DALE PURVES

Music *as* Biology

The Tones We Like and Why

Music *as* Biology

The Tones We Like and Why

DALE PURVES

HARVARD UNIVERSITY PRESS

Cambridge, Massachusetts

London, England

2017

Jacket art: Courtesy of Thinkstock

Jacket design: Tim Jones

978-0-674-54515-1 (alk. paper)

978-0-674-97296-4 (EPUB)

978-0-674-97297-1 (MOBI)

The Library of Congress has cataloged the printed edition as follows:

Names: Purves, Dale, author.

Title: Music as biology : the tones we like and why / Dale Purves.

Description: Cambridge, Massachusetts : Harvard University Press, 2017. | Includes bibliographical references and index.

Identifiers: LCCN 2016020279

Subjects: LCSH: Music—Physiological aspects. | Musical perception. | Music and science.

Classification: LCC ML3820 .P87 2017 | DDC 781.1—dc23

LC record available at https://lccn.loc.gov/2016020279

FOR READERS WHO LOVE MUSIC BUT ARE MYSTIFIED BY MUSIC THEORY

Contents

Preface

The aim of this book is to argue that when considered in terms of biology, the whys and wherefores of tonal music are easier to understand than when considered in terms of mathematics, physics, psychology, or conventional music theory. The centerpiece of the argument is the ecological benefits that arise from recognizing and responding to conspecific vocalization. While biology provides a limited perspective for tackling a topic as complex as musical tones and their effects on listeners, there is a good reason for this stance: the sources of tones in the world in which we evolved are almost exclusively animate, and the animals whose tonal utterances are most important to us are other human beings. The foundation of the argument is thus that the human sense of tonality and its current uses in speech and music have arisen to extract information—and there is plenty of it—in human vocalizations.

Even if one accepts this thesis and the evidence for it, musicologists may well take umbrage at the idea that tonal music can be understood in biological terms. This approach, however, is no more antithetical to music appreciation than understanding visual perception as biology is to an appreciation of painting and sculpture. The chief advantage of music over other art forms is the extraordinary opportunity provided by musical preferences established over centuries across different cultures and traditions. Whereas visual aesthetics obviously exist—why else art museums?—preferences in this domain are difficult to codify, a number of attempts notwithstanding. In contrast, musical practice and theory over millennia have amply documented the tones and tone combinations humans worldwide like to play and hear.

An even better reason for focusing on tones and their relationships is that tonal phenomenology, while central to music, has never been fully explained. The tone combinations that humans find consonant, the limited number of scales used to divide octaves, the small numbers of notes in scales, the importance of a reference tone, the uniqueness of octaves, the emotions elicited by particular scales, and the variations in tonal palettes across cultures raise many unresolved questions. The gist of what follows is that musical tones examined from the perspective of biology provide plausible answers.

Finally, the intended audience. My hope is that anyone interested in music—professionals, beginners, and those with little or no musical training—will get something out of the book. Thus, I have tried to steer a course between too much detail and too little, perhaps not always successfully. Although the basics are provided, I have not attempted to distill the voluminous literature on acoustics, auditory physiology and anatomy, music theory, and music history. In addition to balance, I wanted to keep the focus on the core of the argument. For those who wish to dig deeper, recommended reviews, books, and articles are given in footnotes and at the end of each chapter, and a more complete bibliography at the end of the book.

As a dedicated but mediocre musician, it has been fun as well as useful to think hard about these issues. Whether in the end readers deem the arguments about music right or wrong, I trust they will have some fun as well.

1

Sound Signals and Sound Stimuli

A THEME THROUGHOUT THESE CHAPTERS is that the demands of human biology, rather than the physical world, determine what we hear and that music provides one way to make this case. Whereas vision research began to move beyond the idea of seeing physical reality in the nineteenth century with Hermann von Helmholtz's theory of "unconscious inferences" arising from experience, auditory science has been more closely allied with the idea that the goal of sensory systems is to recover and represent the physical properties of the environment. Like the visual system, however, the auditory system has no access to the physical parameters of the objects and conditions in the world that give rise to sound signals. Thus, defining what we hear in terms of mechanical energy at the ear is a good start, but ultimately misleading. A logical place to begin thinking about this issue is to consider the reasons why sensory systems are saddled with this quandary; how the quandary might be resolved; and how the apparent resolution bears on auditory preferences, tonal music in particular. To do so, however, requires reviewing some basic facts about sound and the human auditory system.

The Transformation of Sound Signals

Sound signals are physically determined patterns of mechanical energy in the local environment that exist whether or not there is someone around to hear them. The transformation of sound signals into biologically determined *sound stimuli* requires a listener and begins by an enhancement of pressure changes at the ear that fall within a range that is relevant to humans (Figure 1.1; see the Appendix for a primer on the human auditory system). The peculiar shape of the outer ear, the dimensions of the ear canal and the bony apparatus of the middle ear all contribute to filtering and amplifying the sound signals we receive to better succeed in the world, which in biology eventually boils down to reproducing successfully.

The inner ear, or *cochlea,* is where this modified mechanical energy is transformed into neural signals. The major player in the inner ear is an elongated strip called the *basilar membrane*, which lies coiled within a bony shell about the size of a pea. The basilar membrane vibrates in response to movements of the cochlear fluid initiated by the back and forth movement of the middle ear bones at the oval window. Sitting atop the basilar membrane are neural receptors called *hair cells* each with cilia ("hairs") protruding from their apices into an overlying gelatinous strip called the tectorial membrane. About 3000 *inner hair cells* in each cochlea detect this mechanical displacement, whereas about 12,000 *outer hair cells* sharpen the location of displacement along the basilar membrane by changing its compliance as a result of innervation (feedback) that arises centrally. The displacement of the cilia changes the membrane potential of hair cells, releasing neurotransmitter molecules onto the endings of auditory nerve axons that carry the resulting location and pattern of hair cell displacement to the central auditory nervous system.

A sound signal can be perceived when the resulting pressure changes displace the cilia on the inner hair cells by as little as 0.3 nm, which is about the diameter of an atom of gold. The force required to elicit such movements is equally minuscule, on the order of 10^{-12} newtons / m^2. At the threshold of human hearing, this force corresponds to a displacement of about 10^{-11} meters, the power involved being on the order of a trillionth of a watt / m^2. The upper end of the audible range of human hearing is about than 10^{12} times greater than this threshold, and it is limited by the damage done to hair cells by very high-energy sound signals.

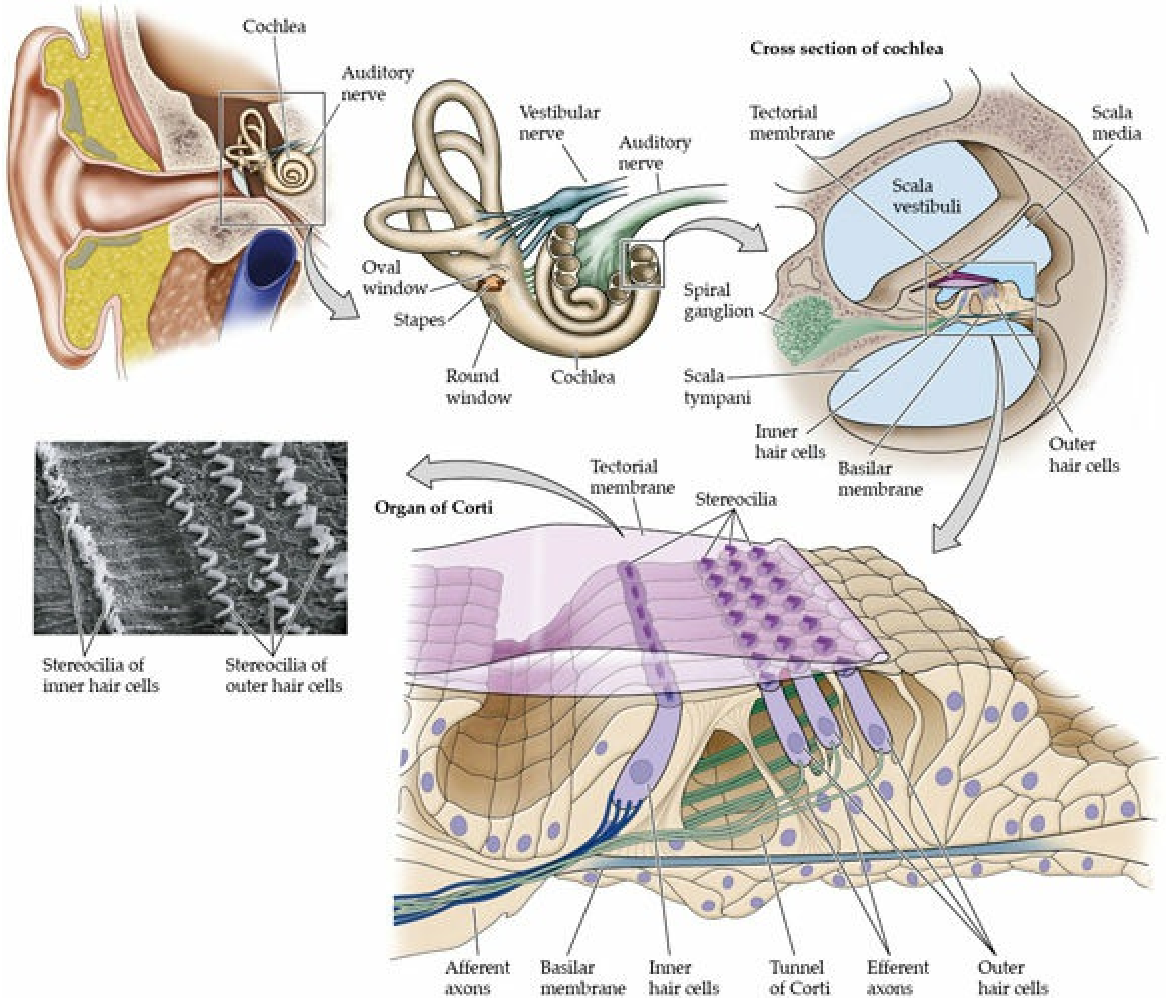


FIGURE 1.1 The peripheral auditory apparatus. This series of images shows the outer, middle, and inner ear, with the cochlea cut in cross section. Blowups show the basilar membrane and the receptors (hair cells) that initiate auditory processing and perception when action potentials are elicited in the afferent axons of the auditory nerve. (From Purves et al., 2012.)

