FOUNDATIONS IN MUSIC PSYCHOLOGY

Theory and Research

edited by

Peter Jason Rentfrow and Daniel J. Levitin

Foundations in Music Psychology

Theory and Research

edited by Peter Jason Rentfrow and Daniel J. Levitin

The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Stone Serif by Westchester Publishing Services. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Rentfrow, Peter J. | Levitin, Daniel J.
Title: Foundations in music psychology : theory and research / edited by Peter Jason Rentfrow and Daniel J. Levitin.
Description: Cambridge, MA : The MIT Press, 2019. | Includes bibliographical references and index.
Identifiers: LCCN 2018018401 | ISBN 9780262039277 (hardcover : alk. paper)
Subjects: LCSH: Music—Psychological aspects. | Musical perception. | Musical ability.
Classification: LCC ML3830 .F7 2019 | DDC 781.1/1—dc23 LC record available at https://lccn.loc.gov/2018018401

10 9 8 7 6 5 4 3 2 1

Contents

5 Musical Intervals, Scales, and Tunings: Auditory Representations and Neural Codes

Peter Cariani

Introduction

Much of the world's music is *tonal music*—music that uses combinations and sequences of pitched sounds to produce its effects on listeners. This chapter describes basic properties of musical scales and tuning systems and examines their major characteristics in terms of the psychophysics of pitch, musical interval perception, consonance, tonal hierarchies, and auditory neural representations.

The first section surveys the structure of musical pitch space and tonal relations. The second describes major features of common scales and tuning systems and takes up questions related to why scales and tuning systems take the forms that they do. The last examines what characteristics of pitch space and scale design might be explicable in terms of early auditory neural representations of pitch. The discussion moves from an introduction to musical pitch perception to descriptive musicology (or "music theory") to psychophysical and auditory neural models.

Tonal music is music whose focus is on patterns of pitch. Musical *melodies* involve trajectories through subjective pitch spaces, whereas musical *harmonies* involve combinations of pitches, both concurrent and successive, within them. Musical *intervals* are the pitch frequency–ratio relations that determine the perceptual geometries of those spaces. In a given musical system, *scales* determine the musical intervals available for use. *Tuning systems* fix the exact note-frequencies and/or their interval ratios associated with the notes of scales, either by explicitly specifying them or by prescribing methods for tuning instruments that implicitly determine them.

Our goal is to address the question "Why these notes and not others?" (Johnston, 2009)—why musical scales might be structured the way they are. Choices of the scales and tunings used in a given musical culture are shaped by both culture and nature. Cultural constraints involve cultural histories and musical practices (which voices and instruments are used to fulfill which musical, psychological ends), whereas natural, biological constraints involve auditory neural processes that subserve pitch perception and cognition.

Cultural, social-psychological factors can be approached through comparative musicological and ethnomusicological studies of musical cultures (e.g., Nettl, 2015). Musical cultures differ in their social uses of music, musical preferences, and available instruments. Although notable but isolated exceptions can be found (McDermott et al., 2016), near-universals exist among tonal systems in the world (Gill & Purves, 2009; Savage, Brown, Sakai, & Currie, 2015; Trehub, Becker, & Morley, 2015). Musical perceptions tend to be less dependent on culture than musical preferences. Perceptual universals that are shared by the vast majority of listeners of all musical cultures include distinctions of musical pitch, octave equivalence, consonance/dissonance, and ability to recognize transposed melodies. Some individual exceptions involve amusic listeners who cannot recognize melodies and listeners who direct attention to different perceptual aspects of sounds (e.g.. pitch height vs. chroma, roughness vs. harmonicity). This chapter focuses on the common aspects of musical scales and tuning systems and discusses their possible origins in temporal neural codes and representations that subserve auditory perception.

A central and abiding question in the psychology of music has involved the origins of integer frequency ratios that were discovered by the Pythagoreans and others (Rameau, 1722/1971; Helmholtz, 1885/1954; Révész, 1954/2001; Boomsliter & Creel, 1961; Plomp et al., 1965; Schellenberg & Trehub,1996; Burns, 1999; Lester, 1999; Tramo et al., 2001; Green & Butler, 2002; Thompson, 2013). Neural representations based on temporal codes in early stages of auditory processing may explain some of these basic features of musical tonality. In addressing the details of auditory representations, this discussion is meant to complement excellent existing introductory surveys and reviews (see Burns & Ward, 1982; Dowling & Harwood, 1986; Handel, 1989; Kendall & Carterette, 1996; Burns, 1999; Thompson, 2013).

Tonal Music

In approaching musical scales and their design, it is helpful to first consider how and why humans use music. Listening to music involves attending to temporal patterns of sonic events for the purpose of influencing internal mental states in desired ways. Music listening can be for pleasure, emotional engagement, cognitive interest, novelty, relaxation, arousal, movement (dance, exercise), beauty, or a host of other reasons, individual and social. Music itself can be broadly defined as those organized sound patterns that can be used in such deliberate ways by a given person or group to achieve their desired psychological, experiential ends. Scales and tuning systems are means to these ends.

Music is relational and purposive in that it is defined in terms of its intended psychological use. Music listening therefore contrasts with involuntary exposure to sounds that induce undesired effects, such as annoyance, irritation, stress, discomfort, and pain. One individual's music is another's unwanted "noise," and even for a single individual, the very same sound patterns can be regarded as music in some situations, when desires are fulfilled, and as annoying noise or even torture in other situations, when desires are blocked or negated.

Musical listening in its purest form is listening in the aesthetic mode—intentionally modifying experience in a manner consistent with one's current purposes and preferences. In contrast to speech, but like poetry, listening to instrumental music evokes and provokes, rather than communicating explicit messages. Thus, "Using music to order a pepperoni pizza for home delivery is unlikely to meet with success" (Janata, 2004, p. 203). In contrast to ecological modes of listening, whose primary purpose is to gather information about the external world for orientation and action, music's purpose is to modulate internal mental states for individual and/or social ends.

Music is conveyed through the medium of sound, such that its perception is mediated mainly by the auditory system. Each musical event evokes multiple auditory perceptual attributes: loudness, duration, relative timing, pitch, timbre, and the spatial attributes of direction, distance, and apparent size. Some of these attribute classes, such as pitch and timbre, have multiple dimensions. Most of these are associated primarily with one set of related acoustic parameters: pitch (dominant periodicity), loudness (intensity), and duration (duration). Timbre is more complex, encompassing different perceptual qualities that covary with spectral shape, onset characteristics, and fluctuations in amplitude, frequency, and phase.

Tonal music is music whose primary focus (its foreground) involves changes in pitch. In tonal music, pitch sequences (melodies) or combinations (harmonies) are the most salient features that distinguish one musical piece from another. Much of the world's music, including most genres of Western music, is tonal in the sense that it is "pitch-centric" music.Tonal music can be contrasted with music that focuses on rhythmic, timbral, or vocal patterns. In rhythm-centered music, such as West African drumming, the main sonic changes of interest and expression involve rhythms—i.e., event timing patterns, rather than melodies or harmonies. In timbre-centered music, such as ambient, electroacoustic, and nonsensical Dadaist phonetic music,[1] the focus is on successions of changing timbral sound qualities. In vocal-centered or spoken music, such as lexical music, chant, and rap, recognizable words and their meanings are paramount, with tonality, rhythm, and timbre playing subsidiary, supporting roles.

A commonsense test of which aspects are most essential is to flatten one or more dimensions by selectively eliminating changes in pitch, rhythm, timbre, or lyrics and evaluating whether a given piece of music has retained its most important essentials—i.e., whether it is still recognizable as the same piece or whether it still retains musical interest.

Musical tonality encompasses those aspects of music that depend on pitch relations, both successive and simultaneous, between notes of a scale.[2] In the conventional music notation of European-derived “Western” musical cultures (see 5.2b later in this chapter), pitch is the vertical dimension, and time is the horizontal dimension. Tonality includes melody and harmony. Melody involves successions of pitches in time, whereas harmony involves pitch relations between concurrent or sequential notes. Melodies involve patterns of movements within pitch space, whereas harmonies characterize the relational structure of that space.

Musical Pitch

The umbrella term *musical pitch* has several meanings. Musical pitches are pitches used in music, for musical purposes, and/or played by musical instruments. In other contexts, musical pitches are those types of pitches whose properties support chroma relations that enable recognition of musical intervals, melodies, and harmonies. A *musical note* is a particular auditory event that produces a clear pitch percept in listeners (e.g., the last note *A Day in the Life* by the Beatles), but it can also refer to a particular pitch class (e.g., the note C_3).

Most pitched-notes in music are produced by human voices and musical instruments that are designed to produce clear pitches. In acoustical terms, these pitches are almost invariably produced by harmonic complex tones with periodic waveforms (for introductions, see Plack & Oxenham, 2005; McDermott & Oxenham, 2008; de Cheveigne, 2010; Moore, 2013; Oxenham, this volume). Although rare in music and nature, pure tones also produce strong pitches that can convey melodies and harmonies just as effectively. For both pure and complex tones, the pitches that are heard correspond very closely to the repetition rates of these periodic waveforms. If the waveform is complex, then this repetition rate is the fundamental frequency (F0), whereas if it is a pure tone, individual harmonic, or frequency component that consists of a single sinusoid, this repetition rate is its frequency (f), and $f = F0$.

An extensive literature exists on the physics and acoustics of musical instruments (Benade, 1990; Rossing, Wheeler, & Moore, 2002; Forster, 2010; Hartmann, 2013; Heller, 2013). Most musical instruments that produce strong, clear pitches involve vibrating strings (e.g., pianos, violins, guitars), air columns (e.g., organs, woodwinds, brass), flexible structures (reeds), and membranes (voices) that produce harmonic complex tones. Less periodic sounds, such as inharmonic complex tones, are produced by bells, kettle drums, metallophones, and lithophones. Inharmonic tones and various kinds of noise stimuli produce weaker pitches with lower pitch saliences, and although such sounds are only very rarely used in musical contexts, some are nevertheless capable of carrying recognizable melodies and harmonies.

Terminology of Pitch Types

To be clear, *note-pitches* will refer to the pitches that are perceived, whereas note-frequencies will refer to the repetition rates of sounds. Pitch is thus a subjective, perceptual quality, whereas repetition rate, frequency, and fundamental frequency are intersubjectively measurable physical-acoustical properties. In operational, experimental psychology terms, being a subjective quality, *note-pitch* is measured in terms of an overt pitch judgment—the physical frequency (F0, *f*) of a reference tone—that a listener has chosen as the best pitch match to a given note. The reference tone is usually a pure tone presented at a specified sound level.

The terminology concerning musical pitch and pitch space can be complicated and confusing because there are several vantage points from which pitches can be regarded and classified. Many labels, such as *spectral pitch*, *virtual pitch*, *periodicity pitch*, and *residue pitch*, make implicit theoretical assumptions about the nature of the neural representations and mechanisms that are thought to produce them. When trying to make sense of the extensive and diverse literature on pitch, it is helpful to keep in mind these many alternative perspectives:

1. *Production/use*: Which sound sources evoke which pitches (e.g., musical instrument pitches, piano-pitches, human voice pitches, *musical pitches* that are produced for tonal music),
2. *Acoustics*: Which types or aspects of sounds produce them (pitches of pure vs. complex tones, *f-pitches* matched to individual frequency components vs. *F0-pitches* matched to fundamentals of groups of harmonics),
3. *Auditory perception*: Perceptual functions they enable (e.g., *musical pitches or chroma-pitches* that support perception of chroma relations, such as recognition of intervals and transposed melodies),
4. *Presumptive auditory representation*: Which auditory representations might subserve particular pitch percepts (e.g., *spectral vs. periodicity pitch*),
5. *Presumptive neural substrate*: Which neural codes and mechanisms produce particular pitch percepts (*cochlear place pitch* vs. *interspike interval pitch*), and
6. *Music-theoretic pitch class* (chroma-equivalence class): The music-theoretic function of a particular pitch in a musical context.

Basic Pitch Attributes

Many basic aspects of scales reflect the organization of underlying perceptual qualities associated with pitch, i.e., the basic, dimensional geometry of pitch space. Musical pitch has four basic attributes: *pitch height*, *pitch chroma*, *pitch strength/salience*, and *vibrato*.

By far the most important attributes for tonal music are pitch chroma and relative height because these attributes form the basis for melody and harmony. The vast majority of listeners perceive pitch height and chroma in relative rather than absolute terms. They can evaluate whether one pitch is higher or lower than another (*relative pitch height*), whether one pitch is stronger than another (*pitch strength* or *salience*), and whether two pitches bear some distinctive similarity or relation (e.g., an octave apart) to each other (*relative pitch chroma, musical* *interval discrimination/recognition*). Pitch salience (Zwicker & Fastl, 1999) can be regarded as an auditory analogue of color saturation in vision, but few musical genres make use of variations in this attribute.

Vibrato, or lack thereof, involves qualities related to the constancy of sound periodicity. Slow oscillations in periodicity create qualities of wobbling, wavering, fluctuating, or fluttering pitch that can vary in frequency extent (usually fractions of a semitone, 0.5–3 percent of F0) and rate of oscillation (usually 5–8 Hz or less) (Sundberg, 1994; Vurma & Ross, 2006). Vibrato is found in the vocal and instrumental music of many musical cultures worldwide and is used mainly for emotional expression and ornamentation.

Musical note-pitches are highly stable. Musical pitch is highly invariant with respect to acoustic parameters that dramatically alter other perceptual attributes, such as loudness, duration, timbre, and sound direction. Timbral alterations created by changes in gross spectral shape, spectral tilt, attack/decay, and frequency-onset characteristics have little effect on musical pitch. Because of these strong perceptual invariants, different musical instruments and voices can reliably play the same pitches, and scales and practical tuning systems can be constructed that are largely independent of these parameters. The small deviations from these invariances, such as pitch shifts with sound level, are always much smaller for the harmonic complexes of musical instruments than for pure tones. Fortunately these shifts are not large enough to make a practical difference for the vast majority of musical tuning procedures (Dowling & Harwood, 1986, pp. 50–51).

Lastly, several different pitches can coexist together in the auditory scene. These are multiple pitches that can be heard in single notes and groups of notes (chords). Although the strongest pitch of a single musical note usually lies at its fundamental frequency, multiple weaker pitches related to individual harmonics above the fundamental can be heard out, provided that these harmonics are separated by at least 20 percent in frequency (de Cheveigné, 2010; Plomp, 1976). As harmonic number increases and harmonic spacing in logarithmic terms consequently decreases, pitch saliences of individual harmonics successively weaken until harmonics above the 5th can no longer be or aurally resolved, i.e. they can no longer be "heard out" as separate pitches (Plomp, 1976).[3]

When multiple notes are sounded together in chords, several pitches can also potentially be heard (Parncutt, 1989). Some of these pitches correspond to the fundamentals of the individual notes, whereas others can be heard that correspond to the common fundamental frequency of all of the notes. Figure 5.10c later in this chapter shows model predictions of various pitches that can be heard in a major triad. Because the fundamental frequency of the chord is the highest common subharmonic of the note-F0s, and each note-F0 is the highest common subharmonic of the harmonics of each note, the fundamental frequency of the note-F0s is also the fundamental frequency of all of the harmonics of the notes. This "grand fundamental" of the notes is also known as the *fundamental bass* of the chord. This is Rameau's *basse fundamentale*, which has played an important role in the development of theories of harmony (Rameau, 1772/1971; Riemann, 1905/2011; Terhardt, 1974, 1984). The stronger this fundamental bass, the more unified the pitch, and the more stable the chord in terms of possible competing pitches. Neural models for harmony that are based on such mutually reinforcing subharmonic patterns are discussed in the last section of this chapter.

Pitch Space

Musical tonality depends on pitch relations, whose structure can be considered in terms of the geometry of pitch space. Two-factor models of pitch space represent pitch height as a linear dimension (figure 5.1a) and chroma as a circular dimension (figure 5.1b, top), with the rising pitches of musical scales represented as a helix (figure 5.1b–d) in a cylindrical space (Bachem, 1950; Burns & Ward, 1982; Deutsch, 1999; Drobisch, 1852; Révész, 1954/2001; Shepard, 1964, 1982a, 1982b; Ueda & Ohgushi, 1987; Ward, 1999). Some depictions, such as the "tonal bell" of Ruckmick (1929) (figure 5.1c) and the *oto-no-horin* of Yatabe (1962) (figure 5.1d), use the radius of the helix to denote the strength of chroma-related qualities (*chroma salience*) and the frequency range over which they exist. This existence range extends roughly from just below C_1 (33 Hz) to just above above C_8 (~4.2 kHz). More elaborate geometric forms, such as multiple helices embedded in toruses (see figure 4.8), have been conceived to attempt to capture the neighborhood perceptual distance relations between other note and key relations within the octave (Shepard, 1982a, 1982b; Krumhansl, 1990).

Musical scales universally organize notes sequentially according to pitch height, from low to high, but they also incorporate the circular dimension of pitch chroma, repeating their organization with each successive octave.

Musical intervals are distinctive chroma relations. Notes an octave apart share the same chroma class and are perceived as similar. Pairs of notes separated by other frequency-ratios constitute other characteristic chroma relations. *Intervallic structure*, the pitch chroma relations of successive notes to the tonic, is the most important determinant of the essential form of a melody: it is what distinguishes a particular melody from others and that makes it recognizable and memorable.

For most listeners, pitch chroma distinctions can be clearly discerned only for periodicities (*f*, F0) roughly between 25 Hz and 5 kHz. Above C_8, at 4.2 kHz, pitch chroma distinctions, such as octave similarity and recognition of other intervals, are severely degraded and tend to disappear altogether (figure 5.1c and 5.1d). Notably, some individuals have higher upper limits of chroma perception (Burns, 1999).[4] The disappearance of pitch chroma at high frequencies leaves only crude pitch height distinctions. The resulting chroma-less patterns of the rising versus falling pitch transitions of melodies are called *pitch contours*.

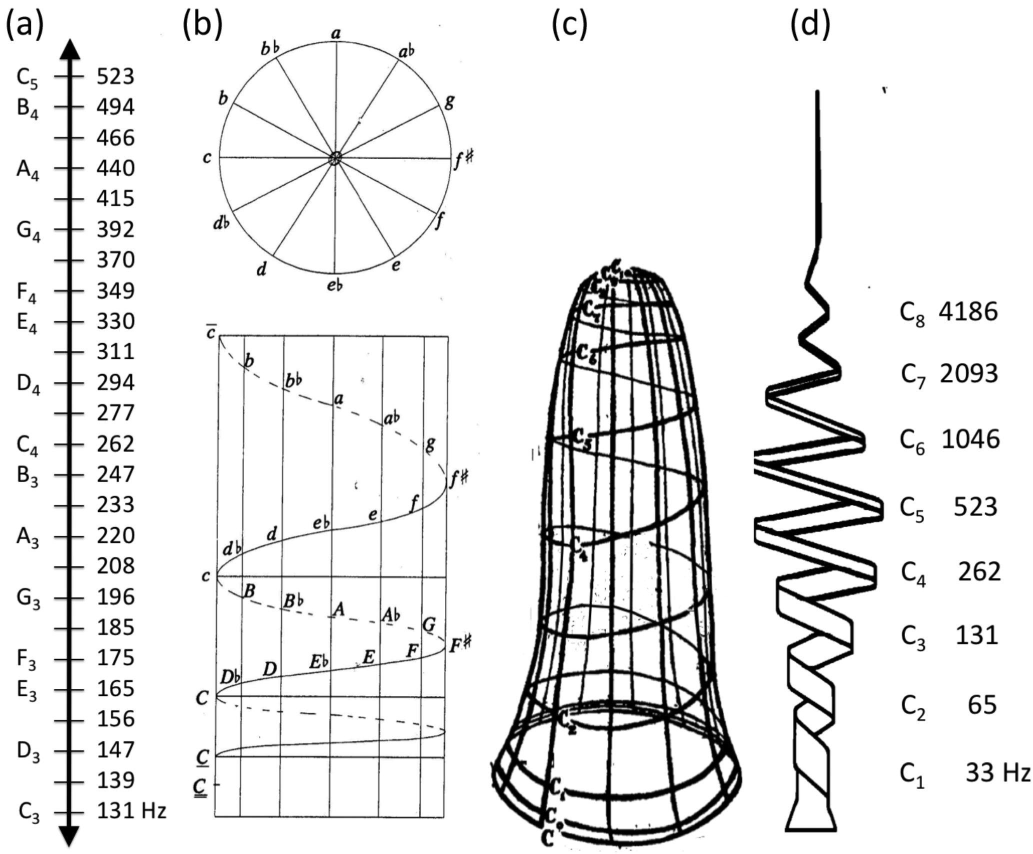

Figure 5.1
Four low-dimensional geometrical depictions of pitch space. (a) A linear map of pitch height in two middle octave registers from C_3 to C_5. (b) Top. Circular dimension of pitch chroma, which projects similar pitches within each octave into the same class. Bottom. Helical, "two-component" representation of pitch (Révész, 1954), with the vertical dimension representing pitch height and the angle around the vertical axis representing pitch chroma (after Drobisch, 1852, reprinted in Butler, 1992). (c) The tonal bell of Ruckmick (1929). The structure is an attempt to incorporate the attribute of "volume" or "breadth" (see also Boring, 1942, pp. 375–381) and the existence region of tonality in one structure. (d) A thinning helix representation that indicates the diminishing dimension of chroma at higher frequencies (redrawn from Yatabe, 1982, as reprinted in Ueda & Ohgushi, 1987). Corresponding note frequencies are shown for the C chroma class. Note that Ruckmick's C_0 is our present-day C_1 (33 Hz). See also pitch helices in figures 4.1 and 11.1.

The frequency limits of musical tonality thus determine the ranges of pitches that can be used for tonal musical purposes—that is, to produce recognizable musical melodies and harmonies. These limits explain why musical scales do not cover the entire frequency range of hearing (~20 Hz–20 kHz). The F0-frequency range of notes on the piano, 27 Hz–4.2 kHz, is also approximately coextensive with the frequency range for musical tonality (figure 5.2) for essentially the same reason: it spans the range of usable musical pitches that can support chroma relations.

Despite preservation of their pitch contours, melodies transposed into octaves above 4–5 kHz are perceived to be highly distorted or even completely unrecognizable—"transposition behavior becomes erratic" (Atteneave & Olson, 1971, p. 147). They go on to say:

> Perhaps the most provocative finding of these studies is the abrupt deterioration of musical transposition that occurred at about 5,000 Hz, both in the musical subjects of Experiment I and in the non-musical subjects of Experiment II. Something changes rather dramatically at this level; phenomenally it is identifiable as a loss of musical quality, whatever that may be. (p. 163)

What disappear above 5 kHz are chroma relations, the substrates for octave similarity, musical intervallic relations, and melodic structure. Chroma relations thus appear to be critical for robust melody recognition, with patterns of pitch contours playing much weaker roles.

The existence region for musical tonality has potentially strong implications for the nature of the neural codes and representations that subserve *musical pitch*, i.e., those types of pitched sounds that can support chroma relations (see Burns, 1999; Atteneave & Olson, 1971). Types of sounds that produce relatively strong musical pitches are low-frequency pure tones (50 Hz $< f <$ ~4.5 kHz) and harmonic complexes (~25 Hz $<$ F0 $<$ ~1.2 kHz) consisting of perceptually resolved harmonics (typically three or more low harmonics, $n \leq 5$, with wide harmonic separations, $\Delta f > 20\%$). Weaker musical pitches can be produced by sets of perceptually unresolved harmonics (higher harmonics, $n > 5$, with narrow harmonic separations $\Delta f < 20\%$), band-pass noise, pattern-repetition-noise (e.g., iterated ripple noise, maximum length sequences), and amplitude-modulated noise. Nevertheless, all of these stimuli support pitch chroma relations; that is, they support octave matching and they can carry a recognizable melody. The upper frequency limit of such musical pitches coincides with the limit of statistically significant spike timing ("phase-locking") in the auditory nerve (Johnson, 1978), suggesting that chroma-relations might be based on temporal representations (see neural coding discussion below and figure 5.7 later in this chapter).

