding Helmholtz's theory of consonance. *Am. J. Sci.* **58**: 1–12.
70. ZWICKER, E., G. FLOTTORP & S.S. STEVENS. 1957. Critical band width in loudness summation. *J. Acoust. Soc. Am.* **29**: 548–557.
71. YOST, W.A. 2000. *Fundamentals of Hearing*. Academic Press. San Diego, CA.
72. KAMEOKA, A. & M. KURIYAGAWA. 1969. Consonance theory part I: consonance of dyads. *J. Acoust. Soc. Am.* **45**: 1451–1459.
73. PRATT, C.C. 1921. Some qualitative aspects of bitonal complexes. *Am. J. Psychol.* **32**: 490–515.
74. BRUES, A.M. 1927. The fusion of non-musical intervals. *Am. J. Psychol.* **38**: 624–638.
75. GUTHRIE, E.R. & H. MORRELL. 1928. The fusion of non-musical intervals. *Am. J. Psychol.* **40**: 624–625.
76. SCHELLENBERG, E.G. & S.E. TREHUB. 1996. Natural musical intervals: evidence from infant listeners. *Psychol. Sci.* **7**: 272–277.
77. TRAMO, M.J. 1990. Impaired perception of relative pure tone pitch following bilateral lesions of auditory cortex in man. *Soc. Neurosci. Abstr.* **16**: 580.
78. SEASHORE, C.E., D. LEWIS & J.G. SAETVEIT. 1960. *Seashore Measures of Musical Talents*. The Psychological Corporation. New York.
79. TRAMO, M.J., A. GRANT & L.D. BRAIDA. 1994. Psychophysical measurements of frequency difference limens for relative pitch discrimination reveal a deficit following bilateral lesions of auditory cortex. *Soc. Neurosci. Abstr.* **20**: 325.
80. JOHNSRUDE, I.S., V.B. PENHUNE & R.J. ZATORRE. 2000. Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* **123**: 155–163.
81. BHARUCHA, J.J., M.J. TRAMO & R.J. ZATORRE. 1993. Abstraction of the missing fundamental following bilateral lesions of auditory cortex. *Soc. Neurosci. Abstr.* **19**: 1687.
82. ZATORRE, R.J. 1988. Pitch perception of complex tones and human temporal-lobe function. *J. Acoust. Soc. Am.* **84**: 566–572.