The pressure changes of sound signals at the ear consist of rarefactions and condensations of the local atmosphere (Figure 1.2). If these variations are periodic, as in the figure, they are described in terms of the *frequency* of repetition in cycles per second, or Hertz, 1 Hz being one cycle per second. Young adults can hear frequencies of up to about 20,000 Hz. With age, however, sensitivity to higher frequencies diminishes because of the damage suffered by the non-regenerating hair cells over a typical lifetime. Other mammals have higher (or lower) response ranges, depending on their size and biological needs. Some bats, for instance, are sensitive to frequencies of up to 200,000 Hz, with a lower limit at about the upper limit of human hearing.

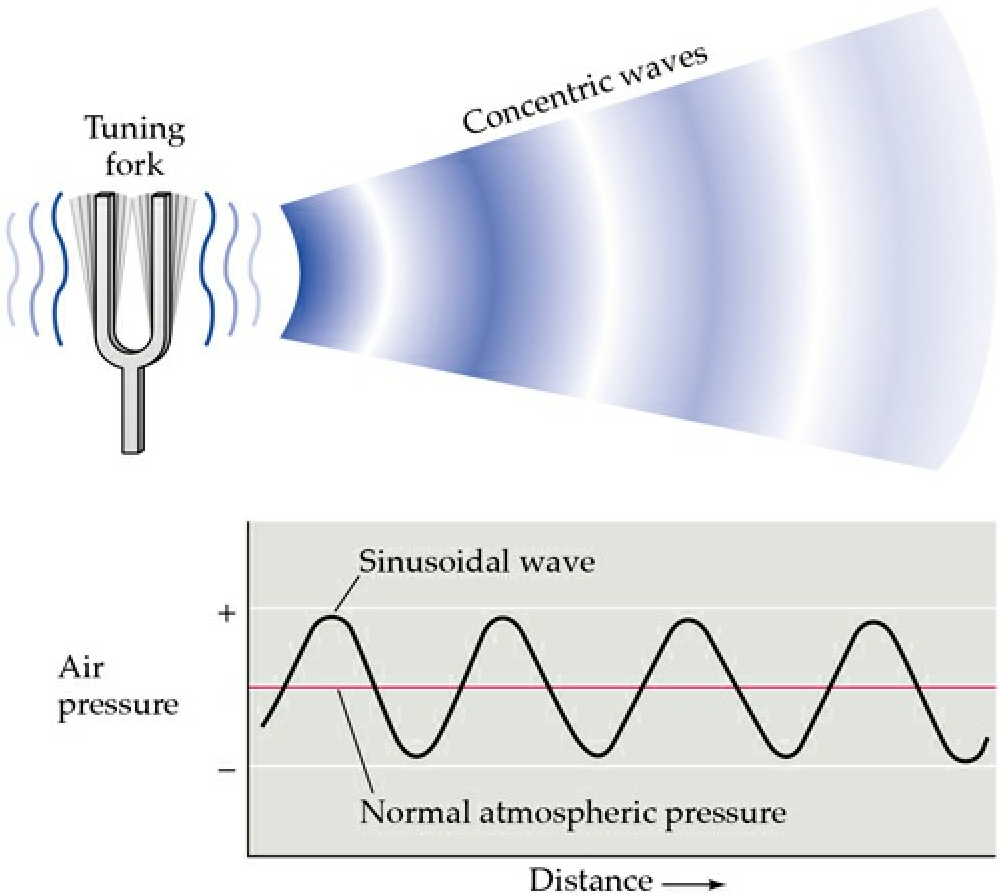


FIGURE 1.2 The sound signal generated by a tuning fork. When the tines of the fork move in one direction, they create a local increase in air density, and when they move in the other direction a local rarefaction. The resulting series of compressions and rarefactions—changes in pressure—produce a highly regular sound signal called a *sine wave.* Like other wave phenomena, sound waves can be described in terms of their form (e.g., simple, like a sine wave, or complex, like most natural sound signals); their frequencies (in Hz); their amplitudes (in decibels); and their phase (the wave considered in terms of 360 degrees over one cycle). Rather than being carried along a two-dimensional surface like the ripples on a pond, however, sound waves propagate in three-dimensional space, creating spherical shells of pressure change emanating in all directions. (From Purves et al., 2012.)

Based on studies using sine waves, the low-frequency end of the range of human hearing is usually given as 20 Hz. The reason we can't hear lower-frequency sine waves is their slow rise time; that is, they don't change quickly enough to displace the cilia on the hair cells, and thus they fail to produce action potentials in the auditory nerve and sound percepts. This slowness is why we don't hear atmospheric pressure changes associated with the weather. We can, however, hear a single, more or less instantaneous pressure change (a "click") that has no recurring frequency. Moreover, if the amplitude of sine waves with frequencies less than 20 Hz is increased, lower frequencies can be heard.

Natural Sound Signals

As implied, audition has most often been studied using *sine waves* (also called "pure tones”) because of their conceptual and mathematical simplicity. Indeed, the popularity of sine-wave stimuli is one reason that physical paradigms have continued to dominate much auditory science. However, atmospheric disturbances generated by vibrations at a single frequency within the range of human hearing rarely, if ever, occur in nature. Thus, sine waves can be a misleading tool when thinking about the sound signals we evolved to hear, or how the auditory system actually operates in linking objectively measurable sound signals to subjective auditory perceptions.

The pressure changes that typically trigger auditory responses are caused by objects that vibrate in far more complex ways than tuning forks. The sound signals we hear as rattles, squeaks, scratches, thumps, clicks, and the like have little or no orderly repetition and are loosely categorized as *noise*. If, on the other hand, the complex pressure changes in the local atmosphere repeat at rates between about 50 and 5000 Hz (see Chapter 3) and are sufficiently intense, the result for the listener is the perception of a *tone*. Their complexity notwithstanding, tones are heard as a single *pitch*, rather than as a bumpy series of peaks and valleys, as occurs when the repetition rate is below this more or less middle range of human hearing. If the signal exceeds about 5000 Hz, as it often does in whistles, sirens, or shrieks, we can still identify a pitch, but not one that is considered tonal (or musical).

Sources of Tones

Tones can have a single frequency, as in the case of sine waves arising from the tuning fork in Figure 1.2. Far more common, however, are complex signals with multiple frequencies that nonetheless repeat systematically. A variety of objects produce such complex vibrations when acted on by a force; musical examples include a taut string on a string instrument, or a column of air in a wind instrument. The *resonance* of such objects—defined as their natural mode of vibration when disturbed by a force—is determined by mass, material composition, shape, and tension. These factors combine to account for the difference between the pitch of tones produced by a low string or a high string on a guitar or other string instrument.

Manufactured tonal sources have existed as musical instruments for thousands of years. But in natural auditory environments, tonal sounds arise almost exclusively from the signaling apparatus of animals, providing a basis for distinguishing animate from inanimate sound sources. Animate tones include the stridulations of insect appendages, the calls of amphibians (e.g., frogs), the songs of many birds, and the vocal sounds of mammals, including humans, all of which provide a means of social communication. Inanimate mechanical forces (e.g., wind and water) acting on resonant structures rarely combine to produce periodic sound signals. Thus, animals—our fellow humans in particular—are the principal source of periodic sound energy in the environments in which we evolved. As discussed in Chapters 3 and 4, this fact takes on particular importance in understanding music and why we like it.

Sound Signal Spectra and Harmonic Series

Periodic sound signals, or indeed any sound signal, can be characterized by its *spectrum.* A spectrum is a graphical representation of the distribution of a signal's intensity (amplitude) plotted on the y-axis as a function of frequency on the x-axis (Figure 1.3). Spectra are usually determined by Fourier analysis, a mathematical tool that decomposes signals into their component frequencies. Since natural sound signals typically change over time, spectra are derived from sound signals sampled over a relatively brief interval (e.g., 100 milliseconds); they can thus be thought of as a "snapshot" of how the energy in a sound signal is distributed. It is important to distinguish this more or less momentary analysis of energy distribution as a function of frequency from the analysis of a signal over an extended time (often referred as the "time signal"). For example, the peaks of the waves in Figure 1.2 recur over ongoing time at some frequency; the spectrum of the signal sampled over some definite time interval would have a single peak at the frequency of its repetition rate.