Listeners are able to distinguish major from minor pure tone triads up to a mean frequency of roughly 3 kHz and down to a mean frequency of just above 100 Hz (Biasutti, 1995). On the low-frequency end of the tonality existence region, clear F0 pitches and their chroma relations for harmonic complexes disappear around F0 = 20–25 Hz (Pressnitzer, Patterson, & Krumbholz, 2001; de Cheveigne, 2010). Below this lower limit, stimuli with dominant periodicities in the 10–20 Hz register are perceived as infra-pitch (Warren & Bashford, 1981; Warren, 1999) and

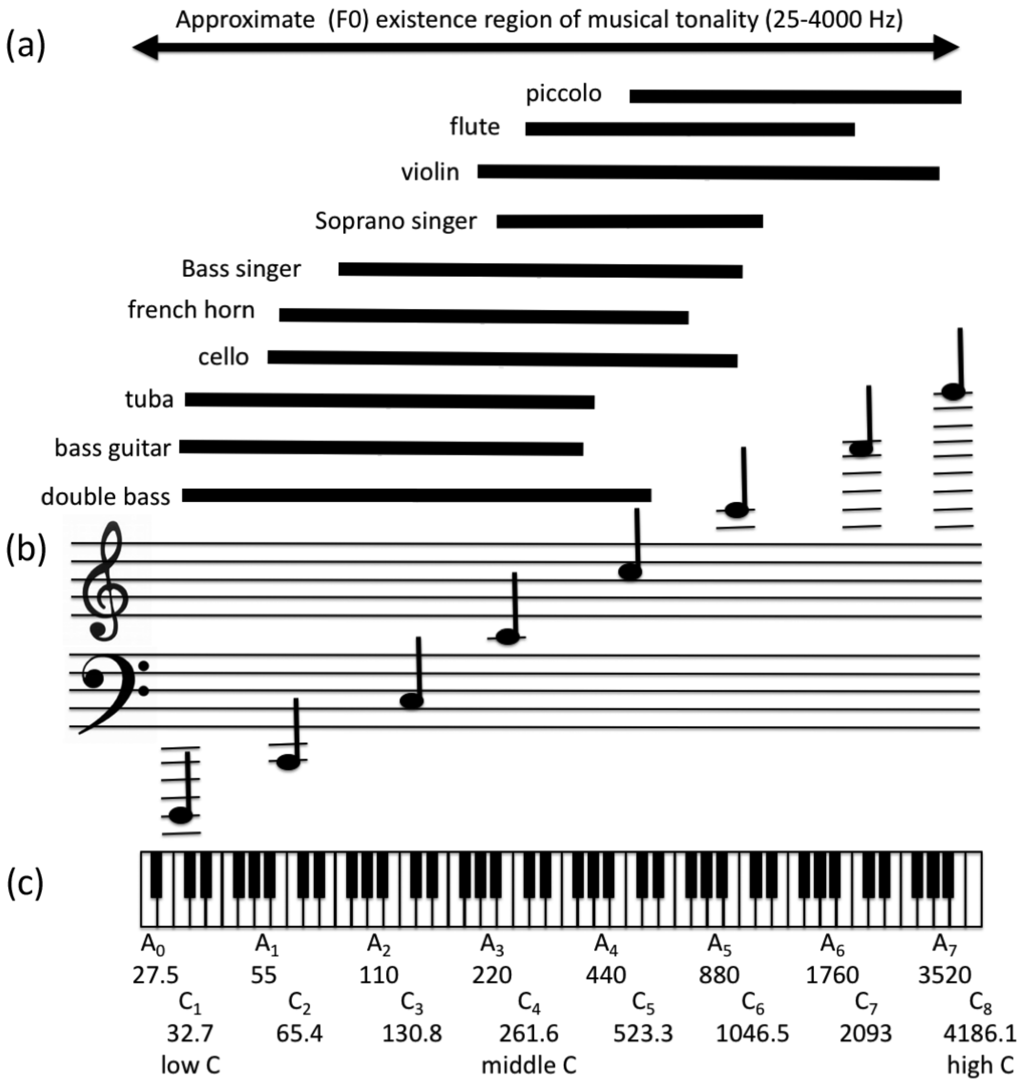

Figure 5.2
Frequency range of musical pitch. (a) Existence region for musical tonality (octave matching, recognition of transposed melodies) and approximate F0-pitch ranges for selected musical instruments. (b) Mapping of the F0-pitch range for an eighty-eight-key piano onto Western music notation. (c) Piano keyboard and corresponding numerical frequency values for F0-pitches, in Hz. The piano keyboard corresponds to the chromatic scale (twelve notes per octave), typically using equal temperament tuning (table 5.1), with adjacent notes a semitone (~6%) apart. Note that the repeating pattern of piano keys mirrors the repeating octave-based pattern of pitch chroma sequences and scales.

have qualities such as "motor-boating." For repetition rates of less than ~10–12 Hz, individual auditory events can be distinguished, and the repeating temporal patterns are perceived as rhythmic patterns of events rather than infra-pitches or pitches.

Consonance, broadly construed, describes perceptual qualities and their aesthetic-hedonic preferences that are related to sequential or concurrent combinations of pitched-notes. The perceptual qualities can be grouped into those related to roughness and those related to harmonicity. *Roughness* (also known as *sensory dissonance*) is produced by the beating of nearby harmonics, whereas qualities related to *harmonicity* (pitch multiplicity, stability, and unity) involve the degrees to which neural representations associated with multiple pitches either mutually reinforce or compete with each other (see discussion of harmonicity in chapter 1). Arguably, aspects of consonance related to harmonicity are closely related to harmony, and therefore influence choices of sets of notes in tonal systems (scales). Aspects related to roughness are relatively more important for tuning systems.

The first note of the scale, called its root or *tonic*, determines the tonal context or reference point for all other notes of the scale that follow. The note corresponding to each musical interval has a characteristic perceptual relationship to the tonic, such that different musical intervals constitute different perceptual distances to the tonic and to each other. In music theory, each interval is said to have a unique tonal function; in music psychology, each interval has a unique location in a *tonal hierarchy* (Krumhansl, 1990, 1991; Bigand, 1993; Krumhansl, 2000; Krumhansl & Cuddy, 2008; Russo, this volume). Tonal hierarchies establish near-far neighborhood relations between notes, chords, and keys that depend on the current tonal context; that context is the tonal center (tonic) that has been established in auditory short-term pitch memory by preceding notes.

Most listeners can make only rather crude estimates of absolute note-pitches presented in isolation, on the order of identifying which octave register a pitch lies (five accurately identifiable pitch regions [Pollack 1952]; average errors of judgment of "tone height" in the range of five to nine semitones [Bachem, 1950]). In contrast, the few listeners who have *absolute pitch* (AP) can reliably identify individual notes of chromatic scales (Levitin, this volume), that is, with semitone resolution and on the order of seventy-five separate absolute pitch categories (Burns, 1999). Some AP listeners can also make absolute pitch identifications finer than a semitone (e.g., how much a note is mistuned regarding its standard frequency). Most AP listeners also have relative pitch, but in some cases strong reliance on AP can interfere with relative pitch, such that AP listeners may do much more poorly than their non-AP counterparts when melodic recognition requires transposition into unfamiliar tonal contexts (Miyazaki, 1992; Miyazaki, 2004).

Pitch discrimination involves distinguishing pitches produced by two acoustic stimuli (musical notes) with different fundamental frequencies (same vs. different pitch). For musical pitches, the finest, just-noticeable-distinctions (jnd's) are roughly proportional to the F0s involved (Weber's Law, $\Delta F0/F0 \sim$ constant), typically on the order of 1 percent or less in frequency (~1/6th of a semitone). Under ideal listening conditions and with some training, pitch acuity can be still an order of magnitude better than this. Thus, for most listeners, the ability to discriminate pitches is roughly two to three orders of magnitudes finer than the ability to identify them in absolute terms. Even for AP listeners, pitch discrimination is still greater than an order of magnitude more sensitive than their ability to identify absolute pitches.

Most Musical Pitch Perception Is Relative, Involving Relations between Note Pitches

Listeners can make relative judgments regarding whether a given pitch is higher or lower than another (*relative pitch height*). Frequency differences required for these identifications are roughly comparable to those for pitch discrimination, albeit with considerable (and some very counterintuitive) individual differences.[5]

Listeners can also make estimates of *relative pitch distance*, as, for example, what pitch is "half" or "twice" as high as another. The *mel scale* was developed by the psychophysicist S. S. Stevens to attempt to characterize subjective pitch height differences in terms of such magnitude estimations (the vertical scale in figure 5.1c). Pitch space can be depicted in terms of a helix-cylinder with pitch height in mels and chroma as central angles around the cylinder (Ward, 1999, reprinted in Ueda & Ohgushi, 1987). The existence range of musical tonality extends from 0 mels a few semitones below C_1 (33 Hz) to roughly 2500 mels a few semitones above C_8 (~5 kHz). For middle registers the scale is roughly consistent with the proportional, log-frequency scaling of pitch chroma (Zwicker & Fastl, 1999), but outside of this range, especially for the lowest and highest registers, it deviates from them very substantially (Atteneave & Olson, 1971; Hartmann, 1993). The scale is generally viewed as unreliable and unmusical (Rasch & Plomp, 1999) because it is not clear what auditory quality the listener is using to estimate magnitude: different listeners may focus on different aspects of pitch such as pitch height versus chroma or, in neural terms, cochlear place versus temporal cues (Shepard, 1982b).

The vast majority of listeners have *relative pitch*. The term usually means that listeners can perceive relative pitch chroma relations, such that they can discriminate and recognize melodies. Exceptions are those with amusia—that is, those who have poor pitch discrimination and cannot reliably distinguish or recognize melodies (Peretz, 2013, 2016; Quintin, Lense, & Tramo, this volume). Relative pitch is what enables listeners to easily recognize melodies irrespective of the keys in which they are played (i.e., which tonic or starting note is used). Melodies are said to be perceptually invariant with respect to *transposition*. As long as all note-frequencies (F0s) are multiplied by a constant factor, melodies remain highly recognizable as the same patterns of relative pitches. For example, the pitch pattern of the first seven notes of "Twinkle, twinkle little star" in the key of C (notes CCGGAAG) sounds similar to its transposed version in the key of F (notes FFCCDDC). What is preserved are the frequency ratios between the successive notes of the melody relative to the first note (1, 1, 1.5, 1.5, 1.68, 1.68, 1.5). In the transposition, although all frequencies have been multiplied by a constant ratio (1.335), all of the ratios in relation to the first note (the tonic) are preserved, as well as those among all pairs of notes. The melody remains invariant, recognizable as the same melody, much as a triangle's shape remains invariant provided that all of its lengths are multiplied by the same factor.

Musical intervals, most broadly, are the frequency-ratios between pairs of notes (Partch, 1974, pp. 76–85). More narrowly, in the contexts of scales and tonal music, musical intervals also have a more specific meaning as the frequency-ratios of notes to the tonic (i.e., to the first note of a scale or to the tonal center/reference of a melody—its key). Thus, transposition preserves both *musical interval*s (frequency ratios) and *pitch contours* (patterns of rising and falling pitches, i.e., directions of pitch height changes irrespective of their magnitudes). Of these, preservation of musical intervals is by far the more important for melodic recognition—even when contour is preserved, distortion of intervals by more than a semitone rapidly degrades melodic similarity and recognition.

Contour and intervallic expectations appear to be mediated by separate mental processes (Graves, Micheyl, & Oxenham, 2014). Contour depends on changes in pitch direction, whereas intervallic relations depend on pitch ratios, especially ratio relations with tonal centers. Distortions of intervallic structure are more noticeable for familiar and tonal melodies that incorporate tonal-hierarchical relations (Graves & Oxenham, 2017), and less apparent for unfamiliar and less memorable atonal, randomly chosen melodic sequences, where listeners tend to rely more on contour cues (Dowling 1971, 1978). Whereas the pitch acuity and special pitch relations, such as octave similarity, appear to be specific to pitch perception, contour discrimination and recognition have analogues in other auditory percepts (loudness) and in other modalities (visual brightness) (McDermott, Keebler, Micheyl, & Oxenham, 2010a).

Melodic invariance under transposition depends on maintaining constant frequency ratios among the notes.[6] Other transformations, such as shifting the fundamentals of all notes by a constant frequency, preserve melodic contours (up/down changes in pitch direction), but distort intervals (frequency ratios). As these nonmultiplicative (nonproportional, nonlogarithmic) shifts increase, first notes are heard as "sour" or mistuned, and eventually the melody itself becomes unrecognizable. Another transformation, *melodic inversion*, inverts the musical intervals of a melody. For example, a fifth upward ($3/2\,f$) becomes a fifth downward ($2/3\,f$), altering both contour sign and frequency ratio. Melodic inversion immediately violates melodic invariance. If one preserves the pattern of contour signs by octave-shifts of the inverted notes in the direction opposite of the inversion, one has produced an entirely new melody.

Inverted chords involve shifting one of the notes of a note-dyad or -triad by an octave upward or downward, thereby changing the pattern of respective pitch heights of the notes, but preserving the chroma relations and relations vis-à-vis the tonic. Inverted harmonic dyads and triads are recognizable as "the same" interval or chord as their noninverted counterparts (*harmonic invariance under inversion*), and are consequently treated as equivalents in music theory. Melodic sequences inverted this way, with single octave shifts that preserve chroma, are also recognizable, provided that the notes, such as those of the piano, have harmonic overlaps in at least one common octave register. Deutsch's (1972) well-known, unrecognizable "mysterious melody" uses octave shifts with melodies of pure tones such that there are no common registers in which interactions between sequential partials can occur. Likewise, rhythmic patterns of streams of pure tones in different frequency registers tend to separate, whereas those that overlap in the same registers fuse (Bregman, 1990). These phenomena provide strong evidence for auditory processing within roughly octave-wide frequency bands.

Inverting the temporal order of notes, or playing a non-palindromic melody backward, also disrupts melodic recognition. This disruption is likely due to the temporal asymmetry of short-term auditory memory; that is, notes are perceived in relation to their most recent predecessors, their immediate tonal contexts.

After unison (1:1), the octave (2:1) is the most salient musical interval. Octave similarity is thought to be "nearly universal to all cultures of the world" (Dowling & Harwood, 1986). Almost all listeners can perceive tones with F0s an octave apart (a frequency ratio of 2) as being more similar to each other than to other chromatic notes within the octave. Octave judgements are extremely precise, typically within 1 percent, and repeatable as long as listeners attend to chroma and not to competing pitch height cues (Hoeschele, Weisman, & Sturdy, 2012).[7] Using the method of adjustment, octaves can be accurately replicated, with standard deviations on the order of 0.5 percent, whereas those for other intervals within the octave are on the order of 1 percent (Burns, 1999).

Octave matches are most accurate and stable for complex harmonic tones when the spectral composition (timbre) of the two notes is similar. Because complex harmonic tones an octave apart share half of their harmonics, listeners in such situations can match individual resolved harmonics as well as F0-pitches. For complex musical tones this similarity extends over multiple octaves. Virtually all of the world's scales, if they span more than an octave, repeat at octave intervals. Thus, because of octave similarity, pitch space has a circularity in it: by moving upward in pitch, one eventually comes back to a note that is similar to the one at which one started. There is an analogous, circular, repeating structure to interspike interval representations of pitch (figure 5.10 later in this chapter).

Octave stretch involves octave matches of pairs of both pure and complex tones that deviate *very* slightly from F0 ratios of exactly 2. The stretch is greatest for frequencies in the highest register of the piano (C_7–C_8, 2–4 kHz). Listeners match pure tones to slightly stretched octaves (1–4 percent) (Zwicker & Fastl, 1999), whereas for complex tones, this enlargement is much smaller (< 1 percent). In the piano midrange the stretch is on the order of 0.5 percent (2.009) (Dowling & Harwood, 1986).

Musical Intervals

Most listeners identify and recognize different musical intervals (frequency ratios) presented either melodically (sequentially) or harmonically (concurrently), with roughly similar acuity.

In listening to music, sequences of pitches are perceived relative to established tonal contexts, such that notes bear relations to previously presented ones, especially the tonic. In these situations listeners may be able to use pitch cues (alignments between common harmonics and/or subharmonics, as in figure 5.10 later in this chapter) in order to perceive intervals.

Experiments designed to estimate the ability to discriminate frequency-ratios without such pitch cues use isolated, transposed musical intervals. Typically these experiments use a two-interval four-note paradigm in which a first interval N_1N_2 is compared with a second M_1M_2. In order to eliminate use of simple pitch comparison cues (i.e., discriminating the pitch of N_2 vs. M_2), the frequency of M_1 is randomly varied. In this method, tonal context is provided by only one note, whereas in a tonal melodic context, all of the notes in the melody play a role to some extent in reinforcing the tonic as a reference pitch (tonality induction). Where pitch cues are involved, interval discriminations are precise, whereas when these cues are minimized or eliminated, interval discrimination is relatively coarse. By comparison with pitch discrimination and musical intervals in tonal contexts, discrimination of isolated transposed intervals presented sequentially is much coarser, on the order of half a semitone (3 percent) for experienced musicians and typically higher, from 1–3 semitones (6–18 percent) for listeners without musical training (McDermott, Keebler, Micheyl, & Oxenham, 2010a).

Simply because preserving musical intervals is critical for melodic recognition, one should not necessarily assume that intervals are explicitly identified and that melody perception consists of sequences of identified isolated intervals between temporally adjacent pairs of notes. A melody is perhaps better conceived as a cohesive web of pitch relations than a sequence of musical intervals. Burns and Ward (1982, pp. 264–265) caution that “the perception of isolated musical intervals may have little to do with the perception of melody” because “melodies are perceived as Gestalts, or patterns, rather than as a succession of individual intervals, and that interval magnitude is only a small factor in the total percept.”

Several excellent reviews of interval perception as a form of categorical perception exist in the literature (Burns, 1999; Burns & Ward, 1982; Thompson, 2013). Burns (1999) argues that musical interval perception is a form of acquired categorical perception *par excellence*, with category boundaries being even more stable than for phonetic categories. That the categories are learned is consistent with improvement with training, up to a point. “Although the best possessors of relative pitch are able to identify chromatic semitones without error, they are not able to identify the quarter tones between chromatic semitones with perfect consistency” (Burns, 1999, p. 22).

Burns considers whether patterns of interval perception support the existence of "natural" musical intervals based on simple frequency ratios (i.e., just-tuned intervals), but he compares just tuning mainly against the 12-tone equal temperament chromatic scale, which is a very close approximation (within 1 percent) to the just-tuned Pythagorean consonances, twelve equally tempered notes being the closest approximation (see figure 5.5 and discussion later in this chapter). Given the relatively coarse quartertone to semitone (3–6 percent) resolution of sequential interval perception, it is not surprising that interval perception does not necessarily follow one or the other set of tuning standards.

Tonal Context Is Important for Perception of Intervallic Relations

In tonal music, unlike situations involving isolated presentations of transposed intervals, all successive notes are heard in the tonal context formed by the last preceding notes. The building up of a tonal context by several successive notes can influence detection of mistunings in melodies.

In a Japanese study (Umemoto, 1990; see also Graves & Oxenham, 2017), student subjects with differing degrees of musical experience (music vs. psychology majors, *H*igh vs. *L*ow experience) were able to detect quarter-tone (3 percent) mistunings in *T*onal and *A*tonal melodies consisting of five notes. Tonal melodies consist of notes taken from the diatonic scale for a single key, whereas notes of atonal melodies were drawn from the chromatic scale. Detection was always worst for the first tone (LT 15 percent correct, HT 60 percent correct), and best for the last three (LT ~70 percent correct, HT ~90 percent correct). Detection of mistuning in tonal melodies was consistently better than for atonal ones in all conditions by ~20 percent.

Ability to accurately produce musical intervals can also provide insights into the relational structure of pitch space. The accuracy of production of musical intervals in performance (*intonation*) relative to fixed scales depends critically on the nature of the musical genre, because in many genres microtonal inflections are used to convey affective cues, such that strict pitch accuracy would be heard as mechanical and cold. In genres in which notes are concurrent and sustained and a high value is placed on intonation, such as barbershop quartets, musical intervals are replicated with high accuracy (less than 3 cents, Vurma & Ross, 2006; 10 cents, ~0.6 percent, Burns, 1999).

An abiding question has been whether musicians and vocalists, left to their own devices, tend to produce note-frequencies biased towards just-tuned rather than equally tempered (ET) intervals (see next section), but there appears to be no bias toward either just intonation or equal temperament tunings (i.e., the singers align to each other's pitches rather than to a scale, and this 10-cent accuracy is larger than most JI-ET discrepancies; see table 5.1 later in this chapter, last column). Similar high precisions are seen for solo violinists as for singers (Loosen, 1993) in Burns (1999), but again there appears to be no bias toward one or the other tuning system. This indifference may be due to the close approximation of ET to just tunings (<< 1 percent) for those (consonant) musical intervals that would be most critical. The differences between just and ET tunings are usually small in relation to production accuracy and pitch variability, such as vibrato.

Summary

In our view, pitch space geometries, tonal hierarchies, and harmonicity-related aspects of consonance play the most important roles in determining the overall structures of scales, melodies, and harmonies in tonal music, whereas aspects of consonance related to roughness caused by beatings among harmonics are most important for choosing tuning systems for fixed pitch instruments.

Scales and Tuning Systems

Descriptions of scales and tuning systems used in musical practice generally come under the rubrics of music theory, musicology, and ethnomusicology. Out of the continuum of pitch sensations on which we can make thousands of possible fine distinctions, most of the world's tonal music uses relatively small sets of discrete distinguishable pitch classes (see figure 5.3).

Scales

Nineteen different scales and tuning systems from different world musical cultures are shown in figure 5.3 (Justus & Hutsler, 2005). Other recent comparative studies have examined scales within the wider context of all musical properties that might characterize "musicality" (Savage et al., 2015; Trehub et al., 2015).

Scales are defined by distributions of the approximate frequency-ratios of notes within a single octave, which covers a two-fold range of note-frequencies. The horizontal axis is scaled in terms of log-frequency, such that equal frequency-ratios span equal spatial extents. Dark vertical bars indicate the frequency positions of scale-notes, whereas, for comparison and reference, gray bars indicate positions of the Western twelve-note, equally tempered chromatic scale. This scale divides the octave into twelve equal ratio-steps ("*t*welve tone *e*qual *t*emperament" or 12-TET), each of which is a semitone. A semitone is a ~6 percent change in frequency ($2^{1/12} = 1.0595$), and a whole tone consists of two semitones ($2^{2/12} = 1.1225$) or ~12 percent. Adjacent keys on a piano produce notes a semitone apart in fundamental frequency.

Virtually all of the world's scales subdivide the octave into a relatively small number (5–24) of discrete pitches, repeating this octave organization to cover the entire pitch range that is used. The eight piano keys of figure 5.1 span one octave, whereas the eighty-eight keys of the entire piano keyboard in figure 5.2 span almost the entire pitch range of musical tonality,

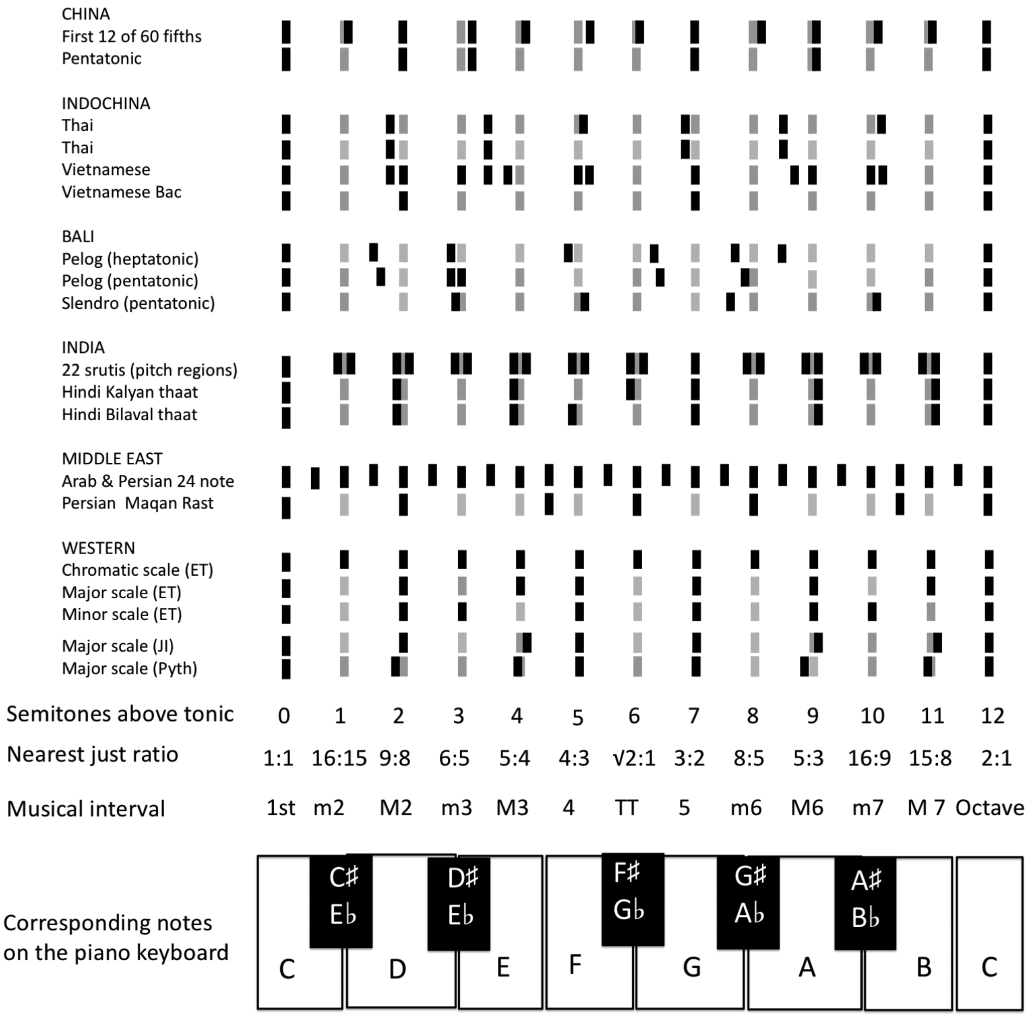

Figure 5.3
Comparison of the scales and tuning systems in music from different parts of the world. The chart of scales was adapted from Justus & Hutsler (2005). Top: An octave range of note-frequencies (the horizontal positions of the black vertical marks) is shown for each scale. The horizontal axis indicates note-positions within the octave, in log-frequency units (semitones). Gray vertical marks indicate the note-positions of the equally tempered (12-TET), Western chromatic scale. Middle: Nearest just intonation ratio and musical interval abbreviations (m: minor, M: major, TT: tritone). Bottom: Approximate equivalent locations of scale notes on the piano keyboard (C_n–C_{n+1}).

about seven and a half octaves, from 27.5 to 4,186 Hz. Above and below these repetition frequencies, octaves, musical intervals and melodies become unrecognizable for most listeners.

The reasons for this octave-based organization of scales involve the two-dimensional linear-circular structure of pitch space (figure 5.3b–d). The spiral structure of pitch space in turn may be due to the periodic nature of neural interspike interval representations of sound (figure 5.7 later in this chapter). Likewise, the upper frequency limits of musical tonality may be due to the operating limits of those temporal neural representations.

Although different scales divide the octave in different ways (figure 5.3), there are some commonalities between numbers of notes and their placements. Pentatonic, heptatonic/septatonic, octatonic/diatonic, and chromatic scales are common, having, respectively, 5, 7, 8, and 12 separate notes distributed throughout in the octave. Some of these scales distribute notes approximately uniformly, in logarithmic frequency-ratio terms, whereas others have decidedly nonuniform distributions.

Consonant musical intervals are incorporated into most of the world's scales. Many world scales share similar notes, with almost all containing notes near the fifth and fourth notes and with many including notes near the sixth and third notes of the major Western diatonic scale (Shepard, 1982b). The frequency ratios of these shared notes relative to the first note, the tonic, are all near simple integer ratios: unison (1:1), octave (2:1), fifth (3:2), fourth (4:3), sixth (5:3), major third (5:4), and minor third (6:5). These simple ratios all produce relatively consonant note combinations. When pairs of notes with these ratios are played together, they are perceived as more consonant (smoother, less rough, more unified) and are preferred by many listeners, who find them more pleasant (euphonious) than other more complex ratios, such as the tritone ($\sqrt{2}$) or the minor second (16/15).

The simple, consonant ratios are referred to as the *Pythagorean consonances*, because these particular divisions of the octave were discovered by the Pythagoreans in fifth century BCE from their experiments with monochords with moveable bridges (Tenney, 1988; Barbour, 2004).[8] These ratios and their accompanying smooth, blended combinations of tones are also thought to have been utilized in the ancient music of other widely separated cultures such as those of China, India, and Persia. Presumably these special intervals were discovered through independent aural experimentation or reached distant corners of the world through cultural diffusion.

Subjective consonance ratings for pairs of notes from two modern studies, one focusing on tonal clarity (Kameoka & Kuriyagawa, 1969b) and the other focusing on pleasantness (McDermott, Lehr, & Oxenham, 2010b), both found the notes closest to the Pythagorean ratios to be most consonant (figure 5.4).

The simple ratios are perceived as more consonant for two reasons: first, because they minimize the beating of partials that produces roughness (sensory dissonance), and second,

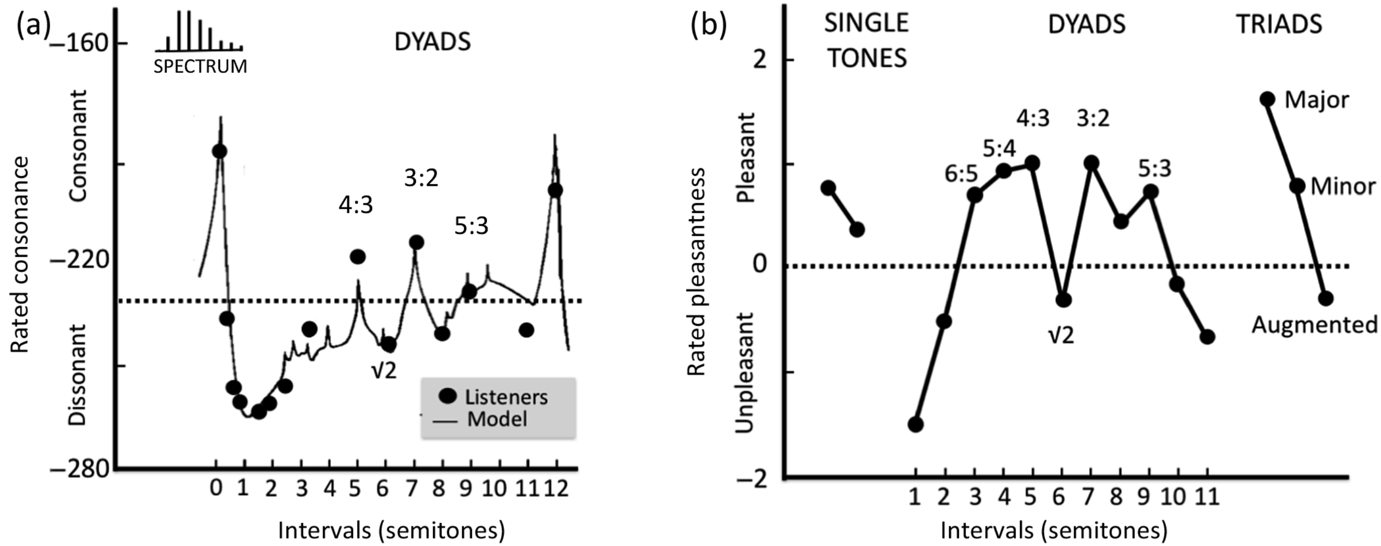

Figure 5.4

Musical consonance. (a) Filled circles: Consonance judgments (*sunda*, “clarity,” vs. *nigota*, “turbidity”) of dyads of complex tones by thirty-one Japanese audio engineers. Lines (redrawn from Kameoka & Kuriyagawa, 1969, figure 7): Predictions of consonance based on a psychophysical roughness model that is similar to the one shown in figure 5.6. Inset: stimulus line spectrum (20 dB scale). (b) Consonance judgments (“pleasantness”) for isolated notes, dyads, and triads of complex tones averaged across stimulus types and 143 college-age subjects (redrawn from McDermott, Lehr & Oxenham, 2010, figure 5S). Stimuli: All notes in both studies drawn from 12-TET scale. The two studies found prominent consonances at or near unison (1:1, 0 semitones), octave (1:2, 12 semitones), fifth (2:3, 7 semitones), fourth (3:4, 5 semitones), major sixth (3:5, 9 semitones), major third (4:5, 4 semitones), and minor third (5:6, 3 semitones).

because they minimize the amount of competition between different pitches (pitch stability). For most musical pitched sounds, these two aspects of consonance closely covary (compare patterns of roughness judgments and estimates in figures 5.4 and 5.6 with pitch stability estimates in figure 5.11 later in this chapter).

The word *consonance* can mean many different things to different people, and its conceptual meaning has changed many times over the course of its long history (Tenney, 1988; Parncutt, 2011a). Some meanings of the term and its opposite, *dissonance*, are related to hedonic or aesthetic auditory *preferences* (e.g., euphony, pleasantness), whereas others are related to attributes of *percepts* (e.g., roughness, smoothness, tonal fusion, pitch stability).

Consonance can also suggest different perceptual qualities and/or tonal preferences to different listeners. Whereas the first study (Kameoka & Kuriyagawa, 1969) focused on common patterns of consonance percepts (tonal clarity) among subjects, the second (McDermott et al., 2010b) focused on individual differences in consonance preferences (pleasantness) between them.[9] In the latter, stimuli were presented that independently manipulated parameters related to roughness (beating of partials) and harmonicity, and it was observed that preference judgments of less musically oriented listeners depended on the degree of beating between harmonics (roughness), whereas those of more musical listeners depended on harmonicity (pitch unity).

Even with differences across individual listeners and variation in the instructive terms that subjects of psychophysical experiments are given, these two studies and scores of others show that there is almost universal agreement among nonmusician listeners over which combinations of notes are more consonant than others. Most human listeners, even young infants (Trainor & Unrau, 2011; Trainor & Hannon, 2013), can readily discriminate between consonant and dissonant intervals. Neural models for consonance are discussed at length in the last section of this chapter.

The Western major diatonic scale (figure 5.3) includes all but one of these consonances, the minor third (6:5). Almost all tuning systems in use, such as just intonation, Pythagorean tuning, and equal temperament, provide either perfect or close approximations (within 1 percent) to these consonant ratios. Both Arabic-Persian and Indian scales contain notes that closely approximate the Pythagorean consonances.[10] The Persian-Arabic scale shown has twenty-four notes that are equally distributed within the octave (24-TET), such that it contains within it the 12-TET chromatic scale. Some scales having more than twelve notes, with "microtonal" steps smaller than a semitone, such as the Arabic-Persian scales and the Indian system of twenty-two srutis (distinctive pitches), similarly contain the notes of the chromatic scale plus other notes and/or frequency regions that subtly deviate from them.

However, even within the limited sample of figure 5.3, there are notable exceptions to the generalization that all tonal systems include approximations to Pythagorean consonances.

These are the Balinese gamelan scales and tunings, which appear to lack fifths (3:2) and sixths (5:3). However, gamelan music is quite diverse, and others have reported gamelan intervals close to minor thirds, fifths, and sixths (Forster, 2010; Duimelaar, 2017). The ethnomusicological literature has a spirited, ongoing discussion of gamelan tunings, whose complexity is compounded by the multiplicity of Indonesian musical styles and the varied, individualized tunings of instruments and ensembles. Some theorists have posited that the metallophones, which produce inharmonic tones, are tuned to minimize the roughness that arises from beating partials (Sethares, 2005), whereas others have asserted that the shimmering, beating patterns are a positive, sought-after aesthetic feature.

Western chromatic scales and note notations Table 5.1 gives interval names, note names, and frequency-ratios for Western chromatic (twelve-note) scales that use just intonation (JI) and equal temperament (ET) tuning. Similar tables have been presented in other reviews (Handel, 1989; Burns, 1999; Rasch, 1999). The scales span an octave range (C_4–C_5), for which absolute frequencies of the notes are given. The first column gives the musical interval name of the note, which is its ordinal scale degree (1st, 2nd, 3rd, . … 7th) in either the Western diatonic major or minor scale.

The second column gives the number of semitone steps from the tonic. The third column gives the note name in Sofége or solfeggio notation, which associates pitches with sung syllables (“solmization”) for remembering and recognizing musical intervals in musical ear training. Although the two uses coincide here, a Solfége note name can mean either a specific pitch (*fixed do* = C) or interval (*moveable do* = tonic). The next column presents the letter names of the notes, and subsequent columns show numerical values of note-ratios and absolute frequencies. The last column computes the percentage frequency differences between intervals of the two tuning systems.

The names of musical intervals associated with different frequency ratios are given in the first column of table 5.1 (see also figure 5.1). These names are derived from the scale degrees of notes in the major and minor diatonic Western scales. Thus, unison, major second (M2), major third (M3), fourth (M4), fifth (M5), major sixth (M6), and major seventh (M7) are the notes of the major scale, whereas minor second, (m2), minor third (m3), fourth (m4), fifthh (m5), minor sixth (m6), and minor seventh (m7) are notes in the minor scale.

The notes of the diatonic major scale are designated using alphabetic letters A–G. Additional sharps (#) or flats (♭) indicate other pitches in the chromatic scale that are 1 semitone (~6 percent in frequency) higher or lower, respectively, from each lettered pitch. In musical letter notation, the successive notes of the Western diatonic scale (seven notes with eight unequal divisions of an octave) are given ascending alphabetic designations A, B, C, D, E, F, and G. The successive notes of the chromatic scale (twelve divisions of the octave) are A, A#, B, C, C#, D, D#, E, F, F#, and G, G#, or alternately, A, B♭, B, C, D♭, D, E♭, E, F, G♭, G, A♭. In equal temperament tuning,

Table 5.1

Notes and intervals in the Western chromatic scale for just intonation and 12-TET equal temperament.

Interval name	Steps (re: tonic)	Solfege	Letter	Just intonation (JI)			Equal Temperament (ET)				JI vs. ET
				Ratio	Decimal	Frequency	$2^{n/m}$	Decimal	Cents	Frequency	(% Diff)
Unison	0	DO	C	1:1	1.0000	**261.63**	$2^{0/12}$	1.0000	0	**261.63**	0
Minor 2nd	1		C♯, D♭	16:15	1.0667	279.07	$2^{1/12}$	1.0595	100	277.18	-0.67
Major 2nd	2	RE	D	9:8	1.1250	294.33	$2^{2/12}$	1.1225	200	293.66	-0.23
Minor 3rd	3		D♯, E♭	6:5	1.2000	313.95	$2^{3/12}$	1.1892	300	311.13	-0.90
Major 3rd	4	MI	E	5:4	1.2500	327.03	$2^{4/12}$	1.2599	400	329.63	0.79
Fourth	5	FA	F	4:3	1.3333	348.83	$2^{5/12}$	1.3348	500	349.23	0.11
Tritone	6		F♯,G♭	45:32	1.4062	367.91	$2^{6/12}$	1.4142	600	369.99	0.57
Fifth	7	SO	G	3:2	1.5000	392.43	$2^{7/12}$	1.4983	700	392.00	-0.11
Minor 6th	8		G♯, A♭	8:5	1.6000	418.60	$2^{8/12}$	1.5874	800	415.30	-0.78
Major 6th	9	LA	A	5:3	1.6667	436.60	$2^{9/12}$	1.6818	900	**440**	0.91
Minor 7th	10		A♯, B♭	16:9	1.7778	465.11	$2^{10/12}$	1.7818	1000	466.16	0.23
Major 7th	11	TI	B	15:8	1.8750	490.55	$2^{11/12}$	1.8877	1100	493.88	0.68
Octave	12	DO	C	2:1	2.000	523.25	$2^{12/12}$	2.0000	1200	523.25	0

Notes: ET is the standard Western 12-TET scale. Frequencies given in Hz for the octave spanning C_3–C_4. The tuning system is anchored at $A_4 = 440$ Hz -> ET C3 = 261.63 Hz (anchoring frequencies in bold). Note letters indicate enharmonic equivalents in the ET scale. Additional just intonation ratios would be possible for some intervals (e.g., 9:5 for the minor 7th, which would differentiate A♯ from B♭).

A# denotes the exactly same pitch as its enharmonic equivalent B♭, but in other tuning systems, such as Pythagorean tuning, these two notes can differ very slightly (Handel, 1989).

In scientific pitch notation, pitches of the same chroma class but in different octaves are given different subscripts (e.g., A_3, A_4; see figures 5.1a and 5.2). Counterintuitively, the octave changes with the transition from B to C: for example, the diatonic ascending pitch sequence goes C_2-D_2-E_2-F_2-G_2-A_2-B_2-C_3-D_3-E_3-F_3-G_3-A_3 and so on. As the scale ascends into a new octave, the letter sequence repeats, thereby replicating the chroma circle within a new range of pitch heights.

Absolute frequencies are given for scales anchored at reference notes $A_4 = 440$ Hz (ET) and $C_3 = 261.61$ Hz (ET, JI). Because most listeners have only relative pitch, and not absolute pitch, the precise frequency of this reference does not change their perception of a scale. Large shifts, however, can move the positions of scale note-frequencies within vocal and instrumental ranges.

A standard reference frequency is obviously critical when multiple instruments with permanent fixed tunings are involved. *Concert pitch* refers to the mapping of a reference notes to note-frequencies for musical ensembles. Modern standard concert pitch sets A_4 at 440 Hz, but some individual orchestras choose different references that range from 436–445 Hz. Some early music ensembles set A_4 either a semitone lower or higher than 440 Hz.

Table 5.1 also shows frequency ratios of equally tempered intervals expressed in cents. Cents are commonly used by psychophysicists and piano tuners to describe fine frequency distinctions within a semitone as well as locations within the octave. The cents metric logarithmically divides the octave into 1,200 cents, each semitone being 100 cents. Cents (x) can be converted to frequency ratios (y) using $y = 2^{x/1200}$ and ratios (y) can in turn be converted to percent frequency differences (z) using $z = 100\ (y - 1)$. The musical interval of the tritone is 600 cents, half an octave, at its geometrical mean. The corresponding ratio of the tritone is exactly $\sqrt{2}$, often approximated by 45/32. The arithmetic mean of the octave is 3/2, at 702 cents, so the equally tempered fifth (700 cents) deviates from the just-tuned ratio by just 2 cents or 0.11 percent.

Modal and non-diatonic scales Modal scales are scales that use the same sets of notes but that have different tonics; i.e., they begin with different notes. The so-called Greek or ecclesiastical modes in Western music consist of diatonic scales. Diatonic scales, as discussed here, are seven-note scales that consist of two half-steps and five whole-steps, with the two half-steps separated by at least two whole-steps. Table 5.2 shows the seven different modes that share the note C as their tonic and the musical intervals associated with each mode. These modes share the same set of seven pitches, the white keys on the piano, and therefore the same set of inter-note intervals. But each mode has a different tone color due to the positions of scale-notes within the tonal hierarchy, i.e., the relation of the notes in the scale to the first

Table 5.2

Musical intervals associated with modern musical modes.

MODE	SCALE	INTERVALS												
		T	m2	M2	m3	M3	4	TT	5	m6	M6	m7	M7	O
Ionian *(major)*	CDEFGAB	●		●		●	●		●		●		●	●
Dorian	DEFGABC	●		●	●		●		●		●	●		●
Phrygian	EFGABCD	●	●		●		●		●	●		●		●
Lydian	FGABCDE	●		●		●		●	●		●		●	●
Mixolydian Adonai malakh	GABCDEF	●		●		●	●		●		●	●		●
Aeolian *(minor)*	ABCDEFG	●		●	●		●		●	●		●		●
Locrian	BCDEFGA	●	●		●		●	●		●		●		●

Notes: The tonic (T) and octave (O) are shaded darker gray, the major consonances (fourth, fifth, and major sixth) in lighter gray. The modes share the same note sets, but the sets of musical intervals in relation to the tonic vary with the individual mode, giving them their characteristic tonalities. Some modes have relatively more dissonant intervals and fewer consonant ones. The Ionian mode is the modern major scale, the Aeolian, the minor scale.

note, the *tonic*. Scales are anchored pitch systems, with the first note being the tonal center, the tonic. The musical interval relations between scale notes and the tonic set up a tonal hierarchy of perceptual-distances from the tonic. As the first note of the scale, the tonic establishes a tonal context for all notes that follow; if the scales have unequal step sizes, then different modes include different sets of musical intervals having different relations to the tonic. Despite having the same notes, the different orderings of semitone and whole tone steps produce different sets of frequency-ratios vis-à-vis the tonic, thereby imposing different "structural hierarchies on the set of pitches" (Dowling & Harwood, 1986, p. 116).

The seven modes thus contain different sets of musical intervals in relation to the tonic. The modes differ in their mixtures of consonant and dissonant intervals, giving them different sets of tonal possibilities. Examining table 5.2, one can see that all the modes share a common tonic (T) and octave (O). Six of the seven have the fifth, and six have the fourth, but only five have both. Three (Ionian, Dorian, Mixolydian) have the consonances of the fourth, fifth, and major sixth (M6). Five have a whole step for the second note in the scale, two (Phrygian, Locrian) have a half step instead. Two modes (Lydian, Locrian) have the tritone. The effect of these various tonal palettes consisting of different sets of musical intervals is to give characteristic colors to the melodies and harmonies that are produced using them.

The Ionian mode is the modern Western major scale, whereas the Aeolian mode is the modern minor scale. The modal system developed by music theorists in the Middle Ages was inspired by what was then known about ancient Greek scales and also by ecclesiastical modes, which were diatonic scales used for chants. The modes were arrangements of whole tone and semitone steps that spanned an octave. The various Greek and church modal systems were codified into the modern system of modes shown in table 5.2 and assigned Greek place names. Alternate modes are often used as alternatives to major scales (Ionian) or minor scales (Aeolian), and their use can be found in traditional folk, jazz ("modal jazz"), and some classical music. Many of these Western modal scales have counterparts in non-Western musical cultures (Gill & Purves, 2009), for example, in Indian thāts.

Still more scales become possible if constraints on numbers of notes and distributions of whole and half steps are relaxed. These scales are non-diatonic, and some that are commonly used in world music are presented in table 5.3. As with the modal scales above, each scale has its own characteristic tonal coloring due to the musical intervals that are available and their relations within the tonal hierarchy.

Tuning Systems

Whereas the scale includes the approximate locations within the octave of the pitches of the tonal system, *tuning systems* determine the precise ratios between fundamental frequencies (F0s) scale-notes. Tuning systems provide methods for precisely fixing the pitches of

Table 5.3

Common non-diatonic scales and their associated musical intervals.