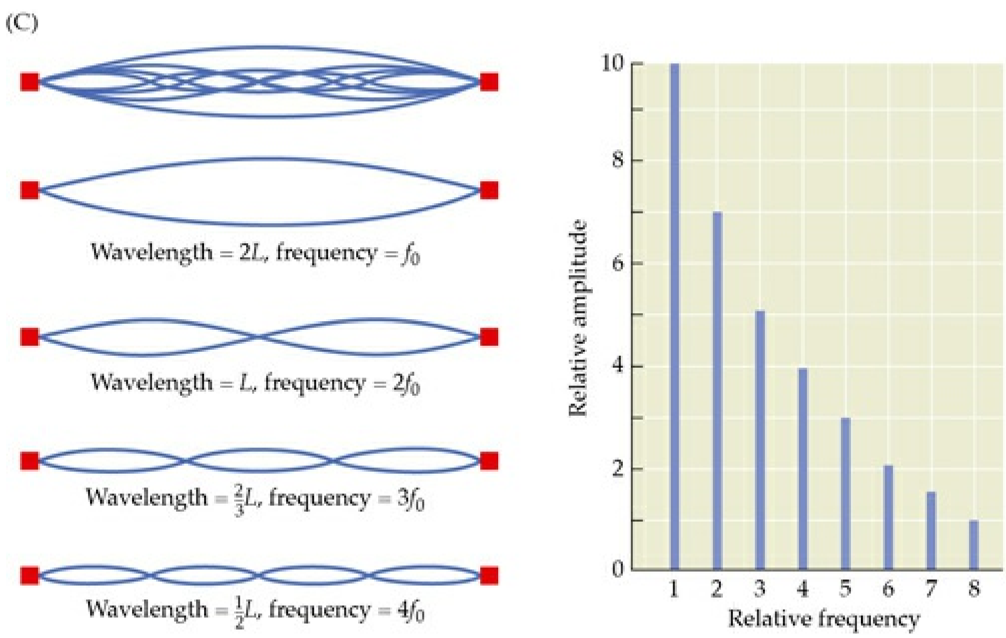


FIGURE 1.3 The harmonic series evident in the spectrum of a vibrating string. The illustration on the left indicates the multiple modes of vibration of a plucked string fixed at both ends (as for any string instrument) and the relative amplitudes of the excursions in the standing wave that results. The graph on the right shows the spectrum determined by Fourier analysis. The numbers on the abscissa indicate successive harmonics in the harmonic series; "1" indicates the vibration that entails the full length of the string, called the fundamental frequency, "2" the vibration at half of the full length, "3" the vibration

at a third of the full and so on. As indicated, the amplitudes of the excursions decline progressively over the harmonic series. (From Purves et al., 2012.)

A musical example of a sound signal spectrum is the vibration of a taut sting, as shown in Figure 1.3. When plucked, a *standing wave* is generated that vibrates in a series of modes due to the reflection of the waves at the fixed ends and their interaction. The greatest excursion occurs over the full length of the string and generates the most energetic component in the spectrum. The frequency of this mode is referred to as the *fundamental frequency* of vibration (somewhat confusingly called "F0" in the literature, despite the fact that it is the first harmonic in the series). The next most powerful mode of vibration is at half the length of the string, the next at a third of the length, the next at a quarter of the length and so on, thus forming a *harmonic series.* These modes occur because the reflection of the wave from the ends of the string creates a series of "nodes" that cancel out all vibratory frequencies that are not integer multiples of the frequency of the full length of the string (think of water waves bouncing back and forth between the sides of a bath tub). Since natural signals that produce periodic sound stimuli—the human vocal apparatus in particular—tend to vibrate in a similar manner, a taut string and the harmonic series it produces is a far more relevant model of the natural periodic signals in music and animal vocalization than the tuning fork and the sine waves it generates.

In determining a spectrum, the sound signal of interest must be sampled over a window of time appropriate to its component frequencies. For example, in Figure 1.3 the window would have to be at least twice as long as the time between the repeating cycle of the first harmonic of the vibrating string (the fundamental frequency). If the string in the figure had a fundamental frequency of 440 Hz (middle, or "concert," A on a piano) this interval would about 2.3 milliseconds. The sampling window would therefore be at least 5 milliseconds, since a shorter sample would not include the lowest harmonic frequency. A different problem arises with respect to sampling the high frequencies in a spectrum. In this case the *rate* of sampling would have be at least twice the frequency of highest frequency to avoid the "aliasing" artifacts that arise when the sampling rate is too slow (referred to as "undersampling"). A typical sampling rate for signals pertinent to human hearing is 44 kHz, which is somewhat more than twice the nominal 20 kHz upper limit of hearing in young adults.

Finally, keep in mind that a harmonic series is both an arithmetic and a geometric progression: the sequence of harmonic numbers 1, 2, 3, 4, … *n* is an arithmetic progression, whereas the way we hear tones (pitch) is a geometrical progression (e.g, 1,2,4,8, … *n*). Thus, the same perceived change of a low-frequency sound signal entails a smaller change in frequency than does a high-frequency sound signal.

Noise

Tonal sounds are especially important in the chapters that follow. Nevertheless, focusing on signals that generate periodicity (vibrating strings, air columns, animal vocalizations) can also be misleading. As already mentioned, most natural stimuli—a babbling brook, a breaking twig, rustling leaves—have little or no periodicity, and are lumped together under the somewhat pejorative rubric of *noise.* Noise can be ongoing, as in the examples of running water and wind, or essentially instantaneous, as in the snap of a breaking twig or a thunderclap arising from the brief displacement of air molecules by heat energy.

If the distribution of amplitudes as a function of frequency is perfectly uniform, the signal is called *white noise.* By definition, then, the spectrum of white noise shows no periodicity at all, the energy at every frequency over the range of human hearing being equal. The opposite extreme is a sine wave, in which the energy in the signal is narrowly distributed at a single frequency. Vision is bound by two similar extremes: "white light" refers to a distribution of light energy that is uniform enough to elicit the visual perception of whiteness, whereas monochromatic light has a single colored energy peak, similar to the spectrum of a sine tone.

Obstacles in Determining the Sources of Sound Signals

The pressure changes at the ear that we experience in perceiving sound stimuli are determined by a variety of physical factors (Figure 1.4). The major contributors are: (1) the mechanical force that acts on an object capable of generating a sound signal; (2) the properties of the object that determine its resonance when a force is applied (e.g., mass, material composition, density, shape, tension); and (3) how conditions in the local environment influence the sound signal that reaches the listener (e.g., the decline in signal intensity that occurs inversely with the square of distance traveled, and the absorbance and reflection of the signal by local objects). In addition, interactions with concurrent sound sources with different phases can amplify or reduce the pressure changes that reach the ear. All these factors underscore the challenge of responding appropriately to sound signals. The peripheral auditory system lacks any obvious means to disentangle these messy, conflated factors, let alone measure their contributions independently. Nonetheless, we routinely behave in response to sound signals as if we "knew" this information.