NAME	SCALE	INTERVALS												
		T	m2	M2	m3	M3	4	TT	5	m6	M6	m7	M7	O
Major Pentatonic	CDEGAC	●		●		●			●		●			●
Minor Pentatonic	ACDEGA	●			●		●		●			●		●
Arabic Heptatonic	$\mathrm{CC^{\#}EFGG^{\#}BC}$	●	●			●	●		●	●			●	●
Chinese Pentatonic	$\mathrm{CEF^{\#}GBC}$	●				●		●	●				●	●
Chinese Whole tone	$\mathrm{CDEF^{\#}G^{\#}A^{\#}}$	●		●		●		●		●		●		●
Byzantine Double harmonic major Arabic major	$\mathrm{CDD^{\#}F^{\#}GG^{\#}BC}$	●	●			●	●		●	●			●	●
Roma ("Gypsy") Double harmonic minor	$\mathrm{CDD^{\#}F^{\#}GG^{\#}BC}$	●		●	●		●		●	●		●		●
Hungarian-Roma minor	$\mathrm{CDD^{\#}F^{\#}GG^{\#}BC}$	●		●	●			●	●	●			●	●
Ukranian Dorian	$\mathrm{DEFG^{\#}ABCD}$	●		●	●			●	●		●	●		●
Neapolitan major	$\mathrm{CC^{\#}D^{\#}FGABC}$	●	●		●		●		●		●		●	●
Neapolitan minor	$\mathrm{CC^{\#}D^{\#}FGG^{\#}BC}$	●	●		●		●		●	●			●	●
Phrygian dominant Spanish Roma	$\mathrm{EFG^{\#}ABCD}$	●	●			●	●		●	●		●		●
Blues	$\mathrm{ACDD^{\#}EGA}$	●			●		●	●	●			●		●

scale-notes, either by mathematical rules that specify ratios of note frequencies or by instrument tuning methods.

Tuning is most important for blending simultaneous notes together. Tuning is especially critical when the instruments involved have fixed pitches (e.g., frets and keyboards), the notes are concurrent and sustained, and environments have long reverberation times (e.g., pipe organs in cathedrals). Systems of exact, preset tunings are least critical in situations where musicians adjust and modulate pitch as they play, as with human voices and unfretted stringed instruments. In solo passages, intonational variability in playing melodic sequences can easily exceed the subtle differences between the notes of different tuning systems.

Consonance considerations become paramount when multiple sustained notes are sounded together. Whereas interval discrimination is relatively coarse, pitch discrimination is finer by an order of magnitude or more. Because we can hear the very low frequency beating of slightly mistuned notes, detection of mistunings between concurrent notes can be even more precise (Burns & Ward, 1982; Burns, 1999; McDermott et al., 2010a). To put these into perspective, A. J. Ellis, the English translator of Helmholtz's *On the Sensations of Tone* (1885), compared the detection of melodic mistunings relative to unison (pitch discriminations) with those of concurrent notes:

> No ear has yet succeeded in hearing the interval of 1 cent between two notes played in succession. Even the interval of 2 cents requires very favorable circumstances to perceive, although 5 may be easily heard by good ears, and 10 to 20 ought to be at once recognized by all singers and tuners. When the two notes are played at the same time, these 2 cents make a distinctive difference in consonances, and 5 cents are found to be out of tune.[11] (p. 487)

Just intonation Table 5.1 presents the frequency ratios and absolute frequencies for a just-tuned chromatic scale anchored at A4 = 440 Hz. In the just-tuned chromatic scale twelve notes that are roughly, though not exactly, equally distributed within the octave are chosen. The resulting unequally spaced intervals are constructed using ratios of small integers (*ratio* column). Decimal equivalents (*decimal* column) are shown for comparison with those of equal temperament. The intervals in the just scale contain exact approximations of the Pythagorean consonances (3:2, 4:3, 5:3, 5:4, and 6:5) as well as less consonant ratios (16:15, 9:8, 45:32, 16:9, 15:8).

Not shown in table 5.1 is a third tuning system, Pythagorean tuning, which shares many intervals with just intonation systems. The Pythagorean scale, invented ca. 200 BCE, was widely used until roughly 1600 (Handel, 1989; Guthrie, 1987; Partch, 1974; Bibby, 2003; Barbour, 2004). The scale is based on series of perfect fifths (3:2) and octaves (2:1), such that the ratios take the form ($3^n/2^m$) or their reciprocals. Thus, Pythagorean tuning preserves perfect fourths (4:3), fifths (3:2), and octaves (2:1), and it has the merit of being easy to tune by ear. However, because the octave cannot be evenly divided using these operations, some uneven sized intervals ("wolf intervals") with audible deviations ("Pythagorean commas") on the order of 1.5 percent emerge.

Pythagorean tuning also produces notable deviations from just intonation for thirds. Both Pythagorean major thirds ($3^4/2^6 = 81{:}64 = 1.2656$) and minor thirds ($2^5/3^3 = 32/27 = 1.185$) differ from their just-tuned counterparts ($5{:}4 = 1.25$ and $6{:}5 = 1.2$ respectively) by 1.2 percent, making those note combinations sound dissonant, discouraging the use of thirds in chords.

The history of tuning systems, from ancient to modern, is intricate (von Helmholtz, 1885/1954; Rasch, 1999; Barbour, 2004), frequently lending itself to polemics (Partch, 1974; Duffin, 2007). As musical practice began to use more thirds, music theorists searched for adjustments to Pythagorean tuning. The increasing use during the Renaissance of keyboard and fretted string instruments, as well as the desire to change keys (modulate) at will, drove theorists and musicians to experiment with various adjusted tuning systems, such as meantone temperament and equal temperament.

Equal temperament Equally tempered (ET) scales divide the octave equally, in logarithmic frequency-ratio terms, thereby distributing notes with exact uniformity. A scale with n such steps is designated an n-TET (n-*T*one *E*qual *T*emperament) or n-EDO (*E*qual *D*ivision of the *O*ctave) system. To divide the octave into m equal ratio-steps to compute ratios for an m-TET scale, each equal ratio step should equal the m-th root of 2, i.e., $2^{1/m}$. To compute the ratio corresponding with the n-th step in the m-TET scale, this stepsize is raised to the n-th power, such that the ratio equals $2^{n/m}$. The standard Western equal temperament chromatic scale is a 12-TET or 12-EDO system ($m = 12$). Thus in table 5.1, the seventh step in the chromatic scale, a musical interval of a fifth, yields an equal temperament ratio of $2^{7/12} = 1.4983$, which is quite close to the just intonation ratio of $3/2 = 1.5000$ ($\Delta = 0.11\%$).

Twelve-tone equal temperament (12-TET) appears to be an optimal means of dividing the octave using a small number of divisions in terms of closely approximating the Pythagorean consonances (see discussion in "Design Principles" section below and figure 5.5). For most notes in the scale, these tunings are fairly close approximations to each other (last column in table 5.1), the smallest differences being for fifths and fourths ($\Delta \sim 0.1\%$), and the largest ones for minor 3rds and major 6ths ($\Delta \sim 0.9\%$). The differences between JI and ET tunings are therefore fairly subtle, less than 1 percent (1/6 a semitone, ~16 cents) for all intervals. For the most consonant intervals (fourths and fifths) these deviations are much smaller, on the order of 0.11 percent, rendering the differences almost completely inaudible. Audible differences between these systems are least apparent in melodic contexts, in which notes are played sequentially, and most apparent when sustained multiple notes (chords) are sounded at the same time.

Equal divisions of the octave appear to have been used in ancient Chinese music and in the music theory of Aristoxenus. Twelve-tone equal temperament (12-TET) systems were proposed and used in Western music beginning in the early sixteenth century (by Henricus Grammateus, Giacomo Gorzanis, and Vincenzo Galileo), but it was not until much later that these systems were precisely described in numerical terms, first in 1584 by Zhu Zaiyu (朱載堉, Ju Tzayyuh, Chu-Tsaiyu), a prince of the Ming court in China (Picken, 1957; Barbour, 2004), then by Simon Stevin in Flanders, circa 1605 (Bibby, 2003) and finally with Mersenne's codification of the theory in his *Harmonie Universelle* in 1637 (Rasch, 1999; Barbour, 2004).

This standardization of scale step-size has the practical effect of making frequency ratios between notes in the scale independent of their relation to a particular tonic, which in turn permits free writing in and modulation to multiple keys (tonics) without subtle key-dependent alterations of frequency ratios.

Design Principles

Why do our scales have the structure that they do—that is, why these notes and not others? The reasons divide into explanations based on universal biological constraints, such as the structure and function of human and animal auditory systems, and those based on specific cultural practices. The two kinds of explanations are complementary. Those grounded in auditory physiology necessarily focus on features that are common to all scales, whereas those grounded in ethnomusicology necessarily focus on cultural differences and the specific social factors that might explain them.

Although the acoustics and the auditory system provide some constraints on how different combinations of musical notes will sound, the aesthetic choices of which specific combinations of pitches and timbres will be utilized often depend heavily on common practices and meanings of a musical culture. The particular scales adopted by a musical culture reflect a mixture of auditory constraints, cultural conventions, aesthetic ends and preferences, and in some cases the musical instrument technologies that are available.

Which notes and note frequencies are chosen for a given type of music is an aesthetic decision whose purpose is to bring about desired effects on listeners (the fine arts as aesthetic engineering). Music theory lays out a general set of design principles and provisional, prescriptive rules for how to achieve specific perceptual, cognitive, hedonic, and aesthetic ends within given cultural or subcultural contexts.

Scales determine the framework of pitches that are utilized in a given practice of tonal music. They can be considered in terms of specifying the primitive categories of tonal grammars (Handel, 1989). Perceptual and cognitive constraints on scale construction and tuning have been proposed (Dowling & Harwood, 1986, pp. 92–95; Burns & Ward, 1982; Burns, 1999); these are included in this discussion of basic design principles (constraints).

Scale design choices include continuous versus discrete sets of pitches, numbers of pitches, F0-range, scale repetition span (e.g., octave, tritave), and placement of notes within the repetition span. Tuning constraints involve more precise specifications of either pitch frequencies or instrument tuning methods. These are soft, defeasible constraints, such that there can be tradeoffs between scales optimized for one or another set of musical goals. Scale structure determines in part how "good" these scale notes and note combinations sound (in terms of whatever aesthetic criteria are adopted), the expressive range of the pitch and interval set, how easily specified tunings can be implemented, and how easily various musical modulations can be achieved. As with any other tool, there are no objectively "perfect" scales, only those that work better or worse for particular musical ends in particular contexts.

Most tonal systems consist of discrete pitches, with seven or fewer unequally spaced pitches per octave consisting entirely of "small intervals" of a fifth or less (Savage et al., 2015):

1. *Discrete versus continuous pitches*. Discrete musical scales provide a common framework for coordination of pitch, much as meter provides a framework for coordination in time. A discrete set of common pitches enables groups of musicians to play together with a minimum of pitch mismatches or beating (Burns, 1999).
2. *Numbers of pitches*. Burns (1999) has theorized that the relatively small number of discrete pitch categories might make melodies easier to distinguish and remember, just as the small number of discrete phonetic classes might make speech easier to understand.
3. *Pitch range*. All scale notes should lie within the existence region of musical tonality and encompass the pitch ranges of the musical instruments and voices of the musical genre.
4. *Octave organization*. Strong octave equivalence mandates that the scale should subdivide each octave and repeat, such that corresponding notes in each octave have chroma equivalents in other registers. The experimental Bohlen-Pierce scale subdivides the tritave (3:1), repeating at tritave intervals. It has just intonation and equal temperament (typically 13-TET) versions.
5. *Consonant and dissonant intervals*. Inclusion of both consonant and dissonant intervals in the scale allows for a range of consonance contrasts that can create tension-relaxation dynamics and emotional meaning.
6. *Equal versus unequal scale steps*. Equal temperament tuning systems divide the octave into equal frequency steps. Unequal scale steps that arise from just or Pythagorean tunings create noticeable tone colorations—due to small distortions of intervals—when playing music in keys other than the one that was used to tune the instrument. Equal temperament permits keyboard instruments to play in any key without retuning, ensuring intervallic uniformity amongst key modulations.
7. *Ease of tuning*. Some tuning systems are more difficult to implement than others, and this can depend critically on the types of instruments and tuning procedures that are used (fixed vs. variable pitched, harmonic vs. inharmonic instruments, electronic vs. acoustic,

digital vs. analog, tuning by ear vs. pitch analyzer). Just-intonation tuning by ear can be achieved simply by minimizing audible beats and roughness, whereas equal temperament is considerably more complex. When piano tuners equally temper a piano by ear, they listen for specific beat rates between particular notes and systematically adjust string tensions until the tunings converge on a set of uniformly spaced notes.

8. *Tuning accuracy of consonant intervals.* If smoothness (minimal roughness) and blending (maximal pitch fusion) are desired, especially for groups of concurrent, sustained notes, then tuning systems that provide nearest approximates to just intervals best achieve that end (Duffin, 2007; Partch, 1974). As roughness and harmonicity preferences prescribe, consonance is maximized when frequency ratios of concurrent note pairs approach those of small integers, i.e., for just tuning. Small deviations from just ratios are most noticeable and/or objectionable for consonant intervals: unisons, octaves, fifths, fourths, and major sixths. Such deviations are less noticeable for dissonant intervals, such as tritones, seconds, minor sixths, and sevenths, possibly because of the greater amount of roughness already present and the shallower roughness gradients at those points (figure 5.6d). Note frequencies should therefore ideally provide reasonably close (<< 1 percent) approximations to the just-tuned ratios of the major consonant intervals. Notably, although twelve-tone equal temperament gives close approximations for fifths (–0.11 percent mistuning) and fourths (+0.11 percent), it mistunes minor and major thirds, as well as major sixths by almost 1 percent (m3: flat 0.90 percent, M3: sharp 0.79 percent; M6: flat 0.91 percent). Audible differences between just intonation and equal temperament are quite subtle for nonmusicians, but can much more apparent for musicians with highly trained ears.
9. *Number of equal subdivisions of the octave.* Equal temperament systems have one scale parameter: How many divisions of the octave? But why a twelve-fold division? In equal tempered scales, the number of subdivisions (n-TET) chosen should provide reasonable approximations to at least a few consonant ratios.

From the perspective of Western 12-TET scales, systems with more than twelve tones are often called "microtonal" because their steps are smaller than a semitone. Although most world music at this point in history is written using a 12-TET tuning system, Arabic music uses a 24-TET system and Indian music uses a twenty-two-tone system, both of which include exact or close approximations to all of the Western 12-TET notes (figure 5.1). A number of 12-TET inclusive systems are mixtures of equal temperament intervals plus just intonation intervals for consonant ratios (see, e.g., discussion of Persian scales in Helmholtz, 1885/1954). Burns (1999) notes that "three of the major non-Western musical systems (Indian, Chinese, and Arab-Persian) have inclusive scales approximately equivalent to the Western chromatic scale, and, hence, have the same propensity for the perfect consonances (octaves, fourths, and fifths)."

Music has been written for many different equal temperament systems, and there is a considerable literature, historical and contemporary, that explores the manifold musical possibilities of these alternative systems, past, present, and future (Sethares, 2005).

Different equal frequency ratio subdivisions of the octave yield note tunings that give better (closer) or worse (more distant) approximations of interval frequency ratios associated with consonance maxima (see figures 5.4 and 5.6). Note-frequency ratios and the mean deviation ("mistuning") between just and ET systems ("tuning error") for major Pythagorean consonances (4th, 5th, and major 6th) are plotted in figure 5.5 as a function of the number of equal ratio divsions of the octave (n-TET).

Considering n-TET systems up to eighteen notes per octave, the 12-TET, 7-TET, and 5-TET systems provide the best approximations to the major consonances (figure 5.6b), with 12-TET

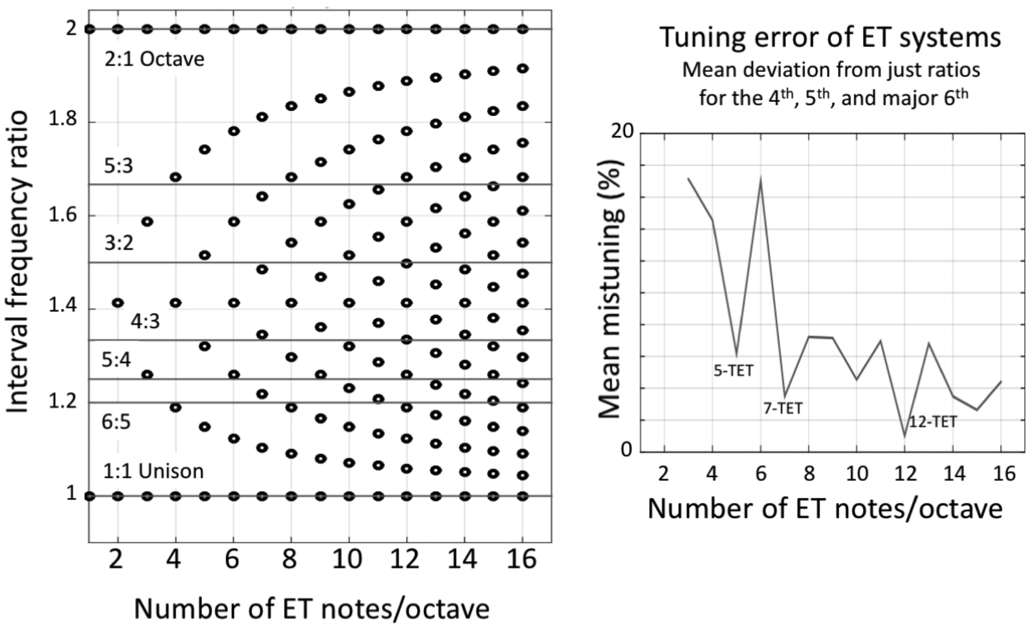

Figure 5.5

Equal temperament approximations to major consonances in the octave. Left. Comparisons of musical interval ratios for different equal logarithmic divisions of the octave (2–16 TET systems). Circles mark ET intervals associated with scale notes; lines indicate ratios of prominent consonant intervals (fifth (3:2), fourth (4:3), major sixth (5:3), major third (5:4), minor third (6:5) (see Sethares, 2005, p. 58). Right. Average tuning error to consonances (fifth, fourth, and major sixth) as a function of the number of equal divisions in the octave (NTET), indicating how well the ET system approximates these consonances. Prominent relative optima are seen for the Western 12-TET system, the ET 7-TET heptatonic scale, and the ET 5-TET pentatonic scale.

the closest. The 12-TET system is currently the most widely used worldwide, but 5-TET and 7-TET systems were used in ancient China and are common in existing world musics (Burns, 1999; Justus & Hutsler, 2005). These include 5-TET xylophones from Thailand and Uganda. Although aesthetic and practical criteria for tuning Indonesian gamelans varies widely, 5-TET and 7-TET approximations can be found, with slendro scales resembling a 5-TET system (Sethares, 2005). Thus equal temperament is not a peculiarly Western invention.

Auditory Neural Models

The psychology of music has engaged in longstanding questions of nature versus nurture and the origins of our affinity with music. These questions involve which aspects of music perception, cognition, and preference are determined by relatively fixed, near universal mathematical-physical, biological, neural-psychological constraints, on the one hand, and which are determined by culture-dependent experientially mediated developmental and statistical learning processes, on the other. Various origin questions involve which aspects of our relations to music are due to extrinsic, music-specific directed processes of evolutionary natural selection, and which are due to intrinsic structural-functional properties of nervous systems.

Temporal codes are possible structural features of nervous systems that could make them amenable to modulation via the temporal patterns of stimulation. This would make music an effective stimulus for modulating many different types of internal psychological states, and the auditory system a particularly effective modality for impressing temporal patterns on neural populations.

Since the Pythagoreans, the pervasive role of frequency-ratios in intervals, scales, and tunings has led Platonically inclined theorists to attribute harmonic structure in music to mathematical order embedded in the natural world. Cognitivist theories tend to attribute this structure to acquired knowledge of musical conventions through repeated exposure—schemas acquired through enculturation.

Bayesian theories attribute this structure to experience and associative learning (Temperley, 2007). Some Bayesian theories attribute scale design to ratios present in the resolved harmonics of voices (Schwartz, Howe, & Purves, 2003; Schwartz & Purves, 2004; Gill & Purves, 2009). However, theories that ascribe the origins of pitch, consonance, and harmony entirely to learned associations run into difficulties. Those that rely on exposure to human voices have difficulty explaining the existence region of musical tonality, which extends well beyond the F0-range of human voices, spoken and sung, by 1–2 octaves in each direction. Extended learning periods also do not appear to be required for music perception, as infants at very early ages are already predisposed to making consonance/dissonance distinctions and recognize melodies, well before the effects of enculturation are seen (Trainor & Unrau, 2011; Trainor & Hannon, 2013). Clearly, statistical learning of human voice patterns and musical cultural norms play some role in influencing tonal expectancies and preferences, but it appears less likely that they play essential, formative roles in enabling more basic aspects of music perception. Instead, the statistical expectancies of voices, music, and environmental sounds gleaned from short-term exposures (Loui, 2012) to longer-term enculturations may depend critically on prior perceptual organization of sensory information by the auditory system. Auditory theories tend to attribute structure in music to the structure of neural representations of sound and to bottom-up automatic grouping mechanisms. Because statistical and structural explanations typically address different aspects of music perception, they are not usually mutually exclusive. A full account of music perception and cognition will likely require incorporation of both types of causes.

This section presents two auditory models most directly related to intervals, scales, and tunings: a psychophysically based model for roughness, and a temporal neural model for musical pitch, chroma relations, consonance, and harmony.

Models for Roughness

Roughness is an aspect of consonance that arises from beatings of nearby harmonics in the cochlea. This theory was originally proposed by Hermann von Helmholtz in his landmark 1863 book on acoustics, music, and hearing, *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (Helmholtz, 1885/1954) and refined by Plomp and Levelt in the 1960s (Plomp & Levelt, 1965). Until very recently, roughness has dominated discussions of consonance (Pierce, 1992; Sethares, 2005). The literature on roughness as it relates to consonance is complex, and there are many other detailed criticisms that can be made of general theories, specific models, and psychoacoustic methods employed (Tramo, Cariani, Delgutte, & Braida, 2001, Machinter, 2006).