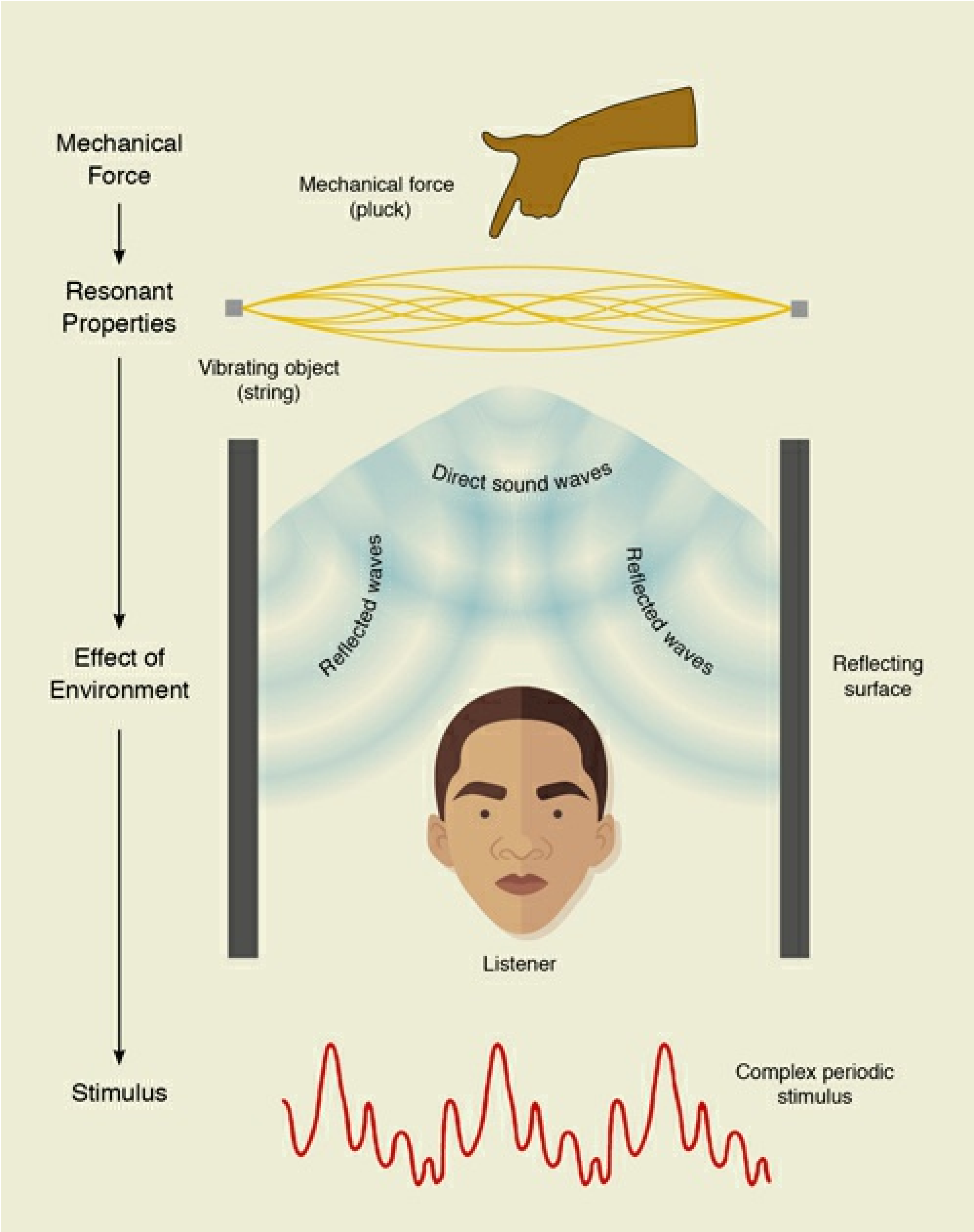


FIGURE 1.4 The major factors that determine sound signals at the listener's ear. A variety of physical interactions affect sound signals at the ear, making it difficult to understand how listeners can use this conflated information to sort out the sources of sound signals in the environment and behave appropriately.

Although what we hear seems to directly represent the sources of sound signals in the environment, understanding how and why we hear what we do presents a fundamental problem. The quandary is how a biological sensing system like audition, which has no way to disentangle the physical parameters of sound signal sources, can nonetheless generate successful behavior.

An Empirical Approach

These confounding observations mean that auditory perceptions must be generated in some way other than by recovering and representing the physical sources of sound signals. As indicated in Figure 1.4, the information is simply not available in sound stimuli, despite the fact that physical causes obviously determine the nature of sound signals.

Some indication of how this fundamental problem for sensory systems may be resolved has emerged in vision research. In vision, trial-and-error responses assign perceptual values empirically, a strategy that circumvents the need to represent the physical properties of objects and conditions in the world as such. By using feedback from what works, stimuli can elicit useful responses, despite the absence of the objective measurements that biological sensors are unable to provide (see Chapter 9). This strategy as it applies to audition is diagrammed in Figure 1.5.

In general terms, the biology underlying this way of thinking about what we hear and why is well known: it is simply Darwinian principles in auditory action. Random changes in the organization and function of ancestral auditory apparatus and neural circuitry persisted—or not—according to how well a given variation served the survival and reproductive success of the animals whose preneural apparatus and neural circuitry harbored that variant. Any configuration of peripheral auditory structure and central neural processing that mediated successful responses to sound signals would eventually disseminate throughout the population, whereas less useful auditory apparatus, circuit configurations, and operations would not. This biological feedback loop would progressively organize the basic *qualities* we perceive (loudness, pitch and timbre; see Chapter 2) according to the empirical occurrence of stimulus patterns. As explained in Chapter 9, the strategy works because it establishes an objective-to-subjective mapping via biological machinery that does not depend on measurements of physical reality. The role of the physical world in this understanding of sensory neurobiology is simply an arena in which preneural apparatus, neural circuitry, and perceptions are empirically tested by the criterion of reproductive success.

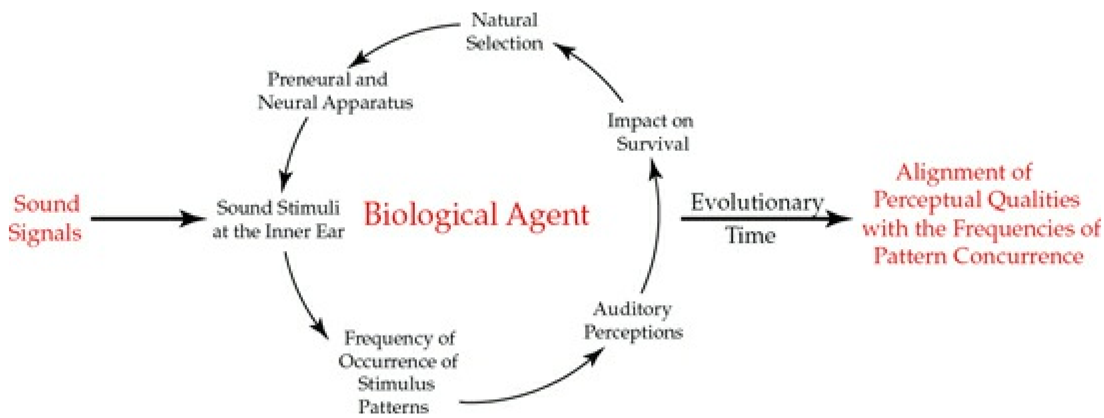


FIGURE 1.5 Auditory perception based on feedback from experience. As explained in Chapter 9, the frequency of occurrence of different sound signal patterns provides a link between objective and

subjective domains that can promote survival and reproductive success without ever representing the physical parameters of the signal sources in the world.

What, then, is the best way to describe what we hear? A start would be to abandon our overwhelming sense that the auditory and other sensory systems reveal the physical world, or that they evolved to do so. What we hear in response to sound signal—including musical tones—are percepts determined by the history of biological success rather than the physical characteristics of signal sources or local pressure changes at the ear. The central issue going forward is whether this seemingly odd way of defining perception can be used to better understand musical phenomenology.

Conclusion

Sound signals are easy to study in physical terms, which has encouraged a largely physical paradigm in auditory research. However, the inaccessibility of the physical sources of sound signals to auditory animals implies that understanding what we hear and why requires thinking about audition in terms of empirical success over the course of evolution and lifetime learning rather than as a system that measures the physical properties of sound signals. The argument in the chapters that follow is that understanding music and its appeal may be better informed by this biological framework rather than by a framework based on mathematics, physics, or music theory.

Additional Reading

Forster, C. M. L. (2010). *Musical Mathematics.* San Francisco: Chronicle Books.

An encyclopedic account of the math and physics pertinent to music, showing among other things that the descriptions here are grossly simplified versions of more complicated issues.

Heller, E. J. (2013). *Why You Hear What You Hear: An Experimental Approach to Sound, Music and Psychoacoustics.* Princeton, NJ: Princeton University Press.

An account of audition and music from the perspective of someone who takes audition to be understandable in physical terms.

Plomp, R. (2002). *The Intelligent Ear: On the Nature of Sound Perception.* Mahwah, NJ: Erlbaum.

A thoughtful consideration of the auditory issues debated over much of the 20th century.

Rossing, T. D., R. F. Moore, and P. A. Wheeler (2002). *The Science of Sound,* 3rd ed. San Francisco: Addison-Wesley.

A good textbook on hearing and sound signals.

Schnupp, J., I. Nelken, and A. King (2012). *Auditory Neuroscience: Making Sense of Sound.*

Cambridge MA: MIT Press.

Another fine textbook that provides much more detail about many of the topics discussed here.

Warren, R. M. (1999). *Auditory Perception: A New Analysis and Synthesis.* Cambridge, UK: Cambridge University Press.

A monograph that focuses on sound perception.

2

The Perception of Sound Signals

Our awareness of sensory stimuli—a reasonable definition of *perception*—is characterized by the subjective qualities we use to describe what we see, feel, taste, smell, or hear. The basic qualities for vision are lightness, brightness, color, form, depth, and motion; for the tactile senses, touch, pressure, vibration, and pain; for taste, sweet, salty, bitter, and sour; and for odors, a list that includes descriptors such as floral, pungent, and putrid. In audition, *loudness, pitch,* and *timbre* are the terms that describe the perceptual qualities elicited by sound signals in the environment that biology transforms into sound stimuli. Loudness is the subjective sense of the intensity of sound pressure levels at the ear. The pitch of a sound signal, if it has one, is the sense of higher or lower on a subjective scale that depends on periodic repetition of the pressure changes in the signal. And timbre is the term applied to the complex attributes that distinguish the quality of sound signals that elicit the same sense of loudness and pitch. The purpose of this chapter is to introduce these perceptual qualities, their peculiarities, and the ways in which they pertain to music.

Loudness

Loudness describes the perception of sound signal intensity, which is measured as *sound pressure level* at the ear, a physical parameter whose units are newtons / m^2 (a newton is the amount of force needed to accelerate a 1-kilogram mass 1 meter per second). In practice, however, intensities are usually given in *decibels,* units named after Alexander Graham Bell. The decibel (dB) is also a physical unit of sound pressure, but defined in terms of human hearing. Thus, a sound pressure level of 0 decibels is the average hearing threshold for young adults responding to a sine tone at 1000 Hz (equivalent to about 0.00002 newtons / m^2). Devices that measure sound pressure are therefore calibrated ("weighted") for particular purposes. If the aim is specifically pertinent to human hearing and perception, the so-called "A-weighting" is used, which mimics the frequency-dependent sensitivity of the human auditory system, much as photometers mimic the sensitivity of the human visual system rather than the absolute amount of photon energy reaching the detector. Settings with other weightings are used to assess, for example, the adequacy of concert halls, or the safety of sound pressure levels in a workplace.

Since humans respond to sound signal intensities over an enormous range of local pressure changes, the decibel scale (and our perception of sound intensity) is logarithmic (Figure 2.1). As a result, a small change in decibels represents a large change in physical intensity. For example, an increase of 6 dB represents a doubling of sound pressure level at the ear. The upper limit of perceivable intensities is about 120 dB, an intensity of about 60 newtons / m^2. At this level, the changes in pressure are painful and damage hair cells, although lower intensities can also be destructive if the exposure is chronic. Since hair cells don't regenerate, any loss leads to some degree of permanent deafness.

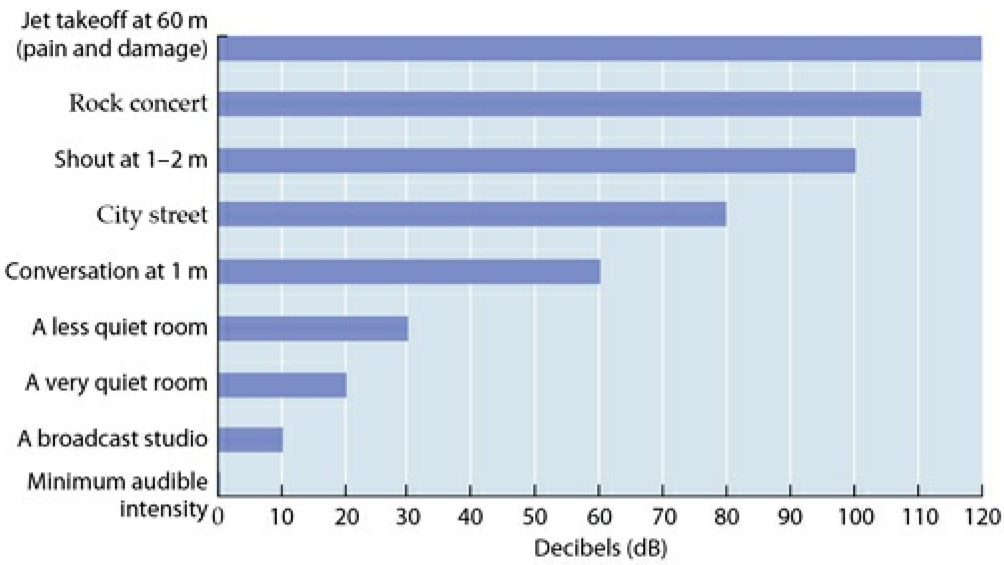
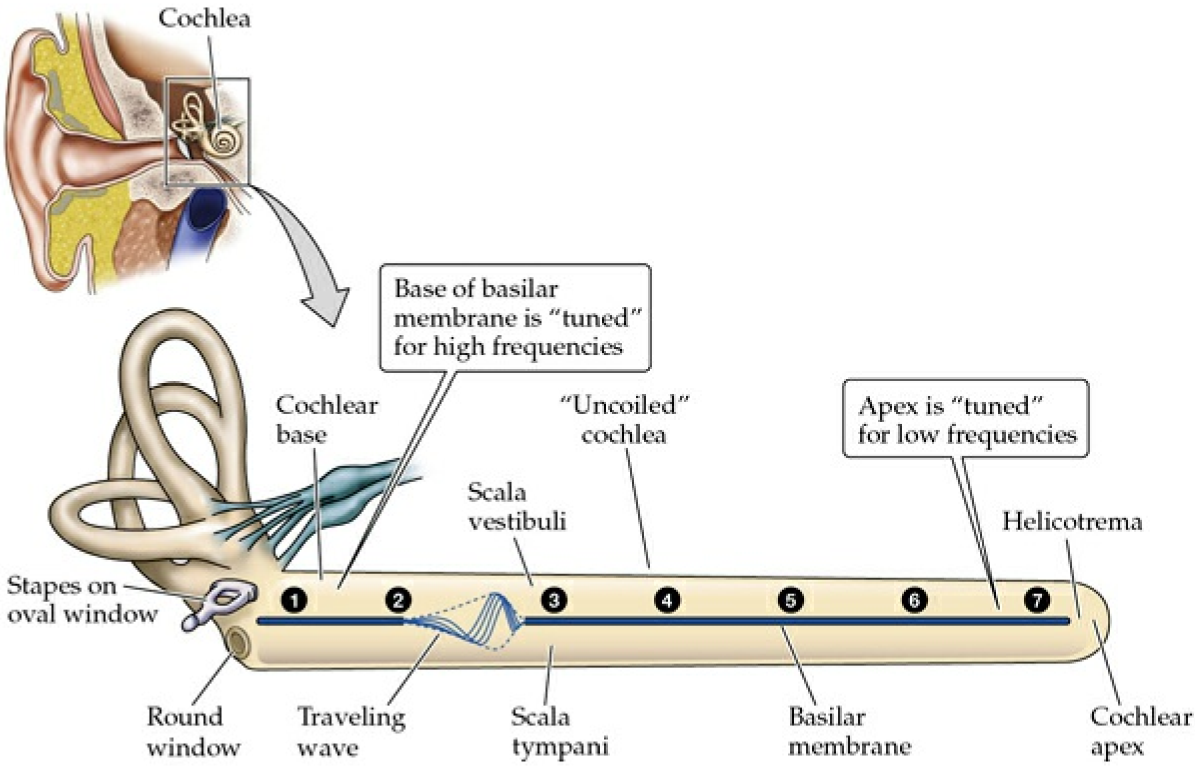


FIGURE 2.1 Examples of sound signal pressure levels expressed in decibels. (From Purves et al., 2013.)

The Peculiar Relationship between Sound Pressure and Loudness

The perception of loudness begins with signal modifications imposed by the preneural apparatus of the ear. As described in Chapter 1, the properties of the external and middle ear have evolved to amplify and filter sound signal energy before it reaches the stage of neural processing, as is typical of the peripheral apparatus in any sensory system. The end result in the cochlea is a *traveling wave* that proceeds along the basilar membrane of the inner ear. The shearing motion of the membrane caused by the traveling wave deflects the hair cell cilia, initiating action potentials in afferent auditory nerve axons and the subsequent neural processing that we end up perceiving as relative loudness (Figure 2.2).



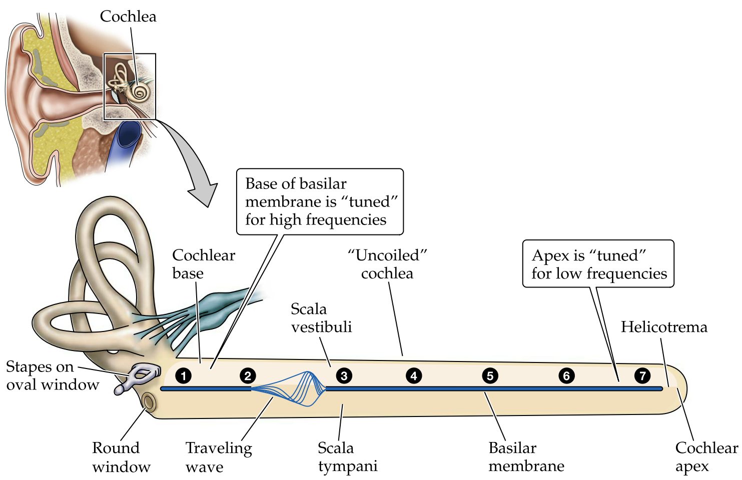


FIGURE 2.2 Response of the basilar membrane to a tonal stimulus. The movement that activates the neural components of the auditory system is a *traveling wave* along the basilar membrane that continues for as long as a tonal sound signal persists (a sine tone in this example). The vibration of the membrane differs from the standing wave generated by the taut string in Figure 1.3 because it is attached only at the oval window end of the cochlea (1), being free to move at the apical end (7). An analogy is the wave that moves along a whip held at one end but free at the other. The greater width and stiffness of the basilar membrane near the oval window allows it to respond to higher frequencies, whereas the more compliant region toward the cochlear apex responds to lower frequencies. The perception of loudness is initiated (but not simply commensurate with) the amplitude of the peak (or peaks) in the traveling wave along the membrane. Activation by a harmonic series or non-tonal sound signals occurs in the same way, although the movements of the membrane are more complex. (From Purves et al., 2012.)