What are the neural correlates of roughness? Figure 5.6 illustrates how the neural activity patterns associated with roughness are thought to be generated. A sensation of roughness is created when pairs of pure tones close together in frequency are presented simultaneously to the same ear. When two musical notes are sounded, harmonics from the two notes interact (figures 1.4 and 5.6a), and beating patterns at the difference frequencies (Δf) of nearby pairs of harmonics are produced. The sets of beating patterns change as a function of musical interval (F0-ratio). If the two frequencies lie within a critical band, i.e., less than ~20 percent apart, they interfere (beat), creating low-frequency oscillations in amplitude (envelope modulations) at their beat frequency Δf. The sensation of roughness is related to the perception of this beating in the 20–120 Hz range. If the tones are almost identical in frequency ($\Delta f < 1\%$), they fuse, creating little or no roughness sensation (figure 5.6c). However, if the tones are

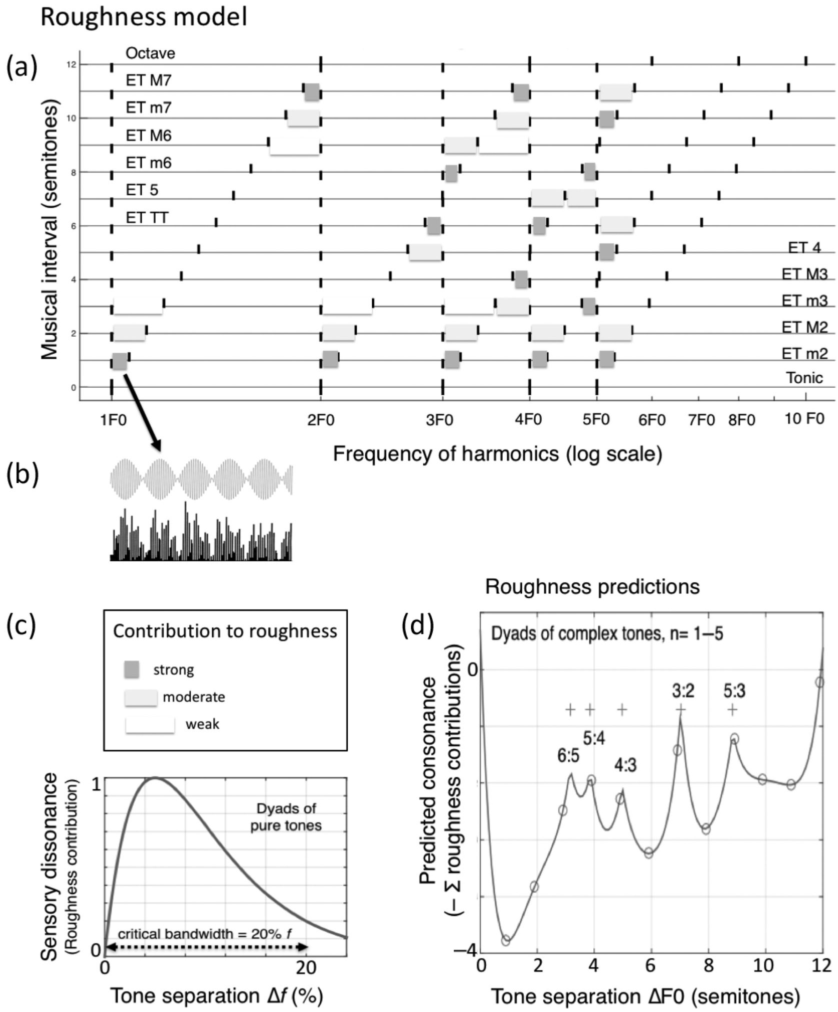

Figure 5.6

A psychophysical model of roughness. The model estimates degree of perceived roughness from frequency proximities of beating harmonics. (a) Interactions between harmonics for dyads of complex tones consisting of harmonics 1–5 separated by different equally tempered musical intervals. The schematic indicates which pairs of harmonics would be expected to produce weak, moderate, or strong contributions to total perceived roughness. (b) Beating between the first two harmonics of a minor second (440 and 469 Hz) and the modulated temporal firing pattern of an auditory nerve fiber tuned to this frequency region (Tramo et al., 2001). See also figure 1.4. (c) Roughness contribution as a function of harmonic separation in terms of fraction of critical bandwidth (CB = 20% *f*). Function shown is Parncutt's approximation to Plomp & Levelt's dissonance factor *g* (Machinter, 2006). Roughness contributions of all pairs of harmonics are then summed to estimate total dissonance (–consonance). (d) Predicted consonance ratings based on roughness summation. Scale is inverted to show consonance. Note consonance peaks for fifths (3:2), fourths (4:3), major 6ths (5:3), major thirds (5:4), and minor thirds (6:5). Circles indicate ET musical intervals; crosses, JI ratios.

slightly farther apart, from 2 to 12 percent, they create strong sensations of roughness. Thus two instruments playing sustained notes slightly out of tune with each other (1–2 percent) can produce noticeably rough sensations.

When two harmonic complex tones are presented together, the partials present consist of integer multiples of the two respective fundamentals ($F0_1$, $F0_2$). Depending on the frequency ratios of the two fundamentals, i.e., the musical interval, and the frequency selectivity (bandwidth) of cochlear filtering, different partials of the two notes that are near each other in frequency beat, causing sensations of roughness.

In his *On the Sensations of Tone as a Physiological Basis for the Theory of Music* Helmholtz outlined a resonance theory of cochlear function and proposed a model of consonance based on the aggregate amount of harmonic interaction in the cochlea. Helmholtz noted that simple F0 frequency ratios of musical tones minimize this interaction, and hence they minimize the sensation of roughness. Further, he noted that equal temperament provides the best all-around approximation to these ratios (best for fifths and fourths, worst for thirds), such that roughness is minimized (Helmholtz, 1885/1954, pp. 312–315).

Auditory masking experiments in the early twentieth century by Harvey Fletcher and colleagues revealed that these interactions are nonlinear; that is, the masking of one pure tone by another drops precipitously when the frequencies of the two tones are separated by more than a “critical bandwidth” (rule of thumb: $\Delta f > 100$ Hz for $f < 500$ Hz, $\Delta f > 20\%$ for $f > 500$ Hz). Following the introduction of critical bands into auditory theory, Helmholtz’s theory was revised by two groups (Plomp & Levelt, 1965; Kameoka & Kuriyagawa,1969a, 1969b) to take into account the roughness contributions. Both groups conducted psychoacoustic experiments that estimated degree of roughness produced by pairs of pure tones as a function of amplitude and frequency separation, in terms of fraction of critical bandwidth. The contributions of all pairs of partials were summed together to estimate the roughness produced by complex tones (the solid curve in figure 5.4a).

To illustrate the operation of these models, a simplified psychophysical model of roughness has been implemented using Parncutt’s approximation to this function (figure 5.6c). Pairs of partials closest to ~4 percent in frequency cause the most roughness, and the roughness caused by the whole complex tone dyad is well predicted by the sum of roughness contributions produced by individual pairs of nearby partials (Kameoka & Kuriyagawa, 1969a, 1969b). Thus, greater degrees of perceived roughness are produced when there are more pairs of beating partials and the beating partials are closer together.

The psychophysically based roughness theory gives a precise and plausible account of why scales might incorporate close approximations to the integer ratios of the most prominent Pythagorean consonances. When tuning musical instruments by ear, one can very accurately zero in on just-tuned intervals (octaves, fifths, fourths) simply by minimizing the degree of beating (roughness) that is heard. For equal temperament, many of the locations of scale pitches in the octave also minimize roughness.

In order to ground psychophysically based roughness models in terms of neural substrates, it is necessary to examine how auditory neurons respond to combinations of harmonics. Over two decades ago, we carried out a systematic study of the neural correlates of consonance in the auditory nerve of anesthetized cats (Tramo et al., 2001). Eight stimuli were used: four pure tone dyads and four complex tone dyads (harmonics 1–6) separated by four musical intervals (minor second [16:15], fourth [4:3], tritone [45:32], and fifth [3:2]). We found strong neural correlates both for roughness and for pitch fusion/stability that replicated the rank ordering of perceptual consonance judgments for both pure and complex dyads.

When partials are close together, they beat, causing low-frequency temporal modulations of firing in auditory nerve fibers whose characteristic frequencies are near those of the beating partials (figure 5.6b). For individual harmonic tones, the beating does not interfere with F0-pitch because all beats are at the difference frequency Δf, which is the fundamental F0. In the case of two harmonically unrelated complex tones, however, the beat rates between different pairs of nearby interacting harmonics are all different and unrelated to the note F0s, such that the temporal beating patterns clash with each other and interfere with all other pitch-related spike periodicities that are related to individual harmonics, fundamentals, and fundamental basses.

As in the psychophysical models, perceptual roughness can be estimated from neural auditory nerve fiber (ANF) responses by quantifying the amounts of low-frequency modulation of discharge rate (10–120 Hz) in different neural frequency regions and summing them together to form a physiologically based estimate of roughness. The relative rankings of these physiological roughness estimates, in auditory nerve (Tramo et al., 2001), brainstem (Bidelman & Krishnan, 2009, 2011), midbrain (McKinney, 2001; McKinney, Tramo, & Delgutte, 2001), and cortex (Fishman et al., 2001) all correlate highly with those from human psychophysical consonance experiments.

As attractive as this theory of beats may seem, it has many general shortcomings. First, roughness cannot account for perceived consonant and dissonant relations between successive pitches (e.g., arpeggios, melodies), because roughness is produced by interactions of concurrent partials in the cochlea (tones that do not temporally overlap do not beat in the cochlea). A short-term auditory memory mechanism is needed to account for melodic consonances. Second, dissonance can be produced without beating partials (Lipps, 1905/1995; Révész, 1954/2001). All nearby, interfering partials can be selectively removed from the tones, but the dyads can still sound dissonant because of their inharmonicity. Inharmonic complex tones—that is, complex tones in which the frequencies of at least some or all partials are not part of the harmonic series (nF0 for n = 1, 2, 3, …)—sound less consonant than their harmonic counterparts, even in the absence of beating partials. Finally, roughness is a perceptual quality distinct from pitch, and as such does not appear to be directly involved per se in musical harmony.

Many theorists, including Helmholtz, recognized these limitations and accordingly proposed dual aspect models of consonance that include both roughness (*sensory dissonance*) and musical consonance. As a proto-Gestaltist alternative to Helmholtz's theory of beating harmonics, a theory of tonal fusion was proposed in the late nineteenth century by Carl Stumpf in his *Tonpsychologie* (Stumpf, 1883/1890; Lipps, 1905/1995; Boring, 1942; Révész, 1954/2001; DeWitt & Crowder, 1987; Schneider, 1997). Modern conceptions of consonance related to these ideas stem from models of musical pitch and competition between multiple reinforcing or competing pitches (pitch multiplicity, pitch stability). Pairs of tones separated by consonant intervals tend to fuse together and evoke one pitch related to their common fundamental, the *fundamental bass*, whereas dissonant intervals produce more distinct competing pitches (pitch multiplicity).

Pitch multiplicity models are based on F0-pitches associated with common subharmonics of all harmonics present, be they individual notes or combinations of notes. In recent decades Terhardt's theory has been the most prominent of these (Terhardt, 1974, 1977, 1984; Parncutt, 1989), but time-domain models based on temporal coding of sounds in the auditory nerve have also been proposed (Tramo et al., 2001; Cariani, 2004a). A temporal neural pitch multiplicity model based on interspike interval representations is presented below.

Temporal Models for Musical Pitch

Temporal theories of consonance and harmony stem from neural representations for musical pitch that are based on temporal coding in the auditory system. This section introduces temporal codes, outlines a temporal theory for musical pitch and pitch multiplicity, shows how the structure of interspike interval representations mirror the spiral structure of chroma relations and scales, and applies the temporal model of pitch multiplicity to consonance (harmonicity) and harmony (pitch stability of chords).

Neural coding is a fundamental problem in neuroscience. For over 150 years, auditory scientists have been debating the nature of the neural coding of sounds in the auditory system (Boring, 1942). The neural coding problem for audition entails identifying which aspects of neural activity subserve particular auditory functions (Cariani, 1999; Cariani & Micheyl, 2012)—that is, what constitute "the signals of the system." The debate has revolved around two complementary types of neural codes, channel codes and temporal codes.

Channel codes rely on patterns of responding neurons (neural channels) to convey informational distinctions. In rate-channel coding, distinctions are conveyed by which neurons fire how often (at which average rates). In the auditory system, rate-channel codes have historically been called "rate-place" codes. These are based on spatial profiles of neural firing rates either as a function of their cochlear place of innervation or location within a neural cochleotopic, tonotopic map. Cochlear place determines the frequency tunings of each of the 50,000 ANFs that constitute the human auditory nerve. By virtue of place-dependent cochlear tunings, the rate-place profile of the population provides a coarse, highly nonlinear representation of the running power spectrum of the acoustic stimulus that changes dramatically with sound level.

In contrast, *temporal codes* rely on temporal patterns of spikes to convey informational distinctions. A simple temporal pattern code is an interspike interval code, in which time durations between spikes convey information. Temporal codes can be found in nearly every sensory modality (Perkell & Bullock, 1968; Cariani, 2001b).

Temporal codes have a long history within auditory psychophysics and physiology (Boring, 1942; de Cheveigné, 2005; Moore, 2013). The main advantages of a temporal theory of hearing stem from the precise, invariant, and robust character of temporal patterns of spikes. Interspike interval distributions are level-invariant in a manner that parallels the precision and stability of pitch perception.

Temporal coding is immediately apparent in neuronal firing patterns at the level of the auditory nerve, illustrated in figure 5.7 (Cariani, 1999). The spike train data was recorded from auditory nerve fibers in an anesthetized cat (Cariani & Delgutte, 1996a, 1996b). The sound (figure 5.7a) is a single format vowel. Histograms showing spike timing patterns of ANFs with different characteristic frequencies (CFs: 0.2–10 kHz) are shown in response to 100 stimulus presentations. The spike timing patterns closely follow the time structure of the positive part of the waveform after it has been filtered by the cochlea.

Thus, in the auditory nerve, spikes are correlated with the stimulus waveform, such that the periodicities in the waveform are impressed on the temporal patternings of spikes. Because of this correlation, also known as "phase-locking," patterns of time durations between spikes—interspike intervals—reflect periodicities in the stimulus. First-order interspike intervals are time durations between consecutive spikes, whereas all-order intervals include those between both consecutive and nonconsecutive spikes. Although models based on first-order intervals successfully predict acuity of pitch discrimination for pure tones as a function of frequency, level, and duration (Siebert, 1968; Goldstein & Srulovicz, 1977; Heinz, Colburn, & Carney, 2001), all-order interval distributions, which include first-order intervals, can account for wider ranges of pitch phenomena.

Global temporal pitch models combine all-order interspike intervals from all CF-regions of the auditory nerve to form a temporal population-based auditory representation. These population-intervals or summary autocorrelations based on all-order intervals among auditory populations yield accurate, precise, and robust predictions for a wide range of F0-pitches (Meddis & Hewitt, 1992; Slaney & Lyon, 1993; Cariani & Delgutte, 1996a; Lyon & Shamma,

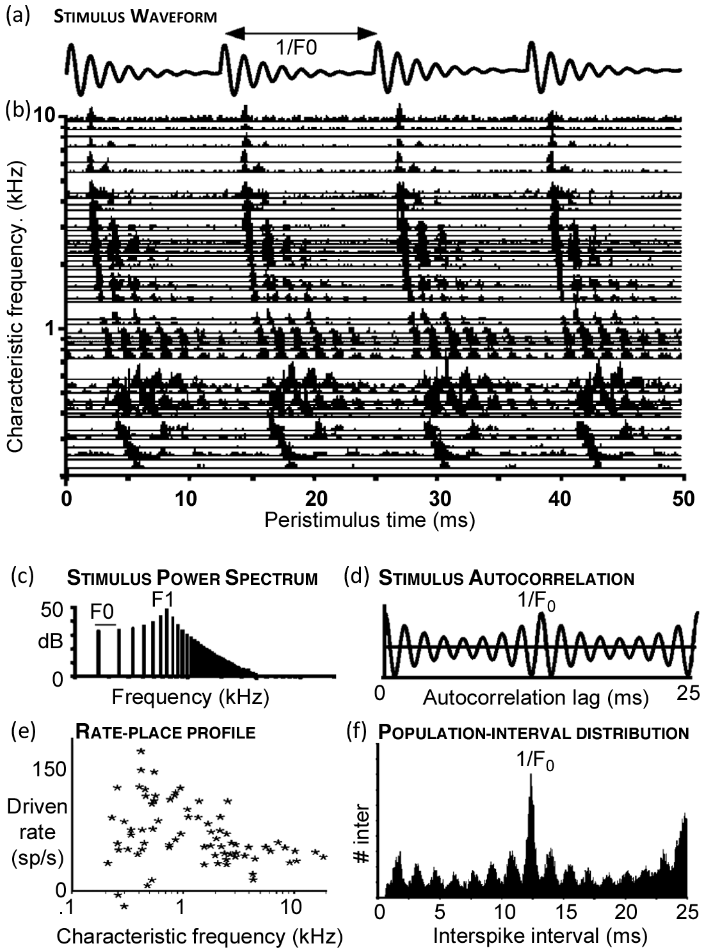

Figure 5.7

Temporal coding of pitch in the auditory nerve. Auditory nerve fiber (ANF) responses to a harmonic complex (single formant vowel, F0 = 80 Hz, pitch period 1/F0 = 12.5 ms, 100 presentations at 60 dB SPL). (a) Stimulus waveform. (b) Peristimulus time histograms of different cat ANFs as a function of characteristic frequency (baseline value). (c) Stimulus power spectrum. (d) Stimulus autocorrelation function. (e) Stimulus-driven rate-place profile of ANFs; i.e., firing rate—spontaneous rate. (f) Population-interval distribution (PID) formed by summing all-order intervals from all recorded fibers (Cariani, 1999). Data from Cariani & Delgutte (1996a).

1996; Cariani & Delgutte, 1996b; Meddis & O'Mard, 1997; Cariani, 1999). These models are descendants of earlier neural temporal hypotheses for pitch (Licklider, 1951, 1959; Moore, 1980; van Noorden, 1982) based on mass statistics of interspike intervals. Currently, global temporal models that use neurophysiologically realistic neuronal responses as inputs provide the strongest, most comprehensive predictions of musical pitches.[12]

In our studies of the neural correlates of pitch in the auditory nerve (Cariani & Delgutte, 1996a, 1996b), all-order interspike interval distributions of individual auditory nerve fibers of all characteristic frequencies were summed together into *population-interval distributions* (PIDs, 6F, a.k.a. *summary autocorrelations*, SACFs). The durations associated with the highest interval peaks in the PIDs predict the F0-period of the stimulus with high accuracy, precision, and robustness. The theory successfully predicts the pitches heard from the neural data over a wide range of F0s and stimulus types, harmonic and inharmonic, periodic and quasi-periodic.

Pitch strength (salience) is qualitatively predicted by the relative height (peak-background ratio) of pitch-related peaks. Finally, the PIDs resemble positive portions of the stimulus autocorrelation functions (ACFs) (compare figure 5.6d with 5.6f), such that these temporal population-based representations can serve as general purpose neural representations of the stimulus power spectrum, up to the frequency limits of useable phase-locking, roughly 4–5 kHz. Other studies have shown that this purely temporal, global interval information is sufficient for representing multi-formant vowels (Palmer, 1992), and therefore also those aspects of musical timbre that depend on low-frequency spectrum.

Global temporal pitch models can predict virtually all pitches produced in tonal musical contexts, which are invariably harmonic and near-harmonic complex tones with F0-periodicities below ~4 kHz. They also predict pitches for other classes of stimuli that can carry a melody (support chroma relations): low-frequency pure tones ($f < \sim 4$ kHz), complex tones with high, unresolved harmonics, inharmonic complex tones (as produced by bells, lithophones, and metallophones), amplitude modulated noise ("nonspectral pitch" [Burns & Viemeister, 1976, 1981]), repetition noise, and spectral edge pitches. With additional assumptions regarding binaural cross-correlation and cancellation operations, population-interval representations can also plausibly account for binaurally created pitches (e.g., the Huggins pitch) that can also support musical melody recognition. The only chroma-supporting pitches that these peripheral temporal representations clearly cannot explain are Zwicker tone auditory afterimages (Gockel & Carlyon, 2016), which likely have a more central origin. It is not yet clear whether Zwicker tones above 4 kHz can support chroma relations.

Some psychophysicists adopt a "two-mechanism" model for F0-pitch (see chapter 1 in this volume), using a strong spectral pattern mechanism for pitches of resolved harmonics and a weak temporal pattern mechanism for unresolved harmonics (Carlyon & Shackleton, 1994; de Cheveigne, 2010). Pitches of unresolved harmonics produce significantly weaker F0-pitches with much coarser discrimination thresholds. However, despite their lower saliences, unresolved harmonics can nevertheless support musical tonality (interval and melodic recognition). Neurons that are selective for pure tone frequencies and the corresponding F0s of both resolved and unresolved harmonics have also been found at the cortical level (Bendor & Wang, 2005), which suggests that they may be tuned to common incoming temporal patterns of spikes that these stimuli share, rather than integrating different types of neural information arising from entirely separate pitch mechanisms.

There are some known functional dissociations between the two classes of pitch. Some amusic listeners, who cannot make musical interval judgments if the note-harmonics are perceptually resolved (i.e., the kinds of stimuli that predominate in tonal musical contexts), can nevertheless make such distinctions for F0-pitches of higher-numbered, unresolved harmonics (Cousineau, Oxenham, & Peretz, 2015).[13]

Neural Basis of Chroma Relations

Temporal codes provide a possible means of explaining the ubiquity and importance of frequency-ratios in music perception (Burns, 1999). Simple ratio theories are usually mentioned in discussions of theories of consonance, and often dismissed, fairly or not, as unphysiological theories or Pythagorean-Platonic numerological fantasies (e.g., Révész, 1954/2001; Plomp & Levelt, 1965; Sethares, 2005; Bowling & Purves, 2015). However, reasonable explanations for frequency ratios that are firmly grounded in auditory neurophysiology are possible.

A number of auditory neurophysiologists and theorists have suggested that chroma relations might have a basis n temporal coding. Modern perspectives couched in terms of temporal codes are essentially neural versions of Galileo's observations concerning the regularities of sounds with simple ratios.[14] Centuries later, Licklider's (1951) temporal autocorrelation model provided a framework that could explain a wide range of auditory pitch phenomena, including the role of simple frequency ratios, through the interactions of common periodicities.[15]

Along similar lines, the auditory neurophysiologist Jerzy Rose, who carried out early investigations on the temporal discharge patterns of auditory nerve fibers, related frequency-ratios to temporal cadences of neural discharge.[16] Roy Patterson (1986) observed that the temporal firing patterns of auditory nerve fibers have a repeating, spiral structure that mirrors that of pitch space and musical scales (see figure 5.10 and the related discussion below).

Perhaps the most ambitious attempt to explain frequency ratios in terms of neural temporal processing was the "long pattern hypothesis" for pitch, harmony, and rhythm of Boomsliter and Creel (1961, 1963, 1970). Their theory was inspired by music perception, Licklider's temporal theory, and the ubiquity of phase-locking in the auditory nerve to both pitch-related periodicities and rhythmic patterns. As with pitch, albeit on slower time scales, strong arguments can be made for temporal coding of rhythm (Cariani, 2002). With their

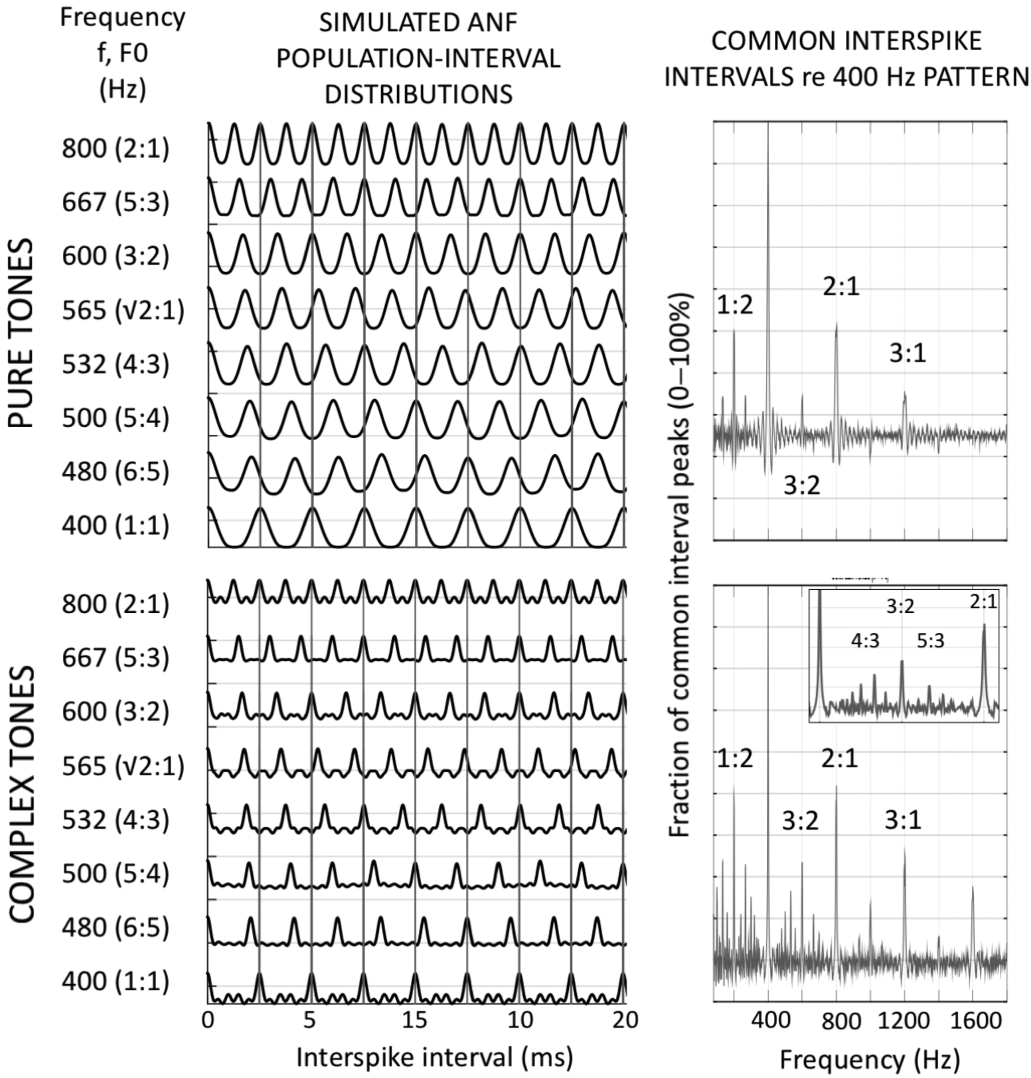

Figure 5.8

Octave similarity and interspike interval representations. Left: Schematic of simulated auditory nerve fiber (ANF) population-interval distributions (PIDs) produced in response to pure tones and complex tones (n = 1–6). The peaks in the PIDs are located at subharmonics (1/*f*, 1/F0) of the pure tone frequencies *f* and complex tone fundamentals F0. Vertical lines indicate peak positions for f, F0 = 400 Hz. B. Fraction of common interval peaks of half-wave rectified ACFs, 80–1800 Hz, 1 Hz steps, lags = 0–40 ms. Note that tones an octave apart (2:1, 1:2) share half their interval peaks; those a twelfth apart (3:1, 1:3) share a third. The inset plot shows the fine structure of common intervals within the octave. The simulation used the Zilany, Bruce, and Carney model (2014). See text notes for parameters.

"harmony wheel," Boomsliter and Creel (1963) graphically show the temporal similarities that exist between periodic patterns related by simple ratios, which was developed into a theory of melody based on sequences of ratios. The theory postulated networks of reverberating delay-loops in the brain that could propagate temporal patterns, such that simple ratios would be self-reinforcing: "The brain exhibits the properties of an apparatus that works by neural mechanisms of temporal recurrence" (Creel & Boomsliter, 1970). Time-domain neural information processing operations and neural timing net architectures along these general lines have since been proposed (Cariani, 2002).

Octave Similarity

Octave similarity is a strong perceptual effect around which virtually all musical scales are constructed. Several auditory scientists have suggested that octave similarity might be grounded in interspike interval distributions. (Ohgushi, 1978, 1983; Patterson, 1986). This is easiest to appreciate if one examines interspike interval distributions for different pure tone frequencies f and complex tone F0-periodicities. Figure 5.10 shows simulated auditory nerve all-order population-interval distributions (PIDs) in response to isolated pure and complex tones ($n = 1 - 6$) with different ratio relations to 400 Hz.

Note first that the neural PIDs that are produced by pure and complex tones that have the same dominant periodicities ($f_{pure} = F0_{complex}$) and that evoke the same pitches also share similar patterns of major interspike interval peaks. This immediately explains the phenomenon of *pitch equivalence*—why pure tones and harmonic complexes with similar fundamental frequencies produce similar pitches despite their very different spectra.

Different stimulus frequencies produce characteristic interspike *interval patterns* that repeat at subharmonics of the stimulus fundamental. The neural correlate of the pitch of a tone is likely to be a *pattern of interspike intervals* rather than preponderance of a single interval—i.e., a pattern of PID peaks rather than one highest peak. Thus, the interval patterns for pure tones of a given frequency bear a high resemblance (major peaks at the same time lags) to their complex tone counterparts.

Octave similarity and *chroma-equivalence* are also explicable in these terms. Tones an octave apart (1:1 vs. 2:1, 400 vs. 800 Hz) might also be regarded as similar, because of the pattern-similarities of their representations. Tones an octave apart (400–800 Hz) share half their interval peaks, and tones separated by other simple frequency ratios also share peaks as well. The vertical lines in the plots of figure 5.10 show the characteristic rhythms of the interval alignments. The plots on the right show the fraction of common peaks (major and minor) that the 400 Hz interval pattern shares with those produced by other tone frequencies. For the complex tones (harmonics 1–5), there are also shared peaks within the octave (inset plot) that are associated with the Pythagorean consonances (3:2, 4:3, 5:3). The rank ordering of these magnitudes comports with those of consonance judgments (figure 5.4) as well as estimates from roughness and pitch fusion models (figures 5.6 and 5.9).

A common, but weak, explanation attributes octave similarity to harmonic overlap. Although two complex tones with F0s an octave apart share half their harmonics, octave similarity is also easily perceived for low-frequency pure tones and harmonic complexes with unresolved harmonics, where there are no harmonics in common. Pure tones and complex tones of unresolved harmonics are also fully capable of conveying melodies and harmonies despite their lack of overlapping distinguishable harmonics.

The forms of the PIDs resemble the positive portions of the stimulus autocorrelation functions.[17] The major peaks all correspond to subharmonic periods (n/*f*, n/F0). All three of these classes of sounds that produce octave equivalences do share common subharmonics (n/F0) that are well represented in all-order interspike interval distributions in the auditory nerve and brainstem. These are shared patterns of subharmonics of the harmonics present, not the harmonics themselves. Subharmonic relations and patterns of interaction provide a potential neural basis for chroma relations.

One can also compute Pearson correlations between simulated PIDs and arrive at similar results (Cariani, 2002). The pure tones show positive correlations only at small multiples (2:1 and 3:1 and their inverses), whereas harmonic complexes also show prominent correlations related to the Pythagorean consonances. The waveforms, power spectra, and autocorrelations of pure tones an octave apart are completely uncorrelated, but when waveforms are half-wave rectified, as they are in the process of auditory transduction in inner cochlear hair cells, their autocorrelations and power spectra have positive correlations. Because this distortion comes after cochlear filtering, it is present in temporal codes, but not place codes. Thus, octave similarity may be a consequence of the temporal neural codes that are used by the auditory system.

Octave stretch is another octave-related effect that may be explicable in terms of neural temporal firing patterns. The stretch is a small deviation from the true octave that is seen in octave matching experiments, usually larger for pure tones than complex tones, and fairly subtle, on the order of fractions of a semitone (Hartmann, 1993). Temporal models analyze small changes in interspike interval distributions (Ohgushi, 1983; McKinney, 1999; Ohgushi and Ago, 2005), whereas other models combine temporal and spectral factors, such that changes in place-of-excitation can weakly influence pitch (Hartmann, 1993).

To summarize, interspike interval distributions have a recurrent structure that mirrors the circularity of chroma relations and the spiral structures of musical scales (Patterson, 1986). The recurrent structure of interspike intervals holds for all periodic sounds, i.e., for both pure and complex tones. When frequency is advanced a full octave, as in figure 5.10, half of the interval peaks become realigned. When frequency is advanced by a fifth, a third of the interval peaks are aligned. Thus, different frequency ratios create characteristic alignments between the interval distributions of their respective tones. Neural discharge periodicities related to common subharmonics provide a potential basis for chroma relations and perceptual similarity relations in tonal hierarchies.

Existence Region of Musical Tonality

For most listeners, musical tonality extends upward in frequency to roughly 4–5 kHz, a frequency limit that is broadly consistent with that of significant phase-locking in the cat auditory nerve (Johnson, 1980). Efforts are underway to develop non-invasive methods for estimating this physiological limit in humans (Verschooten, Robles, & Joris, 2015). If musical pitches and chroma relations are indeed ultimately based on interspike interval information, which depends on phase locking, then its limit directly explains the upper frequency limit of musical tonality. The limit might also explain why animals with predominantly high-frequency hearing, such as birds, do not appear to perceive octave similarities (Russo, this volume; Patel, this volume). Explanations for the lower limit of tonality, roughly 20–25 Hz, are much less clear, but are usually based on the assumption that there are durational limits to neural delays available for processing pitch. The durational, temporal integration limits may vary with characteristic frequency region, with longer delays for lower frequencies and shorter ones for high frequencies (Oxenham, Bernstein, & Penagos, 2004).

Temporal Models for Consonance and Harmony

Pitch-based models of consonance rely on the degree to which different periodicities cooperate (reinforce) or compete (interfere) with each other. The greater the harmonicity of the sound, the greater the degree to which the harmonics present can subsumed into one harmonic series, or equivalently, the periodicities present can be subsumed into one fundamental pattern. The more unified the pitch and its underlying auditory representation, the fewer alternative pitches it implies, and the greater its harmonic stability (Lipps, 1905/1995).

Pitch unity is closely related to *tonal fusion*, the degree that the pitches of sounds fuse together, on which Carl Stumpf's proto-Gestaltist theory was based (Stumpf, 1883/1890; Boring, 1942; Schneider, 1997). Our interpretation combines Stumpfian concepts of tonal fusion with pitch stability. In terms of harmonicity and pitch multiplicity, the more unified and stable the set of pitches evoked by a given stimulus, the more consonant it will sound. The more the pitches that are evoked compete with each other, the less stable is the sound's harmonic interpretation, and the more dissonant it will sound. Stability, here, means predictive certainty—that is, the opposite of pitch ambiguity or harmonic entropy (Sethares, 2005, p. 371). For example, in vision a two-dimensional square is an extremely stable form because it has but one dominant perceptual interpretation, whereas Necker cubes and other reversible figures are unstable because they have multiple, competing perceptual interpretations that dominate with roughly equal probabilities.

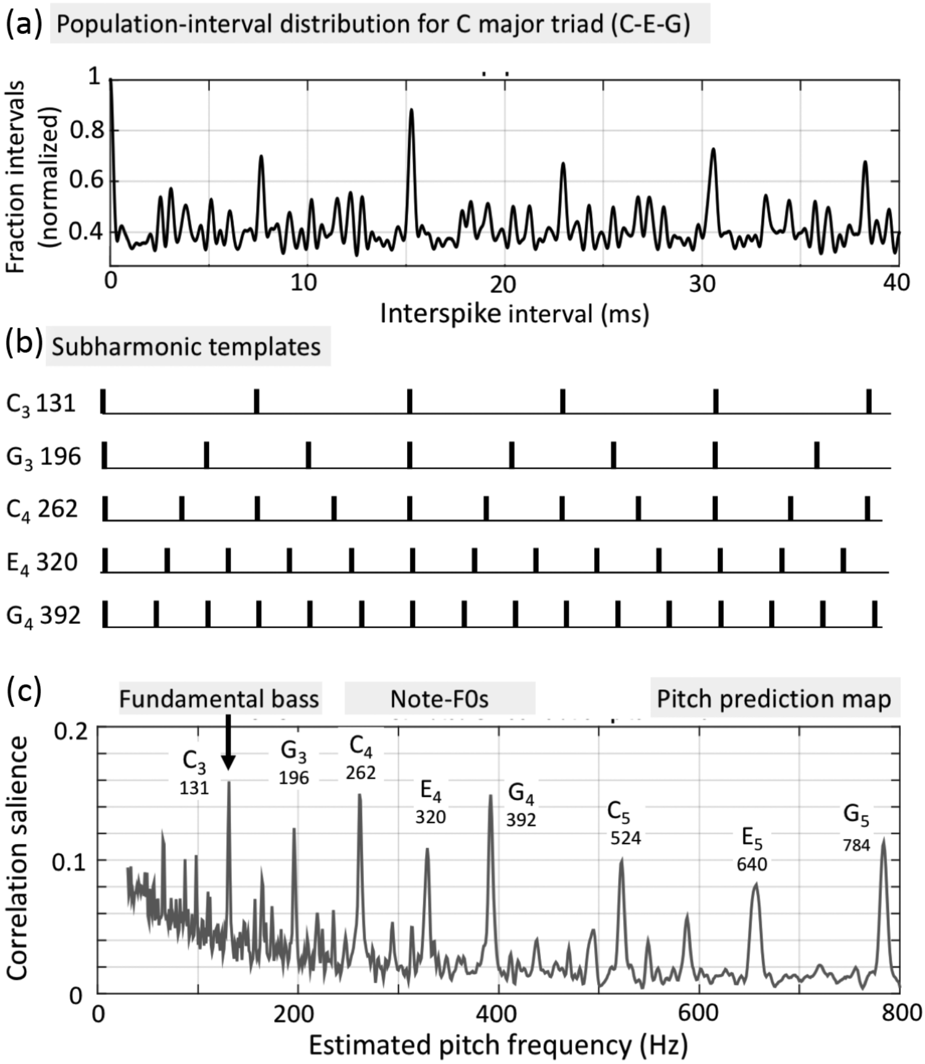

Figure 5.9

Temporal model of pitch multiplicity: estimated consonance of dyads and pitch stability of chords. (a) Simulated normalized auditory nerve population-interval distribution (PID) response to C major triad (C_4-E_4-D_4). (b) Selected subharmonic sieve templates for five selected periodicities that correspond to the three note-F0s plus two subharmonics. The full set contains all frequencies F0 = 30–1000 Hz, 1 Hz steps. Sieve tines had 0.2-ms widths. (c) Map of estimated pitch strengths (salience) of Pearson correlation coefficients (salience) of the PID with all sieves. The highest salience is found for the fundamental bass (arrow), the F0 of all of the notes comprising the chord, at 131 Hz, an octave below the root of the chord (C_4 = 262 Hz). In addition, weaker pitches corresponding to the individual note F0s (C_4, E_4, D_4) and their second and third harmonics (C_5, E_5, D_5, C_6) are predicted.

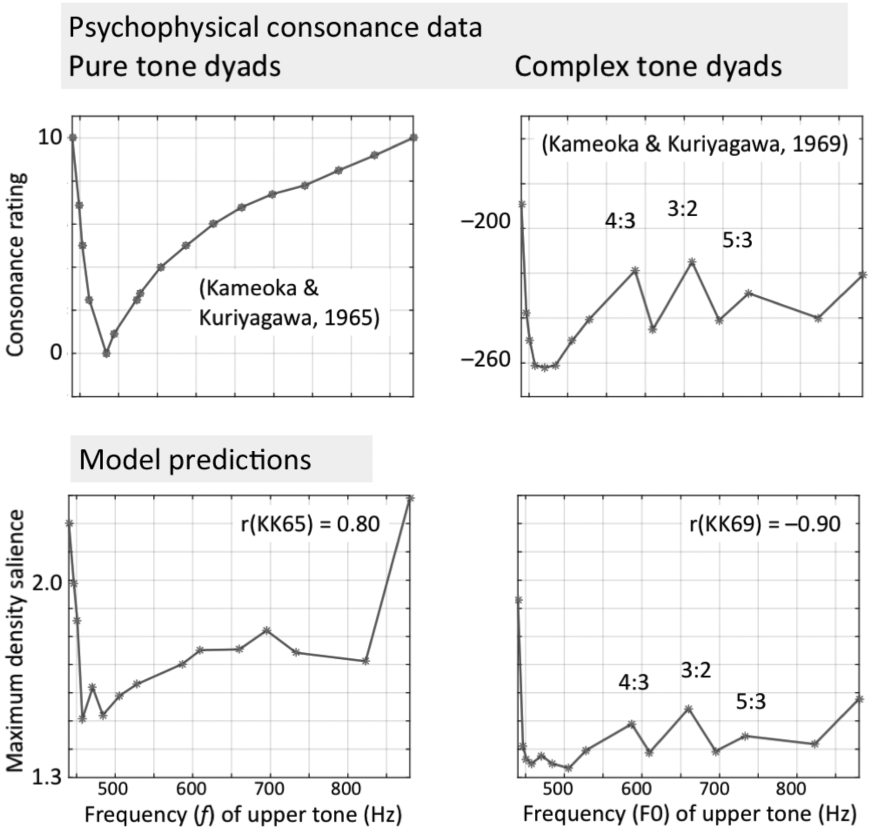

Figure 5.10

Temporal model predictions for consonance. Top plots. Judgments of consonance for dyads of pure tones (left) and complex tones (right) by thirty-one Japanese audio engineers. Pure tone data from figure 1 of Kameoka and Kuriyagawa (1965); complex tone data from figure 7 of Kameoka and Kuriyagawa (1969), same as figure 5.4 above. Subjects were instructed to judge the relative consonance/dissonance of tone dyads according to their "clearness" (*sunda*) vs. "turbidity" (*nigotta*). Bottom plots. Estimates of pitch stability from the temporal pitch multiplicity model. Pitch estimation was based on maximum correlation between subharmonic sieve and population-interval distribution (PID). Maximal salience was estimated using the mean density of intervals in PID bins in the subharmonic sieve for the estimated pitch.

One of the advantages of a theory of consonance based on pitch stability is that it readily couples to music cognition and music-theoretic concepts of pitch entropy/uncertainty, tonal centers, perceptual distance hierarchies, and harmonic tension/relaxation dynamics (Parncutt, 2011a, 2011b). Pattern-similarities between temporal representations similar to population-interval distributions when processed here using self-organizing Kohonen networks replicate neighborhood perceptual distance relations between notes, chords, and keys (Leman & Carreras, 1997; Leman, 2000; Krumhansl & Cuddy, 2008).

Pitch multiplicity The temporal model of pitch multiplicity presented in figure 5.11 predicts the consonance of musical intervals (figure 5.11d) and also estimates the relative pitch

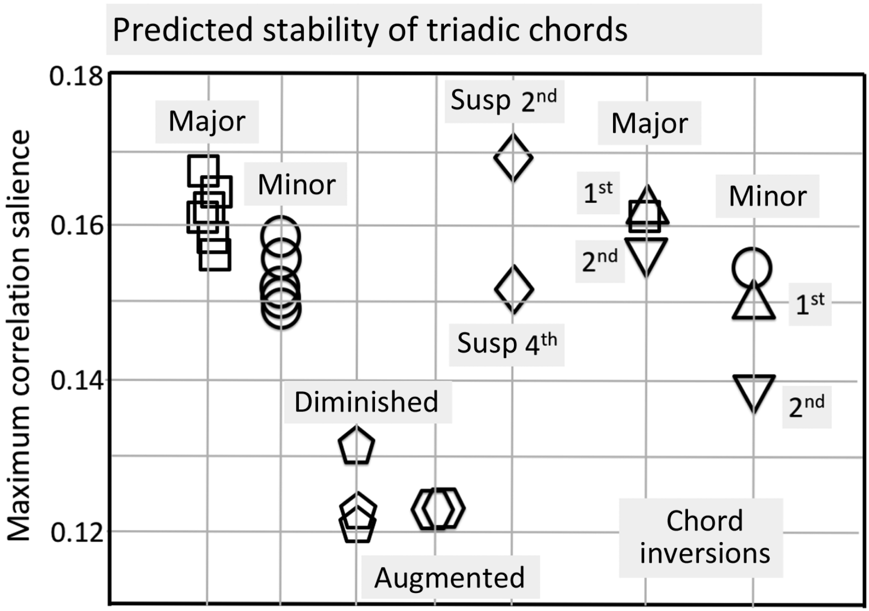

Figure 5.11
Predicted pitch stabilities of triadic chords. Maximal correlation saliences (estimated pitch stabilities) for triadic chords and their inversions from the temporal pitch multiplicity model. All triads had roots in the C_4–C_5 octave: I, II, III, IV, V, VI, VII, D_{maj}, E_{maj}, G_{min}, C_{min}, C_{aug}, S_{susp4}, C_{susp2}, C_{aug}, G_{dim}, A_{maj} and first & second inversions, A_{min} and first & second inversions. Notes consisted of harmonics 1–6, 70 dB SPL simulated level, 317 CF positions, 631 ANFs; all other simulation parameters as in D, which are associated with various periodicities in the neural response. The subharmonic templates for several of these note-F0 periodicities are shown below it (11B) such that sets of prominent PID peaks that correspond to the subharmonic patterns can be easily visualized by comparing the two plots (A, B). The map of estimated pitch strengths (saliences) is shown in (11C), which are the maximal correlation values of the PID with the set of subharmonic peaks associated with frequencies 25–800 Hz, $\Delta f = 1$ Hz.

stabilities of isolated triadic chords (figure 5.11e). Pitch multiplicity involves the ability of listeners to hear more than one pitch for a given stimulus. In addition to estimating the strongest, most likely pitch that will be heard, in order to assess pitch multiplicity and stability, pitch models also require means of estimating the relative strengths of multiple competing pitches.

Early interspike interval models (Meddis & Hewitt, 1991; Cariani & Delgutte, 1996a, 1996b; Meddis & O'Mard, 1997) used the time interval (τ) associated with the first major individual peak in auditory nerve PIDs as an estimate of the period of one pitch that could be heard. However, the strategy of picking individual peaks in autocorrelation-like functions has well-known difficulties with octave confusions. It is also unreliable for pitch estimation when multiple musical notes and/or human voices are presented concurrently.

Our pitch model (Tramo et al., 2001; Cariani, 2004a) evolved from dependence on one predominant interspike interval to a pattern of pitch-related intervals, i.e., a subharmonic series of peaks (n/*f*, n/F0) rather than a single peak (1/*f*, 1/F0). The model estimates the relative pattern-strengths of *interspike interval patterns* associated with different stimulus periodicities and pitches.

The temporal model is illustrated in figure 5.9 in response to a C-major triad. First, spike train responses of hundreds of simulated auditory nerve fibers to a stimulus are simulated (not shown).[18] The all-order interspike interval distributions of these fibers are weighted and summed to compile the estimated population-interval distribution (PID) of the human auditory nerve (figure 5.9a). The current version of the pitch estimation algorithm estimates the strengths of possible pitches by computing the correlations between the simulated population-interval distribution (figure 5.9a) and an array of subharmonic patterns corresponding to different pitch frequencies (figure 5.9b).

The subharmonic pattern with the maximum correlation is taken as the estimate of the most salient estimated pitch, i.e., the pitch most likely to be heard. Maximum pitch salience (strength), which is interpreted as a measure of pitch stability, is taken as the correlation value associated with the strongest pitch (*correlation salience*).

The specific example of a C-major triad (C_4-E_4-G_4) is illustrated in figure 5.9. The population-interval distribution (figure 5.9a) that is produced shows a pattern of major and minor peaks. The pitch with the highest salience corresponds to C_3 (131 Hz), which is an octave above the true fundamental of the chord C_2 (65 Hz). The *fundamental bass* (*basse fundamentale*) is the fundamental of the note-F0s, which is the same as the fundamental of all of the harmonics present in the chord. In music-theoretic terms, however, C_2 and C_3 are in the same chroma class and therefore are functionally identical, so the fundamental bass of the chord has the chroma class of C. The maximal salience of this dominant pitch produced by the chord is then taken as a measure of the pitch stability of the chord. The theory holds that all pitches above a critical salience threshold should be audible, such that other weaker, possible pitches are predicted for the F0s of the individual notes plus some of their lower harmonics.

Tonal consonance We have also simulated the experiments of Kameoka and Kuriyagawa (1965, 1969b). Maximal salience values for different musical intervals formed from dyads of pure and complex tones are shown in figure 5.11. These are the same stimuli with the same spectra as were used by Kameoka and Kuriyagawa (1969b) for their psychophysical consonance experiments (figure 5.4a). This study used an older pitch estimation algorithm based on estimating the pattern-strengths of F0-related interspike intervals using the same subharmonic sieves as in figure (figure 5.9b). The resulting estimated saliences are called *density saliences* in figure 5.11 because they compute the respective densities of intervals in different sets of PID bins. The estimates from the temporal pitch multiplicity model closely match the pattern of consonance judgments that were observed ($r = 0.8, 0.9$), especially for the complex tone dyads ($r = 0.9$). These estimates of pitch multiplicity covary with those based on roughness from beating harmonics (curves of predicted consonance in figures 5.4a and 5.6d). Thus, for most musical tonal stimuli, listeners would be expected to make approximately the same patterns of consonance judgments of consonance, irrespective of whether they attend to roughness or pitch multiplicity cues.