A commonsense expectation is that loudness should vary in direct proportion to physical intensity measured as sound pressure at the ear. This is not the case, however: loudness also varies as a function of frequency, the bandwidth of the signal, signal duration, and other factors. Since less is known about the progressive modifications imposed by the neural stations of the primary auditory pathway, the neural basis for these effects is not clear. An additional obstacle to understanding the perception of sound signal intensity is the fact that the intensities of two sound signals sum in perception only if the signals are similar in most other respects. Otherwise, they are heard as different sources with different loudness values.

The most thoroughly documented example of the complex relationship between the physical measurement of sound pressure and what we perceive is variation in loudness as a function of frequency (Figure 2.3). When listeners indicate the loudness they hear in response to sine tones presented in a laboratory setting, loudness falls off at low and high frequencies within the range of human hearing, being greatest at frequencies between roughly 500 and 5000 Hz. This range of maximum sensitivity coincides with the range of sound signal frequencies that characterize voiced (tonal) speech signals and music, as well as many other sounds in nature. Thus, loudness is not simply a result of the physical intensity of an auditory signal. Like lightness and brightness in vision, loudness depends on the context in which a given intensity occurs. The context in Figure 2.3 is frequency, but other parameters, such as duration and bandwidth, also affect the loudness heard in response to signals that produce the same pressure changes at the ear.

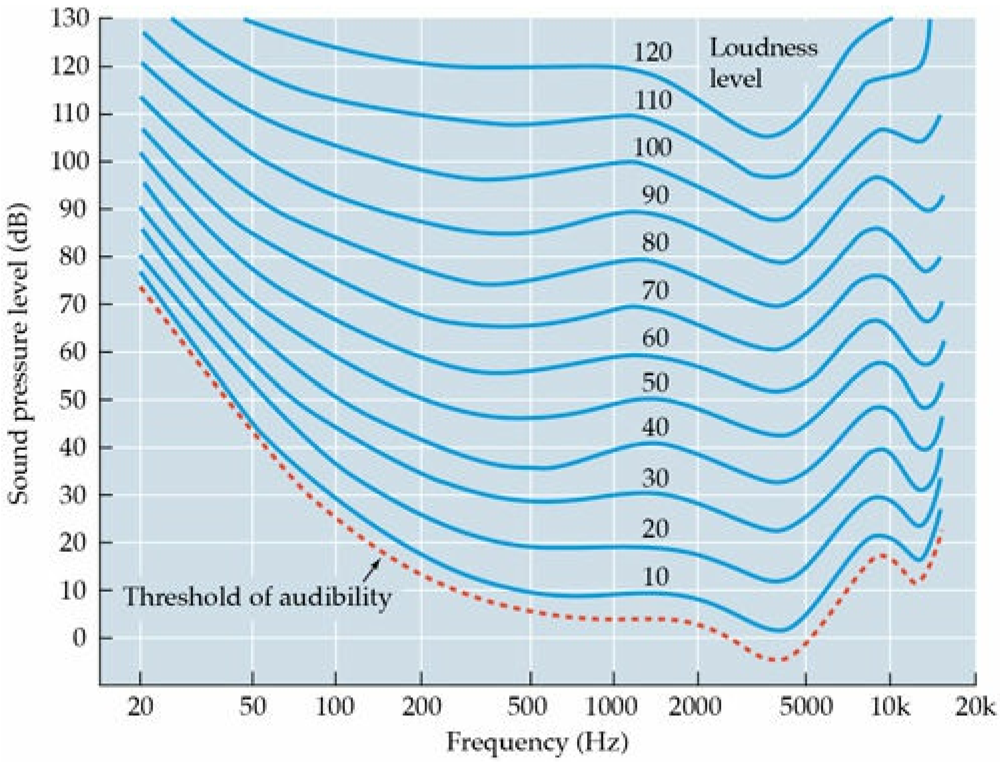


FIGURE 2.3 Variation in loudness as a function of sound signal frequency. Each curve indicates the

sound pressure levels heard as equally loud when listeners are tested with sine tones at different frequencies (each curve is a different level test intensity). The U-shaped curves show that human sensitivity to sound signal pressures is greater at frequencies between ~100 and 6000 Hz than at lower or higher frequencies, a range that includes most of the tonal (repeating) sound signals in speech and music. The dashed red line indicates the absolute threshold of human hearing. (From Purves et al., 2013.)

Pitch and Its Peculiar Relationship to Frequency

Pitch is the perception of a sound signal as higher or lower on a subjective scale that is related to, but not determined by, the repetition rate of the frequencies in a sound signal. Whereas all sound signals have a frequency spectrum, only a minority have a pattern of the recurring peaks and valleys that elicit a sense of pitch. A pitch can arise in response to a sine wave, a tone defined by a harmonic series, or sound signals that repeat in more complex ways within the frequency range of human hearing. As described in Chapter 1, when sound signals lack repetition rates within the span of audible frequencies (~20 to 20,000 Hz) they are loosely classified as "noise."

Of special interest are *tones*, defined as harmonic series whose fundamental frequencies have repetition rates of about 50 to 5000 Hz. The pitches elicited by fundamental frequencies in this range are specific to voiced vowel sounds in speech and the notes used in music, both of which correspond approximately to the tones generated by the eighty-eight keys on a piano (from ~27 to 4186 Hz). Repetition rates outside this tonal range are heard as "bumpy" or "rough" at the low end and "squeaky" or" "shrill" at the high end. The non-tonal pitches that fall beyond this range are, for example, what we hear in response to sirens. Although pitches are heard, they are not ones that are spoken, sung, or played on conventional musical instruments.

Studies of pitch and tonality in modern terms began in the late sixteenth and early seventeenth centuries, mainly from the observations of Vicenzo Galilei and his son Galileo. The Galileis demonstrated that pitch varies as a function of the mass / unit length of a string and its tension, and not simply ratios of string lengths and tensions, as had been thought since the time of Pythagoras in ancient Greece (and probably earlier in China). Galileo thus explained why the same pitch can be generated by a variety of vibrating sources. His insight was that different combinations of the length, density, and tension produce the same pitch if a source produces the same rate of vibration. Galileo therefore proposed that pitch is determined by the frequencies of the sound signals objects produce, and not by ratios as such.

Although knowledge of acoustics and hearing has obviously grown enormously since the seventeenth century, the idea that what we hear is determined by the physical characteristics of the signal at the ear remains a premise in much work on pitch. Over the past century or more, however, evidence has accumulated that pitch is not a simple function of frequency, or of any other physical parameter of the signal at the ear. The findings that have been difficult to explain in terms of sound signal frequencies per se include:

1. The pitch heard in response to a harmonic series corresponds to the lowest harmonic (the fundamental frequency).
2. The fundamental frequency of a harmonic series is heard even when there is no spectral energy in the signal at that frequency, a phenomenon referred to as "hearing the missing fundamental" (Figure 2.4).[1]
3. When the frequencies of a harmonic series are increased or decreased by a constant amount such that they lack a common divisor (which normally defines the fundamental frequency of a harmonic series), the pitch heard corresponds to neither the fundamental frequency nor the frequency spacing between the harmonics (values that are normally the same). This effect is called the "pitch shift of the residue."
4. When the frequencies of only some of the harmonics of a tonal sound signal are changed such that the fundamental of the lower harmonics differs from that of the higher harmonics, the pitch heard typically corresponds to the fundamental of the three or four spectral components closest to ~600 Hz. This phenomenon is called "spectral dominance."
5. Sound signals with waveform repetition rates in the range of ~200–500 Hz evoke a stronger sense of pitch than signals with repetition rates above or below this range. This effect is referred to as "pitch strength."

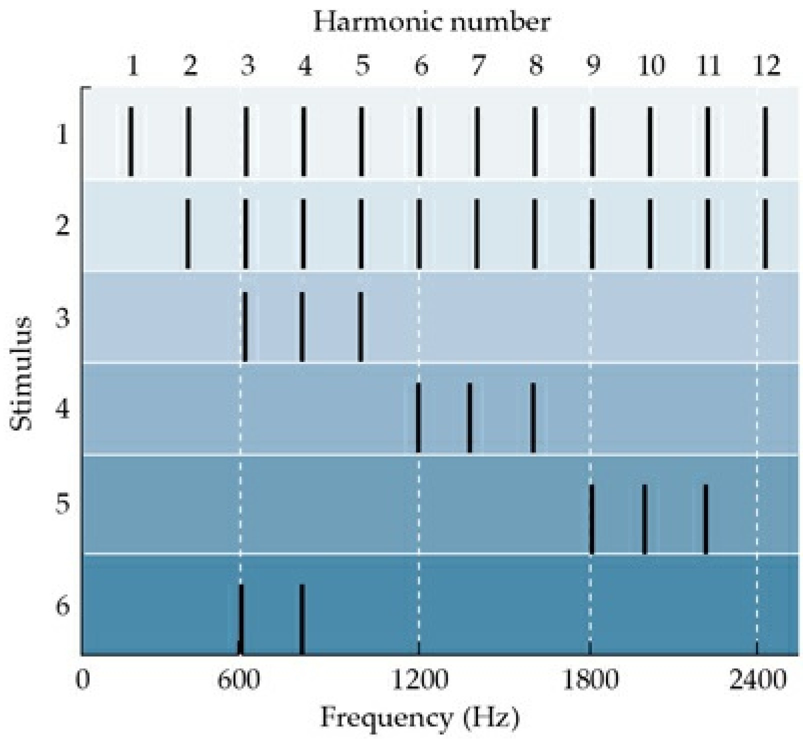


FIGURE 2.4 Hearing the "missing fundamental." Panel 1 diagrams the frequencies of the harmonic series generated by a signal with a fundamental frequency of 200 Hz; the subsequent panels illustrate signals from which different harmonics have been removed. In each case, listeners judge the pitch of the sound signal to be that of the fundamental frequency (i.e., a pitch corresponding to 200 Hz), even though there is no energy at 200 Hz in the spectra shown in the five lower panels. Each signal sounds different, however, because the numbers and frequencies of the upper harmonics produce different timbres. (After Purves et al., 2013.)

Another peculiar aspect of pitch is the distance along the basilar membrane at which two sine tones begin to interfere with each other in perception. This distance is called the *critical bandwidth* and can be measured by psychophysical testing. For a 100-Hz sine tone, when a second "probe" tone comes within about 45 Hz of the 100-Hz tone, listeners notice that the second tone begins to impinge on the perception of the first tone; outside this range, the two tones are simply heard as different. The width of the band in terms of frequency is thus about 90 Hz (since the perception of pitch is nonlinear, this value varies with the frequencies of the tones being tested). When these psychophysical data are translated into physical lengths along the basilar membrane (which extends ~35 mm), the distance within which two sine tones interact is ~1 mm, regardless of the frequency range. This observation is consistent with the fact that each unit length of the basilar membrane reports a smaller range of frequencies as the distance from the oval increases (see Figure 2.2).

What is puzzling, however, is that at any frequency, the physical displacement of the basilar membrane by the traveling wave arising from a sine tone involves regions that are many millimeters in length, as first demonstrated by Georg von Békésy's studies of cochleas taken from human cadavers.[2] To elicit responses in dead cochleas von Békésy had to use very strong sound signals (e.g., sine waves at 100 dB). More recent analyses in living animals show that activation of the outer hair cells mitigates this discrepancy by modulating the compliance of the basilar membrane according to efferent signals arising centrally, thus emphasizing displacement at the locus of the peak illustrated in Figure 2.2. Nonetheless, the difference between responses determined by psychophysical testing and responses measured at the level of basilar membrane movement are difficult to explain. This difference adds to the evidence that what we hear is not just a result of physical sound signals modulated by the peripheral auditory apparatus.

In sum, pitch, like loudness, is strangely related to physical sound signals at the ear and their transduction by the inner ear, and again the question is why.

Timbre

Timbre is defined, more or less by default, as the set of complex perceptual qualities that differentiate sound signals when their intensity and repetition rate are the same. Thus, a clarinet and a bassoon playing the same note and creating pressure changes at the ear that have the same amplitude obviously sound different.

Timbre has no single unit of measurement and is generally accepted as being “multidimensional.” The physical bases of this perceptual quality include the number of harmonics present (which explains why the sound signals diagrammed in Figure 2.4 sound different even though they are heard as having the same pitch); the overall structure of the harmonic series (e.g., wind instruments open at one end generate only the odd harmonics in a series in contrast to vibrating strings; see Figure 1.3); the amplitude profile of the sound signal over time (e.g., its rate of attack and decay); and finally the amount and quality of noise in the signal (i.e., whatever nonperiodic information it includes).

Since these and other factors underlying timbre are difficult to assess, timbre remains the least studied aspect of auditory perception. Nonetheless, the characteristics included under this rubric are important, particularly in vocal and musical sound signals (see Chapter 4).

Evidence from Linguistics

Studies of speech and language have also shown that what we hear differs from the physical characteristics of sound signals in ways that are hard to explain. For example, perception of the same speech sound varies according to the characteristics of the speaker's voice. In studies carried out by linguist Peter Ladefoged and psychologist Donald Broadbent in the 1950s, subjects listened to six versions of the sentence "Please say what this word is," followed by a test sound signal of a *b*-vowel-*t* word that contained one of four different vowels—b*i*t, b*e*t, b*a*t or b*u*t.[3] The six versions of the sentence were presented with the vocal characteristics artificially altered to assess the influence of a speaker's voice on perception. Most subjects identified the test words according to the vocal qualities of the speakers rather than to the acoustical character of the test word signals themselves. Similar studies showed that following an / *al* / sound, listeners tend to hear ambiguous targets as "ga" rather than "da"; conversely, when the targets followed an / *ar* / sound, listeners tended to hear "da" rather than "ga."

Another example is the McGurk effect. When the acoustical qualities of a syllable are coupled with a visual image of the speaker uttering a different syllable, the sound heard is a compromise between the auditory and visual information provided (Figure 2.5).[4] This impressive phenomenon implies that what is seen when hearing speech, a musical performance or any other sound signal can strongly influence what is heard.

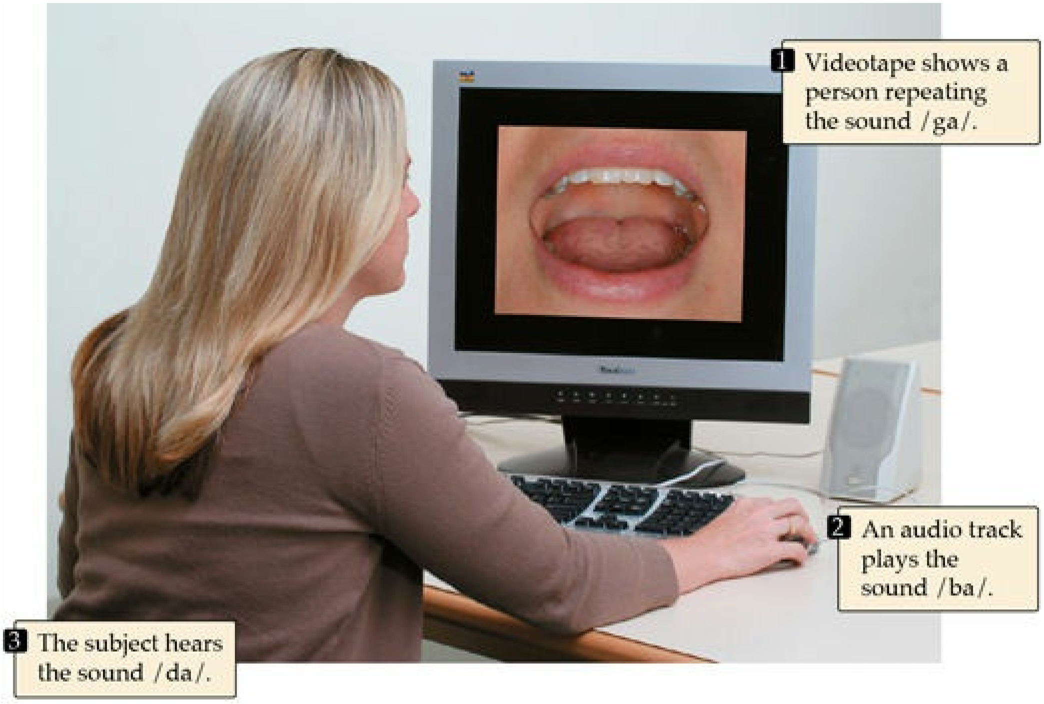


FIGURE 2.5 The McGurk effect. (From Purves et al., 2008.)

Auditory Objects and Scenes

Finally, in keeping with the overall theme of the chapter, what a listener hears is not loudness, pitch, and timbre as such, but something akin to the perception of visual objects—that is, perceptions of *sound* objects that are behaviorally relevant: a voice, footsteps, a banging door, the patter of rain. In vision, objects are organized in scenes, and the analogy in audition is auditory scenes or “soundscapes.”

Researchers interested in this subfield emphasize the importance of grouping and streaming in the perception of auditory scenes. For example, we routinely tune out background noise to follow the speech of a person we are talking with. Listeners thus have the ability to track a pertinent stream amid a welter of simultaneous auditory signals, recognizing the stream as an auditory object of interest, including the notes in a melody and harmonic sequences in music. Such observations further emphasize the idea that audition is determined by the biological significance of sound signals rather than by their physical properties.

Conclusion

The perceptions elicited by sound signals don’t accord with the physical characteristics of sound signals at the ear, although the signals are obviously the real world causes that initiate what we end up hearing. In a sensory system like vision, the inability of the system to measure the physical parameters of light signal sources leads to pervasive discrepancies between visual perceptions and the physical world. Although more subtle, such discrepancies are equally apparent in audition. In both cases, the implication is that what we hear or see is determined by the biological demands of successful behavior rather than by a strategy that reports the physical nature of sensory signals. As taken up in the following chapters, this biological strategy appears to be central to understanding tonal music, and why we like music in the first place.