Bidelman and Heinz (2011) have carried out an extensive investigations of roughness, consonance, and harmony using a temporal model of this type and a pitch estimation algorithm based on interval-densities. They have obtained results that are very similar to those presented here.[19] They also studied the effects of simulated hearing impairment on perception of dyads and triads. Many of their general conclusions parallel those drawn here.

Pitch stability of chords The temporal model can be used to estimate the pitch stability of chords. Different types of isolated chords that include major, minor, augmented, diminished, and suspended 2nd and 4th, as well as first and second inversions of major and minor chords, produce different degrees of estimated pitch stability. Major triads and suspended 4th chords were estimated to be the most stable, followed by minor triads and suspended 2nd chords, with the least stable chords being augmented and diminished triads. These stability rankings are in qualitative agreement with listener ratings (see figure 5.4b; McDermott et al., 2010b) and with music-theoretic principles (Piston & DeVoto, 1987). The model predicts that chord inversions should have relatively subtle effects, if any, on pitch stability.

Although these kinds of neural models can thus potentially handle concurrent harmonic relations, in order to account for melodic, sequential tonal interactions, some sort of auditory short-term memory must be included. Here internal representations of tonal contexts are built up through a persistent, running pitch representation that interacts with new, incoming neural activity patterns (e.g., Huron & Parncutt, 1993; Leman, 2000). The representation, whether realized through persistent firing of feature-selective neurons or regenerated temporal patterns of spikes, would need to persist on the order of a few seconds to integrate temporal melodic, rhythmic, and timbral sequences of events.

Relation to virtual pitch models The temporal multiplicity model has strong correspondences with Ernst Terhardt's virtual pitch model and his theory of harmony (Terhardt, 1974, 1984). His student Richard Parncutt has greatly extended the theory (Parncutt, 1989) and drawn out some of its general implications for harmony (pp. 68–70). Both time- and frequency-domain subharmonic approaches are related to Rameau's concept of the fundamental bass (Rameau 1772/1971). Although Terhardt's model is couched in spectral terms, in the frequency domain, it is based on an analysis of common subharmonics of the partials present, the fundamental being the highest common subharmonic present.

In effect, temporal pitch models based on all-order interspike intervals implement an autocorrelation-like representation that does exactly this, albeit in the time (delay) domain, using interspike intervals rather than acquired subharmonic templates. Like the Terhardt model, the population-interval distribution also reflects the strengths of the different harmonics. Summing interspike intervals carries out a superposition of subharmonics of harmonics that is weighted by their respective amplitudes. The more intense the harmonic, the more ANFs it drives, and the more interspike intervals present that will reflect its subharmonic pattern. In contrast to Terhardt's model, the temporal model also covers musical pitches produced by groups of unresolved harmonics. Whereas Terhardt's theory assumes that subharmonic templates are self-organized through auditory experience, the subharmonics are already directly present in the interspike intervals that are produced directly by phase-locked spiking in the auditory nerve (see also Bidelman & Heinz, 2011).

Historically, theories of harmony, such as Riemann's (1905/2011) based on both harmonic and subharmonic relations (overtones and undertones) have been dismissed out of hand for several reasons related to acoustics, perception, and physiology. First, the subharmonics themselves are not thought to be present in the acoustics. Through Fourier analysis, one sees only frequency components (harmonics), but from autocorrelation one sees mainly peaks at time delays related to all of the subharmonics of those components. Second, pitches corresponding to subharmonics are not usually explicity perceived. Lord Rayleigh (1894/1945) criticized periodicity-based theories of pitch, and implicitly with them theories of harmony based on the undertone series on the grounds that they would predict that we should hear all of the pitches associated with all of the individual subharmonics (which we don't). Finally, subharmonics were not thought to be generated in the cochlea or in neuronal responses. Subharmonics only assume a neural reality when one considers temporal, interspike interval codes. Misconceptions about the volley principle have also led some to incorrectly assert that temporal codes in the auditory system cannot represent periodicities above a few hundred Hz. But for almost a century, it has been known that in the very first stage of auditory neural processing, the auditory nerve, temporal correlates of both overtones and undertones are richly and precisely represented in interspike interval patterns.

Neurocomputational mechanisms Although there appears to be abundant evidence for the existence of spike timing information that could subserve musical tonality, the neurocomputational mechanisms by which this informational might be utilized are much less clear. The temporal pitch multiplicity model is a neuropsychological model in that it predicts auditory percepts based on neural responses, but it does not specify the neural mechanisms by which a central pitch analysis could be carried out. The subharmonic correlation-based pitch estimation method outlined here is thus intended as a means of quantifying the relative strengths of interspike interval patterns associated with different possible perceived pitches, and not as a literal neurocomputational mechanism.

There are deep problems with all neurocomputational models for musical pitch that use specific fixed templates, as most listeners have relative, not absolute pitch. Aside from absolute pitch, pitch perception does not on its face appear to be the result of a template-based recognition process.

Equally problematic are theories that postulate a harmonic spectral pattern analysis at the cortical level. Some theorists hold that all periodicity-related information, temporal- and place-based, is converted to rate-place codes in the ascending auditory pathway by the time it reaches the cortical level (Plack, Barker, & Hall, 2014), and that an F0-spectral pattern analysis could potentially be realized via analysis of those cortical rate-place representations (Fishman, Micheyl, & Steinschneider, 2013). Others point to harmonic template neurons, most of which respond to combinations of high-frequency harmonics (Feng & Wang, 2017) above the frequency existence region for musical pitch.

However, if such a rate-based spectral pattern analyses were carrried out at the cortical level, we should be able to hear strong F0-pitches from low-order resolved harmonics above 4 kHz. According to such hypotheses, the missing fundamental F0-pitch at F0 = 1.5 kHz produced from harmonics 3–5 (4.5–6.0–7.5 kHz) should be every bit as strong as its lower frequency counterpart at F0 = 150 Hz produced from harmonics 3–5 (450–600–750 Hz). The latter produces a strong F0-pitch, whereas the former produces *no* F0-pitch.

At the cortical level, neurons have been found that respond somewhat selectively to periodicities associated with pure tones and combinations of resolved and/or unresolved harmonics (Bendor & Wang, 2005), although, like many neurons in auditory cortex, their firing rate responses also tend to be highly dependent on sound level (nonmonotonic, i.e., responding at only moderate sound levels). The rate responses are therefore unlike pitch percepts, which remain stable and precise at high levels. Perhaps they are the right neurons, but information regarding pitch is embedded in some other aspect of their spike train outputs rather than through average spike rates.

The response properties of these F0-tuned neurons do not necessarily favor spectral pattern over temporal processing, as these responses might be the consequences of either kinds of neural processing at lower-levels of the auditory pathway. Neurons tuned to specific modulation frequencies of unresolved harmonics have been found in the brainstem and midbrain, and putative helical neural maps that form cylindrical tonotopic-periodotopic spaces have been proposed (Langner & Benson, 2015). However, these responses cannot account for the stronger musical pitches produced by pure tones and harmonic complexes. Also, unlike pitch percepts, these modulation tunings degrade at high sound levels.

We are still a long way from understanding how the auditory cortex works in terms of the details of neural representations and operations because we do not yet have a firm grasp of the neural codes that are operant at that level.

Neural timing net theory would hold that the representation of a particular pitch or rhythm is itself a temporal pattern of spikes such that pitch perception does not involve matching to specific, fixed templates; pitches are primarily known only in relation to other pitches, via their neural signal-signal interactions (Cariani, 2002). For relative pitch comparisons, the interval patterns themselves can be compared to those associated with other sounds, provided there is some kind of short-term memory mechanism available that can temporarily hold temporal patterns. Analogous neural correlation operations could also be carried out by neural time-delay architectures (Licklider, 1959; Boomsliter & Creel, 1961), modulation detectors and/or neural autocorrelators (Langner & Benson, 2015), recurrent timing nets (Cariani, 2002), or nonlinear oscillatory networks (Large, Kim, Bharucha, & Krumhansl, 2016).

Neural temporal discharge patterns related to periodicities of fundamentals, harmonics, and subharmonics have been observed in brainstem frequency-following responses, and the model of pitch stability based on maximal salience seems to hold there (Bidelman & Krishnan, 2009; Lee, Skoe, Kraus, & Ashley, 2014). Individual preferences for tones with higher harmonicity also appear to be reflected in temporal response patterns at the level of the brainstem (Bones, Hopkins, Krishnan, & Plack, 2014).

The main difficulty for temporal theories currently involves a lack of understanding of how fine timing information that is abundant and available to midbrain auditory stations would be utilized at thalamic and cortical levels. F0-related spike timing information in the auditory cortex is evident only up to a few hundred Hz (Cariani, 1999; Cariani & Micheyl, 2012; Fishman, Micheyl, & Steinschneider, 2013), which is insufficient for temporal coding of the full gamut of musical pitches.

The psychophysics strongly suggest that this temporal information from the periphery is utilized, but the central neural codes and computations to do it remain obscure (Cariani & Micheyl, 2012). There are also tantalizingly elegant and powerful time-domain scene analysis, discrimination, scaling, and recognition operations that could be carried out if the temporal information were available in some form (Cariani, 2001a; Cariani, 2002; Cariani, 2004b; Cariani, 2015). Despite substantial advances in understanding cortical representations of pitch and harmonic relations (Janata et al., 2002; Warren, Uppenkamp, Patterson, & Griffiths, 2003; Bendor & Wang, 2005; Hyde, Peretz, & Zatorre, 2008; Bizley, Walker, King, & Schnupp, 2010; Fishman et al., 2013; Norman-Haignere, Kanwisher, & McDermott, 2013), the precise nature of those codes and the coding transformations that must exist in the ascending auditory pathway remain enigmas.

Conclusions

Major properties of musical scales and tunings and their relations to the structure of pitch perception and underlying auditory neural representations have been considered and include:

1. The octave-repeating linear structure of musical scales replicates the two dimensional, helical structure of pitch space, with pitch height as its axial dimension, and pitch chroma as its circular dimension.
2. Musical intervals are relative, frequency-ratio relations to the tonic.
3. Musical tonality (octave similarity, musical interval discriminations, tonal hierarchical similarities, and melodic invariance under transposition) depends on pitch chroma relations.
4. The pitch ranges of scales and musical instruments reflect the periodicity and frequency limits of musical tonality, ~25–4,000 Hz.
5. Most scales consist of discrete sets of five to twenty-four notes, enabling pitches to be replicated and for musicians to play the same and/or different compatible pitches together.
6. Most of the world's scales incorporate notes with musical interval frequency-ratios to the tonic at or near the Pythagorean consonances (1:1, 2:1, 3:2, 4:3, 5:3, 5:4, 6:5).
7. *Consonance* can have a variety of meanings related to perceptions of or preferences for particular combinations of tones. Two aspects of consonance involve perceived roughness, which arises from beating harmonics, and perceived inharmonicity, which arises from multiple competing, clashing periodicities that are not harmonically related. Some listeners attend to and prefer sound combinations that are less rough, whereas others prefer those that have higher harmonicity.
8. Just intonation tuning uses ratios of integers. Pythagorean tuning is based on fifths and octaves. Equal temperament (ET) divides the octave into equal frequency-ratio steps. ET regularizes inter-note frequency ratios to eliminate subtle differences in tonal colorations of different keys, thereby enabling free modulation between keys.

9. Tuning systems for instruments with harmonic spectra provide close approximations to these consonances. Just tuning provides exact approximations, whereas twelve-tone, equal temperament (12-TET) provides approximations close enough (< 1 percent) to most of these consonances such that perceived roughness and inharmonicity are minimized, and not noticeable for most listeners in common musical situations. Equally tempered pentatonic (5-TET), heptatonic (7-TET), and chromatic (12-TET) scales give the best approximations, with chromatic being closest.
10. Musical pieces that use different scales or modes produce different tonal colorations because they contain different sets of musical intervals that have different chroma-relations to the tonic that in turn make different sets of tonal hierarchical relations possible.
11. Whereas best pitch discriminations are on the order of 0.1 percent in frequency, best discriminations of transposed musical intervals are much coarser, on the order of a quartertone (3 percent). In both categories musicians tend to make finer discriminations than untrained listeners. Musical interval perception is highly categorical.
12. Whereas the consonance of concurrent sounds is most critical for tuning systems, choice of musical intervals and their relations to tonal hierarchies is more important for scales.
13. Many basic aspects of scales and tunings may be ultimately due to the structure of temporal, interspike interval neural codes for pitch in the early auditory system. Interspike interval models of pitch account for musical note pitches. Pitch chroma relations may be mediated by temporal codes.
14. Octave similarity and the circular organization of pitch chroma may be due to the circular, repeating structure of neural interspike interval distributions in the auditory system. Tones separated by simpler frequency ratios share higher fractions of interspike interval peaks.
15. The upper frequency limit of spike timing information (4–5 kHz) likely determines the upper frequency limit for musical tonality.
16. Temporal pitch multiplicity models and virtual pitch models produce estimates of tonal consonance similar to those of roughness models; both successfully predict the consonance of Pythagorean ratios.
17. Temporal pitch multiplicity and virtual pitch models couple to Rameau's theory of harmony, which is based on relative strength of the fundamental bass.
18. Temporal pitch multiplicity and virtual pitch models estimate the relative tonal stabilities of triads (major and suspended 2nd > minor and suspended 4th > augmented and diminished chords). Virtual pitch models may explain other aspects of tonal hierarchies as well.

Acknowledgments

I would like to thank Nelson Kiang, Bertrand Delgutte, and Mark Tramo for initiating me into the world of auditory neuroscience and its connections to music. Thanks to William Hartmann, Richard Parncutt, Jason Yust, Andy Brughera, Andy Oxenham, and Dan Levitin for useful comments on this chapter. I have very much appreciated the efforts of Muhammed Zilany, Ian Bruce, and Laurel Carney to make their auditory nerve model available to the world at large. Thanks also to anonymous contributors to the Wikipedia entry on Shrutis.

Notes

1. For example, Kurt Schwitter's *Ursonate* (1922–1932).

2. Belgian musicologist François-Joseph Fétis, quoted in Hyer, 1999.

3. *Perceptual resolvability*, as used here, refers to the ability to hear out and accurately matches the pitch of an individual harmonic in a harmonic complex to that of a pure tone of adjustable frequency (Plomp, 1976). There are a host of other theoretical, psychophysically derived measures of resolvability in the auditory literature that are related to frequency selectivity, critical bandwidths, and auditory filter shapes (Moore, 2013; Zwicker & Fastl, 1999).

4. Burns (1999) hypothesizes that listeners may be able to extend the range of usable temporal information through learning (experience). For example, most listeners hear only binaural beats of pairs of pure tones up to ~1,200 Hz, but with training, this limit can be pushed upward more than an octave (Wever, 1948).

5. Frequency difference limens for identification of pitch direction with small changes of frequency are generally comparable to those for pitch discrimination, but there is considerable individual variability, with some listeners (counterintuitively) able to identify pitch direction changes at 50% smaller frequency differences than they require to detect a pitch change, and with others going the other way (Semal & Demany, et al., 2006). Oddly, some amusics can sing with better intonation than they can consciously perceive (Peretz, 2016).

6. Melodic invariance under transposition is analogous to the invariance of rhythmic patterns under changes in tempo, albeit at different time scales (Boomsliter et al., 1961). Both transformations can be achieved though time scale dilation/compression, which preserves temporal ratio relations between F0 pitch periods in melody and between inter-event intervals in rhythm. These auditory temporal pattern invariances are analogous to visual spatial pattern magnification-invariance.

7. Similar ambiguous situations arise for pitch and consonance when individual listeners attend to different aspects of sound. In musical contexts, when asked to make judgements about "pitch," listeners attend to F0-pitches, but if directions are ambiguous, listeners may instead attend to changes in spectral center-of-gravity instead. Musically more experienced listeners tend to focus on F0-pitch differences related to periodicity, whereas less experienced listeners tend to focus on pitch height differences related to spectrum (Shepard, 1982).

8. Although consonance has always encompassed interactions of both sequential and concurrent notes, Tenney (1988) argues that the concept has undergone at least five historical semantic transitions: early conceptions of consonance focused mainly on sequential pitch relations rather than properties of simultaneous notes, such that the association of consonance with roughness is a comparatively modern idea.

9. One should not conflate perceptual distinctions and listener preferences. For example, listeners may be able to distinguish consonant from dissonant chords, yet prefer neither (McDermott et al., 2016). Preferences imply discriminability, but not vice-versa. In general, basic discriminations between musical events are more likely to be determined by auditory constraints, whereas preferences are much more open to influence from acquired learned associations and rewarded cultural norms. On the other hand, many discriminative auditory acuities can be improved by musical training, and some preferences may be innate, near-universals (e.g., sweet vs. bitter tastants, low vs. high frequency tones).

10. See von Helmholtz (1885/1954, pp. 280–284), Forster (2010), and Midya (2015) for more detailed analyses.

11. Note that 1 cent is ~1/100th of a semitone, and 15 cents is roughly a sixth of a semitone or about 1 percent in frequency.

12. Global temporal models can be contrasted with spectral pattern models that recognize patterns of resolved harmonics in spectral representations (Goldstein, 1973; Wightman, 1973; Terhardt, 1974; Cohen, Grossberg, & Wyse, 1994). Those spectral pattern models that successfully predict pitch from realistic neural responses to acoustic stimuli (Srulovicz & Goldstein, 1983) invariably rely on spike timing (interspike interval) information to first form a putative neural central spectrum representation that is then harmonically analyzed to estimate pitch. A major shortcoming of spectral pattern theories is that they cannot account for the (musical) F0-pitches that are produced by unresolved, higher harmonics. Pitch theories that incorporate spectral pattern analysis therefore require a second, temporal mechanism to account for the gamut of musical pitches.

13. The two-mechanism F0-pitch hypothesis would predict that these amusics have intact periodicity analyzers for unresolved harmonics, but an impaired spectral pattern analysis mechanism for resolved ones. A unified temporal hypothesis would predict that amusics lack the additive or multiplicative interactions between multiple frequency regions needed for integration of individual resolved harmonics into the central population-interval representation. Whereas unresolved harmonics create prominent F0-related spike periodicities in the auditory nerve, resolved harmonics produce patterns related to individual harmonics, such that F0 periodicities only become dominant after interspike intervals from different tonotopic regions are combined somewhere in the auditory pathway. A dearth of cross-frequency connections might therefore cause a deficit in perception of the F0-pitch of resolved, but not unresolved, harmonics. Without those interactions and integrations, complex tones with resolved harmonics would be expected to produce unfused, clashing jumbles of pitches of individual harmonics that might cause even individual notes to sound dissonant.

14. "Agreeable consonances are pairs of tones which strike the ear with a certain regularity; this regularity consists in the fact that the pulses so 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” (Galileo, quoted in Plomp & Levelt, 1965, p. 549).

15. “The octave relation, the musical third, fourth, and other consonant intervals are understandable on essentially the same basis. When the frequencies of two sounds, either sinusoidal or complex, bear to each other the ratio of two small integers, their autocorrelation functions have common peaks” (Licklider, 1951, p. 131).

16. “If cadence of discharges were relevant to tone perception, one could infer that the less regular the cadence, the harsher and or rougher or more dissonant the sensory experience. If this were true, the neural data would predict a relation between consonance and frequency ratio because, in response to a complex periodic sound, the smaller the numbers in the frequency ratio the more regular is the discharge cadence. Therefore, our neural data can be taken to support a frequency-ratio theory of consonance” (Rose, 1980, p. 31).

17. For pure tones of frequency *f*, PIDs approximate half-wave rectified cosine functions, i.e., the maximum of 0 or cos $(2\pi f\, t)$ at each time point t.

18. The simulations used the public Zilaney, Bruce, and Carney (2014) model of auditory nerve fibers. Except where noted, similar parameters were used for all of the auditory nerve simulations discussed here (figures 5.8–5.11). For the plots of figure 5.8 300 ANFs were stimulated at 75 dB SPL simulated level, whereas for the simulations of dyads (figure 5.10) and triads (figure 5.11), 320 CF positions with three spontaneous rates per CF position were used (640 ANFs total). Specific model parameters: species (2), i.e., Shera Q human auditory filters, normal ohc and ihc function (1), noiseType (1), human distribution of ANF CFs, distribution of spontaneous rate classes (20% low sr, 30% medium sr, 50% high sr). PIDs were normalized by dividing by the PID mean. Other details concerning the pitch algorithm can be found in Cariani (2004) and Bidelman and Heinz (2011).

19. I developed the method of using dense (ΔF0 = 1 Hz) interspike interval sieves in the late 1990s (Cariani, 2004a) to systematically quantify the pattern-strengths associated with all possible F0-pitches within the tonality existence region. This method used the average density of interspike intervals in the pattern (in sieve bins) to estimate pitch saliences. Bidelman and Heinz (2009) subsequently used this method to predict consonance judgments, and obtained results similar to mine. A few years ago I developed the newer method of computing correlations to subharmonic delay patterns that eliminate pervasive octave confusions that hamper the interval density method and also obviate the need for weighting shorter autocorrelation time lags. Aside from octave errors, the two methods produce comparable results.

References

Atteneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology*, *84*(2), 147–166.

Bachem, A. (1950). Tone height and tone chroma as two different pitch qualities. *Acta Psychologica*, *7*, 80–88.

Barbour, J. M. (2004). *Tuning and temperament: A historical survey*. Mineola, NY: Dover.

Benade, A. H. (1990). *Fundamentals of musical acoustics*. New York: Dover.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, *436*(7054), 1161–1165.

Biasutti, M. (1997). Sharp low- and high-frequency limits on musical chord recognition. *Hearing Research*, *105*(1–2), 77–84.

Bibby, N. (2003). Tuning and temperament: Closing the spiral. In J. Fauvel, R. Flood, & R. J. Wilson (Eds.), *Music and mathematics: From Pythagoras to fractals* (pp. 13–27). New York: Oxford University Press.

Bidelman, G. M., & Heinz, M. G. (2011). Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing. *Journal of the Acoustical Society of America*, *130*(3), 1488–1502.

Bidelman, G. M., & Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *Journal of Neuroscience*, *29*(42), 13165–13171.

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

Bigand, E. (1993). Contributions of music to research on human auditory cognition. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 231–277). Oxford: Oxford University Press.

Bizley, J. K., Walker, K. M., King, A. J., & Schnupp, J. W. (2010). Neural ensemble codes for stimulus periodicity in auditory cortex. *Journal of Neuroscience, 30*(14), 5078–5091.

Bones, O., Hopkins, K., Krishnan, A., & Plack, C. J. (2014). Phase locked neural activity in the human brainstem predicts preference for musical consonance. *Neuropsychologia*, *58*(May), 23–32.

Boomsliter, P., & Creel, W. (1961). The long pattern hypothesis in harmony and hearing. *Journal of Music Theory*, *5*(1), 2–31.

Boomsliter, P., & Creel, W. (1963). Extended reference: An unrecognized dynamic in melody. *Journal of Music Theory, 7*(1), 2–22.

Boring, E. G. (1942). *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century-Crofts.

Bowling, D. L., & Purves, D. (2015). A biological rationale for musical consonance. *Proceedings of the National Academy of Sciences USA, 112*(36), 11155–11160. doi:10.1073/pnas.1505768112.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 215–264). San Diego, CA: Academic Press.

Burns, E. M., & Viemeister, N. F. (1976). Nonspectral pitch. *Journal of the Acoustical Society of America, 60*(4), 863–869.

Burns, E. M., & Viemeister, N. F. (1981). Played-again SAM: Further observations on the pitch of amplitude-modulated noise. *Journal of the Acoustical Society of America, 70*(6), 1655–1660.

Burns, E. M., & Ward, W. D. (1982). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The psychology of music* (pp. 241–269). New York: Academic Press.

Butler, D. (1992). *The musician's guide to perception and cognition* [1 sound disc]. New York: Schirmer Books.

Cariani, P. (1999). Temporal coding of periodicity pitch in the auditory system: An overview. *Neural Plasticity, 6*(4), 147–172.

Cariani, P. (2001a). Neural timing nets. *Neural Networks, 14*(6–7), 737–753.

Cariani, P. (2001b). Temporal coding of sensory information in the brain. *Acoustical Science and Technology, 22*(2), 77–84.

Cariani, P. (2002). Temporal codes, timing nets, and music perception. *Journal of New Music Research, 30*(2), 107–136.

Cariani, P. (2004a). A temporal model for pitch multiplicity and tonal consonance. In S. D. Lipscomb, R. O. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception & Cognition (ICMPC8, Evanston, IL)* (pp. 310–314). Adelaide, Australia: Causal Productions.

Cariani, P. (2004b). Temporal codes and computations for sensory representation and scene analysis. *IEEE Transactions on Neural Networks, 15*(5), 1100–1111.

Cariani, P. (2015). Outline of a cybernetic theory of brain function based on neural timing nets. *Kybernetes, 44*(8/9), 1219–1232.

Cariani, P., & Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology, 76*(3), 1698–1715.

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

Cariani, P., & Micheyl, C. (2012). Towards a theory of information processing in the auditory cortex. In D. Poeppel, T. Overath, & A. Popper (Eds.), *Human auditory cortex: Springer handbook of auditory research* (pp. 351–390). New York: Springer.

Carlyon, R. P., & Shackleton, T. M. (1994). Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms? *Journal of the Acoustical Society of America, 95*(6), 3541–3554.

Cohen, M. A., Grossberg, S., & Wyse, L. L. (1994). A spectral network model of pitch perception. *Journal of the Acoustical Society of America, 98*(2), 862–879.

Cousineau, M., Oxenham, A. J., & Peretz, I. (2015). Congenital amusia: A cognitive disorder limited to resolved harmonics and with no peripheral basis. *Neuropsychologia, 66*, 293–301. doi:10.1016/j.neuropsychologia.2014.11.031.

Creel, W., & Boomsliter, P. (1970). Sensations of tone as perceptual forms. *Psychological Review, 77*(6), 534–545.

de Cheveigné, A. (2005). Pitch perception models. In C. J. Plack, A. J. Oxenham, R. Fay, & A. N. Popper (Eds.), *Pitch. Neural coding and perception* (Vol. 24, pp. 169–233). New York: Springer.

de Cheveigné, A. (2010). Pitch perception. In C. Plack (Ed.), *Oxford handbook of auditory science: Hearing* (pp. 1–51). New York: Oxford University Press.

Deutsch, D. (1972). Octave generalization and tune recognition. *Perception & Psychophysics, 11*, 411–412.

Deutsch, D. (1999). The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 349–411). San Diego, CA: Academic Press.

DeWitt, L. A., & Crowder, R. G. (1987). Tonal fusion of consonant musical intervals: The oomph in Stumpf. *Perception and Psychophysics, 41*(1), 73–84.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341–354.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. Orlando, FL: Academic Press.

Drobisch, M. (1852). Über musikalisch Tonebestimmung und Temperatur [Concerning musical tones and temperature]. In *Abbandlungen der Königlich sächsischen Gesellschaft der Wissenschaften zu Leipzig. Vierter Band. Abhandlungen der mathematisch-physichen Classes. Band 2.* Leipzig: S. Hhirzel.

Duffin, R. W. (2007). *How equal temperament ruined harmony (and why you should care)*. New York: W. W. Norton.

Duimelaar, P. (2017). Relative pitch of tones of several gamelan orchestras in Bali. On website: *Gamelan Bali: Pieter Duimelaar's site on Balinese Gamelan*. Retrieved from http://www.gamelan-bali.eu/pitch_and_tuning.html.

Ellis, A. J. (1885). On the musical scales of various nations. *Journal of the Society of Arts*, *33*(1688), 485–527.

Feng, L., & Wang, X. (2017). Harmonic template neurons in primate auditory cortex underlying complex sound processing. *Proceedings of the National Academy of Sciences USA, 114*(5), E840–E848. doi:10.1073/pnas.1607519114

Fishman, Y. I., Micheyl, C., & Steinschneider, M. (2013). Neural representation of harmonic complex tones in primary auditory cortex of the awake monkey. *Journal of Neuroscience, 33*(25), 10312–10323. doi:10.1523/JNEUROSCI.0020-13.2013.

Fishman, Y. I., Volkov, I. O., Noh, M. D., Garell, P. C., Bakken, H., Arezzo, J. C., … & Steinschneider, M. (2001). Consonance and dissonance of musical chords: Neural correlates in auditory cortex of monkeys and humans. *Journal of Neurophysiology, 86*(6), 2761–2788.

Forster, C. (2010). *Musical mathematics: On the art and science of musical instruments*. San Francisco, CA: Chronicle Books.

Gill, K. Z., & Purves, D. (2009). A biological rationale for musical scales. *PLOS ONE, 4*(12), e8144. doi:10.1371/journal.pone.0008144.

Gockel, H. E., & Carlyon, R. P. (2016). On Zwicker tones and musical pitch in the likely absence of phase locking corresponding to the pitch. *Journal of the Acoustical Society of America, 140*(4), 2257. doi:10.1121/1.4963865.

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America, 54*(6), 1496–1516.

Goldstein, J. L., & Srulovicz, P. (1977). Auditory-nerve spike intervals as an adequate basis for aural frequency measurement. In E. F. Evans & J. P. Wilson (Eds.), *Psychophysics and physiology of hearing* (pp. 337–347). London: Academic Press.

Graves, J. E., Micheyl, C., & Oxenham, A. J. (2014). Expectations for melodic contours transcend pitch. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(6), 2338–2347.

Graves, J. E., & Oxenham, A. J. (2017). Familiar tonal context improves accuracy of pitch interval perception. *Frontiers in Psychology*, *8*, 1753.

Green, B., & Butler, D. (2002). From acoustics to *Tonpsychologie*. In T. Christensen (Ed.), *The Cambridge history of Western music theory* (pp. 246–271). Cambridge: Cambridge University Press.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.

Hartmann, W. M. (1993). On the origin of the enlarged melodic octave. *Journal of the Acoustical Society of America, 93*(6), 3400–3409.

Hartmann, W. M. (1997). *Signals, sound, and sensation*. New York: AIP Press.

Hartmann, W. M. (2013). *Principles of musical acoustics*. New York: Springer.

Heinz, M. G., Colburn, H. S., & Carney, L. H. (2001). Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve. *Neural Computation, 13*(10), 2273–2316.

Heller, E. J. (2013). *Why you hear what you hear: an experiential approach to sound, music, and psychoacoustics*. Princeton, NJ: Princeton University Press.

Hoeschele, M., Weisman, R. G., & Sturdy, C. B. (2012). Pitch chroma discrimination, generalization, and transfer tests of octave equivalence in humans. *Attention, Perception, and Psychophysics, 74*(8), 1742–1760.

Huron, D. B., & Parncutt, R. (1993). An improved model of tonality perception incoporating pitch salience and echoic memory. *Psychomusicology, 12*(2), 154–171.

Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia, 46*(2), 632–639. doi:10.1016/j.neuropsychologia.2007.09.004.

Hyer, B. (1999). Tonality. In T. Christianson (Ed.), *The Cambridge history of Western music theory* (pp. 726–752). Cambridge: Cambridge University Press.

Janata, P. (2004). When music tells a story. *Nature Neuroscience, 7*(3), 203–204.

Janata, P., Birk, J. L., Van Horn, J. D., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The cortical topography of tonal structures underlying Western music. *Science, 298*(5601), 2167–2170.

Johnson, D. H. (1978). The relationship of post-stimulus time and interval histograms to the timing characteristics of spike trains. *Biophysical Journal, 22*(3), 413–430.

Johnson, D. H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *Journal of the Acoustical Society of America, 68*, 1115–1122.

Johnston, I. D. (2009). *Measured tones: The interplay of physics and music* (3rd ed.). Boca Raton, FL: CRC Press.

Justus, T., & Hutsler, J. J. (2005). Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain specificity. *Music Perception, 23*(1), 1–28.

Kameoka, A., & Kuriyagawa, M. (1969a). Consonance theory I. Consonance of dyads. *Journal of the Acoustical Society of America, 45*(6), 1451–1459.

Kameoka, A., & Kuriyagawa, M. (1969b). Consonance theory II. Consonance of complex tones and its calculation method. *Journal of the Acoustical Society of America, 45*(6), 1460–1469.

Kendall, R. A., & Carterette, E. C. (1996). Music perception and cognition. In M. P. Friedman & E. C. Carterette (Eds.), *Cognitive ecology*. *Handbook of perception and cognition* (2nd ed, pp. 87–150). San Diego, CA: Academic Press.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Krumhansl, C. L. (1991). Music psychology: tonal structures in perception and memory. *Annual Review of Psychology, 42*(1), 277–303.

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin, 126*(1), 159–179.

Krumhansl, C. L., & Cuddy, L. L. (2008). A theory of tonal hierarchies in music. In M. R. Jones (Ed.), *Music perception: Springer handbook of auditory research*(Vol. 36, pp. 51–87). New York: Springer.

Langner, G., & Benson, C. (2015). *The neural code of pitch and harmony*. Cambridge: Cambridge University Press.

Large, E. W., Kim, J. C., Bharucha, J. J., & Krumhansl, C. L. (2016). A neurodynamic account of musical tonality. *Music Perception, 33*(3), 319–331.

Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2014). Neural transformation of dissonant intervals in the auditory brainstem. *Music Perception, 32*(5), 445–459.

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception, 17*(4), 481–510.

Leman, M., & Carreras, F. (1997). Schema and Gestalt: Testing the hypothesis of psychoneural isomorphism by computer simulation. In M. Leman (Ed.), *Music, Gestalt, and computing* (pp. 144–165). Berlin: Springer.

Lester, J. (1999). Rameau and eighteenth-century harmonic theory. In T. Christianson (Ed.), *The Cambridge history of Western music theory* (pp. 753–777). Cambridge: Cambridge University Press.

Licklider, J. C. R. (1951). A duplex theory of pitch perception. *Experientia, 7*(4), 128–134.

Licklider, J. C. R. (1959). Three auditory theories. In S. Koch (Ed.), *Psychology: A study of a science. Study I. Conceptual and systematic* (Vol. 1, pp. 41–144). New York: McGraw-Hill.

Lipps, T. ([1905] 1995). *Consonance and dissonance in music* (W. Thompson, Trans.). San Marino, CA: Everett Books.

Loui, P. (2012). Statistical learning—What can music tell us? In P. Rebuschat & J. Williams (Eds.), *Statistical learning and language acquisition* (pp. 433–462). Berlin: Mouton de Gruyter.

Lyon, R., & Shamma, S. (1996). Auditory representations of timbre and pitch. In H. Hawkins, T. McMullin, A. N. Popper, & R. R. Fay (Eds.), *Auditory computation* (pp. 221–270). New York: Springer-Verlag.

McDermott, J. H., Keebler, M. V., Micheyl, C., & Oxenham, A. J. (2010a). Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special. *Journal of the Acoustical Society of America, 128*(4), 1943–1951. doi:10.1121/1.3478785.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010b). Individual differences reveal the basis of consonance. *Current Biology, 20*(11), 1035–1041. doi:10.1016/j.cub.2010.04.019.

McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology, 18*(4), 452–463.

McKinney, M. (1999). A possible neurophysiological basis of the octave enlargement effect. *Journal of the Acoustical Society of America, 106*(5), 2679–2692.

McKinney, M. F. (2001). *Neural correlates of pitch and roughness: Toward the neural code for melody and harmony perception* (Doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

McKinney, M. F., Tramo, M. J., & Delgutte, B. (2001). Neural correlates of musical dissonance in the inferior colliculus. In A. J. M. Houtsma (Ed.), *Physiological and psychophysical bases of auditory function* (pp. 71–77). Maastricht: Shaker Publishing.

Meddis, R., & Hewitt, M. (1991). Virtual pitch and phase sensitivity studied of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America*, *89*(6), 2866–2882.

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

Midya, V. (2015). On mathematical functions for theoretical and experimental distributions for shrutis (micro tonal intervals) and their application in Hindustani music. *eprint arXiv:1502.03679*. Retrieved from https://arxiv.org/pdf/1502.03679.pdf.

Miyazaki, K. (1992). Perception of musical intervals by absolute pitch possessors. *Music Perception, 9*(4), 413–426.

Miyazaki, K. I. (2004). Recognition of transposed melodies by absolute-pitch possessors. *Japanese Psychological Research, 46*(4), 270–282.

Moore, B. C., & Ernst, S. M. (2012). Frequency difference limens at high frequencies: Evidence for a transition from a temporal to a place code. *Journal of the Acoustical Society of America, 132*(3), 1542–1547. doi:10.1121/1.4739444.

Moore, B. C. J. (1980). Neural interspike intervals and pitch. *Audiology, 19*(4), 363–365.

Moore, B. C. J. (2013). *An introduction to the psychology of hearing* (6th ed.). Leiden: Brill.

Nettl, B. (2015). *The study of musicology: Thirty-three discussions*. Urbana: Illinois University Press.

Norman-Haignere, S., Kanwisher, N., & McDermott, J. H. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *Journal of Neuroscience, 33*(50), 19451–19469. doi:10.1523/JNEUROSCI.2880-13.2013.

Ohgushi, K. (1978). On the role of spatial and temporal cues in the perception of the pitch of complex tones. *Journal of the Acoustical Society of America, 64*, 764–771.

Ohgushi, K. (1983). The origin of tonality and a possible explanation for the octave enlargement phenomenon. *Journal of the Acoustical Society of America, 73*(5), 1694–1700.

Ohgushi, K., & Ano, Y. (2005). The relationship between musical pitch and temporal responses of the auditory nerve fibers. *Journal of Physiological Anthropology and Applied Human Science, 24*(1), 99–101.

Oxenham, A. J., Bernstein, J. G., & Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proceedings of the National Academy of Sciences USA, 101*(5), 1421–1425.

Palmer, A. R. (1992). Segregation of the responses to paired vowels in the auditory nerve using autocorrelation. In M. E. H. Schouten (Ed.), *The auditory processing of speech: From sounds to words* (pp. 115–124). Berlin: Mouton de Gruyter.

Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin: Springer-Verlag.

Parncutt, R. (2011a). Consonance and dissonance in music theory and psychology: Disentangling dissonant dichotomies. *Journal of Interdisciplinary Music Studies, 5*(2), 119–166.

Parncutt, R. (2011b). The tonic as triad: Key profiles as pitch salience profiles of tonic triads. *Music Perception, 28*(4), 333–365.

Partch, H. (1974). *Genesis of a music: An account of a creative work, its roots and its fulfillments* (2nd ed.). New York: Da Capo.

Patterson, R. D. (1986). Spiral detection of periodicity and the spiral form of musical scales. *Psychology of Music, 14*(1), 44–61.

Peretz, I. (2013). The biological foundations of music: Insights from congenital amusia. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 141–182). New York: Academic Press.

Peretz, I. (2016). Neurobiology of congenital amusia. *Trends in Cognitive Sciences, 20*(11), 857–867. doi:10.1016/j.tics.2016.09.002.

Perkell, D. H., & Bullock, T. H. (1968). Neural coding. *Neurosciences Research Program Bulletin, 6*(3), 221–348.

Picken, L. (1957). The music of far eastern Asia. I. China II. Other countries. In E. Wellez (Ed.), *New Oxford history of music. Ancient and oriental music* (Vol. 1, pp. 83–194). New York: Oxford University Press.

Pierce, J. R. (1992). *The science of musical sound* (rev. ed.). New York: W. H. Freeman.

Piston, W., & DeVoto, M. (1987). *Harmony* (5th ed.). New York: W. W. Norton.

Plack, C. J., Barker, D., & Hall, D. A. (2014). Pitch coding and pitch processing in the human brain. *Hearing Research*, *307*(January), 53–64.

Plack, C. J., & Oxenham, A. J. (2005). Psychophysics of pitch. In C. J. Plack, A. J. Oxenham, A. N. Popper, & R. Fay (Eds.), *Pitch: Neural coding and perception*. New York: Springer-Verlag.

Plomp, R. (1976). *Aspects of tone sensation*. London: Academic Press.

Plomp, R., & Levelt, W. J. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, *38*(4), 548–560.

Pressnitzer, D., Patterson, R. D., & Krumbholz, K. (2001). The lower limit of melodic pitch. *Journal of the Acoustical Society of America*, *109*(5 Pt. 1), 2074–2084.

Rameau, J. P. ([1722] 1971). *Treatise on harmony* (P. Gossett, Trans.). New York: Dover.

Rasch, R. (1999). Tuning and temperament. In T. Christianson (Ed.), *The Cambridge history of Western music theory* (pp. 193–222). Cambridge: Cambridge University Press.

Rasch, R., & Plomp, R. (1999). The perception of musical tones. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 89–112). San Diego, CA: Academic Press.

Rayleigh, J. W. S. ([1894] 1945). *The theory of sound* (2nd ed.). London: Dover.

Révész, G. Z. ([1954] 2001). *Introduction to the psychology of music* (G. I. C. de Courcy, Trans.). New York: Dover.

Riemann, H. ([1905] 2011). The problem of harmonic dualism (I. Bent, Trans.). In E. Gollin & A. Rehding (Eds.), *The Oxford handbook of neo-riemannian music theories* (pp. 170–193). New York: Oxford University Press. [First published in 1905 in German as “Das Problem des harmonischen Dualismus.” *Neue Zeitschrift für Musik*, *51*, pp. 69–70.]

Rossing, T. D., Wheeler, P., & Moore, F. R. (2002). *The science of sound* (3rd ed.). San Francisco, CA: Addison Wesley.

Ruckmick, C. C. 1929. A new classification of tonal qualities. *Psychological Review*, *36*(2), 172–180.

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences USA, 112*(29), 8987–8992. doi:10.1073/pnas.1414495112.

Schellenberg, E. G., & Trehub, S. E. (1996). Natural musical intervals: Evidence from infant listeners. *Psychological Science*, *7*(5), 272–277.

Schneider, A. (1997). “Verschmelzung,” tonal fusion, and consonance: Carl Stumpf revisited. In M. Leman (Ed.), *Music, gestalt, and computing: Studies in cognitive and systematic musicology* (pp. 115–143). Berlin: Springer.

Schwartz, D. A., Howe, C. Q., & Purves, D. (2003). The statistical structure of human speech sounds predicts musical universals. *Journal of Neuroscience, 23*(18), 7160–7168.

Schwartz, D. A., & Purves, D. (2004). Pitch is determined by naturally occurring periodic sounds. *Hearing Research, 194*(1–2), 31–46. doi:10.1016/j.heares.2004.01.019.

Sek, A., & Moore, B. C. (1995). Frequency discrimination as a function of frequency, measured in several ways. *Journal of the Acoustical Society of America, 97*(4), 2479–2486.

Semal, C., & Demany, L. (2006). Individual differences in the sensitivity to pitch direction. *Journal of the Acoustical Society of America, 120*(6), 3907–3915.

Sethares, W. A. (2005). *Tuning, timbre, spectrum, scale* (2nd ed.). London: Springer.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America, 36*(12), 2346–2353.

Shepard, R. N. (1982a). Geometrical approximations to the structure of musical pitch. *Psychological Review, 89*(4), 305–333.

Shepard, R. N. (1982b). Structural representations of musical pitch. In D. Deutch (Ed.), *The psychology of music* (1st ed., pp. 341–390). New York: Academic Press.

Siebert, W. M. (1968). Stimulus transformations in the peripheral auditory system. In P. A. Kolers & M. Eden (Eds.), *Recognizing patterns* (pp. 104–133). Cambridge, MA: MIT Press.

Slaney, M., & Lyon, R. F. (1993). On the importance of time—A temporal representation of sound. In M. Cooke, S. Beet, & M. Crawford (Eds.), *Visual representations of speech signals* (pp. 95–118). New York: John Wiley.

Srulovicz, P., & Goldstein, J. L. (1983). A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum. *Journal of the Acoustical Society of America, 73*(4), 1266–1276.

Stumpf, C. (1883/1890). *Tonpsychologie*, Vols. 1 & 2. Leipzig: S. Hirzel.

Sundberg, J. (1994). Acoustic and psychoacoustic aspects of vocal vibrato. *STL-QPSR, 35*(2–3), 45–68.

Temperley, D. (2007). *Music and probability*. Cambridge, MA: MIT Press.

Tenney, J. (1988). *A history of consonance and dissonance*. New York: Excelsior.

Terhardt, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America, 55*(May), 1061–1069.

Terhardt, E. (1977). The two-component theory of musical consonance. In E. F. Evans & J. P. Wilson (Eds.), *Psychophysics and physiology of hearing* (pp. 381–390). London: Academic Press.

Terhardt, E. (1984). The concept of musical consonance: A link between music and psychoacoustics. *Music Perception*, *1*(3), 276–295.

Thompson, W. F. (2013). Intervals and scales. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 107–140). New York: Elsevier.

Trainor, L. J., & Hannon, E. E. (2013). Musical development. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 423–497). New York: Elsevier.

Trainor, L. J., & Unrau, A. (2011). Development of pitch and music perception. In L. A. Werner (Ed.), *Human auditory development: Springer handbook of auditory research* (pp. 223–253). New York: Springer.

Tramo, M. J., Cariani, P. A., Delgutte, B., & Braida, L. D. (2001). Neurobiological foundations for the theory of harmony in Western tonal music. *Annals of the New York Academy of Sciences, 930*(1), 92–116.

Trehub, S. E., Becker, J., & Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 370*(1664), 20140096.

Ueda, K., & Ohgushi, K. (1987). Perceptual components of pitch: Spatial representation using a multidimensional scaling technique. *Journal of the Acoustical Society of America, 82*(4), 1193–1200.

Umemoto, T. (1990). The psychological structure of music. *Music Perception, 8*(2), 115–127.

van Noorden, L. (1982). Two channel pitch perception. In M. Clynes (Ed.), *Music, mind and brain* (pp. 251–269). New York: Plenum.

Verschooten, E., Robles, L., & Joris, P. X. (2015). Assessment of the limits of neural phase-locking using mass potentials. *Journal of Neuroscience, 35*(5), 2255–2268. doi:10.1523/JNEUROSCI.2979-14.2015.

von Helmholtz, H. ([1885] 1954). *On the sensations of tone as a physiological basis for the theory of music* (A. J. Ellis, Trans.). New York: Dover.

Vurma, A., & Ross, J. (2006). Production and perception of musical intervals. *Music Perception*, *23*(4), 331–344.

Ward, D. (1999). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 265–298). New York: Academic Press.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences USA, 100*(17), 10038–10042.

Warren, R. M. (1999). *Auditory perception: A new analysis and synthesis*. Cambridge: Cambridge University Press.

Warren, R. M., & Bashford, J. A., Jr. (1981). Perception of acoustic iterance: Pitch and infrapitch. *Perception & Psychophysics, 29*(4), 395–402.

Wightman, F. L. (1973). The pattern-transformation model of pitch. *Journal of the Acoustical Society of America*, *54*(2), 407–416.

Zilany, M. S., Bruce, I. C., & Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *Journal of the Acoustical Society of America, 135*(1), 283–286. doi:10.1121/1.4837815.

Zwicker, E., & Fastl, H. (1999). *Psychoacoustics: Facts and models* (2nd ed.). New York: Springer-Verlag,