Additional Reading

Helmholtz, H. L. F. (1877 / 1954). *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (A. J. Ellis, trans.). New York: Dover.

A nineteenth-century classic that remains relevant today.

Ladefoged, P. (1962). *Elements of Acoustic Phonetics.* Chicago: University of Chicago Press.

A linguist’s take on audition.

Pierce, J. R. (2001). Introduction to pitch perception. In: *Music, Cognition and Computerized Sound* (P. Cook, ed.). Cambridge, MA: MIT Press.

A highly readable summary of pitch phenomenology.

Plomp, R., and W. J. Levelt (1965). Tonal consonance and critical bandwidth. J Acoust Soc Am 28: 548–560.

A classic paper on critical bands.

von Békésy, G. (1962). Three experiments concerned with pitch perception. J Acoust Soc Am

35: 602–666.

An article that explains his model of cochlear function based on studies of inner ears taken from cadavers, which led to von Békésy's Nobel Prize in 1961.

1. This effect is experienced routinely when using a telephone. To allow greater carrying capacity, phone systems don't generate frequencies below 400 Hz or above 3400 Hz, even though the fundamental frequency of an adult male voice is about 120 Hz on average, a female voice about 230 Hz, and the fundamental frequency of middle C on a piano about 262 Hz. This is why voices sound different on the phone, and why earbuds have a larger bandwidth that faithfully conveys musical tones.

2. G. von Békésy, 1961 Nobel Lecture (Nobelprize.org). A highly readable account of this work on the inner ear.

3. P. Ladefoged and D. E. Broadbent (1957).

4. H. McGurk and J. MacDonald (1976). The effect can be experienced in several fine examples posted on YouTube (e.g., "Try The McGurk Effect! - Horizon: Is Seeing Believing? - BBC Two").

3

Human Vocalization

ANIMAL VOCALIZATIONS HAVE A WIDE RANGE of intensities and frequencies that are rich in both periodic (tonal) and aperiodic (noisy) signals, and these signals obviously matter a great deal to us and many other species. Quite apart from any semantic content, the information embedded in human vocalizations includes the probable size, gender, age, and emotional state (angry, loving, bored, etc.) of the speaker. Throw in information derived from timbre, and even individual identity is apparent. Although vocal sounds comprise only a fraction of auditory experience, analyses of speech show it to be structurally similar to a conglomeration of other environmental and animal sounds. All told then, human vocalization provides a good place to begin exploring auditory perception in empirical terms, and, more specifically, how the resulting biological framework informs music by suggesting why humans worldwide prefer particular tone combinations in melodies and harmonies.

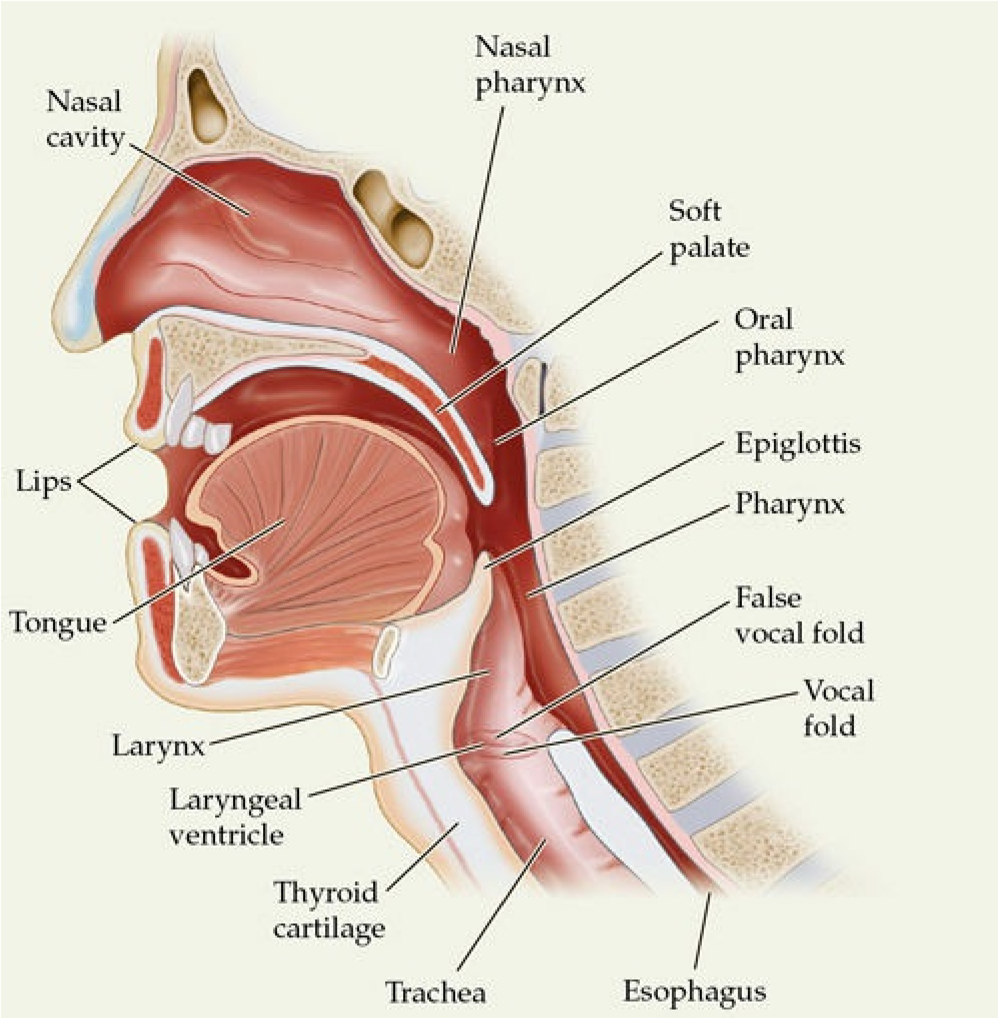
The Production of Vocal Sounds

As discussed in Chapter 1, the greatest excursion of a vibrating string occurs over its full length, the result being a large fluctuation in the amplitude of the sound signal at its fundamental frequency (the first harmonic). The next most powerful mode of vibration is at half the length, called the second harmonic, the next most powerful one third the length, called the third harmonic, and so on (see Figure 1.3). In most instances, however, vibrating strings are attached to other objects, such as the body of a piano, violin, or guitar that amplify and modulate the effects of the vibrating string, enhancing some modes and damping others, as well as adding timbre. Indeed, without these attachments the disturbance of the local atmosphere caused by a vibrating string produces only a faint and not very interesting sound.

The reason for belaboring these points is that the human *vocal folds* (also called the *vocal cords*) are roughly analogous to a vibrating string, while the rest of the *vocal tract* is analogous to the attached body of a musical instrument (Figure 3.1). Speech sounds are thus produced in much the same general way as musical sounds, a fact whose pertinence will become apparent in the next chapter.

Not all musical sounds, however, are quite so simple. While broadly tonal, the sound signals generated by bells, gongs, drums, blades and other metallophone or lamellophone instruments used in traditional Javanese, African and Caribbean music are not exactly the same as the sound signals produced by vibrating strings or the vocal folds. Although the same rules of physics and analysis apply, the spectra are more complex than the simple harmonic series produced by a vibrating string or the vocal folds, with peaks of energy that are less clearly limited to integer multiples of the fundamental frequency.

By definition, the vocal tract includes the entire apparatus for vocalizing shown in Figure 3.1. The air expelled from the lungs streams through the opening between the vocal folds (the glottis) in the larynx. The airstream accelerates through this narrow space (think of the accelerated flow of a river passing through a narrow gorge). The resulting decrease in lateral pressure causes the cords to come together until the pressure buildup in the lungs forces them open again. This cyclical process gives rise to a vibration whose frequency is determined primarily by the muscles that control the tension on the vocal folds and the pressure in the thoracic cavity. Varying the tension on the folds is used change the pitch when speaking or singing. The fundamental frequency of vocalization ranges from about 50 to 400 Hz, depending on the gender, size, age, and emotional state of the speaker; the normal range of speech is about an octave and of singing about two octaves (although training can increase this latter value quite a lot).



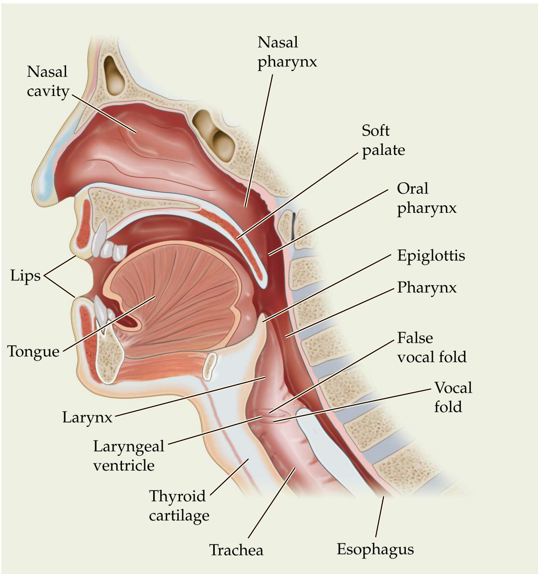
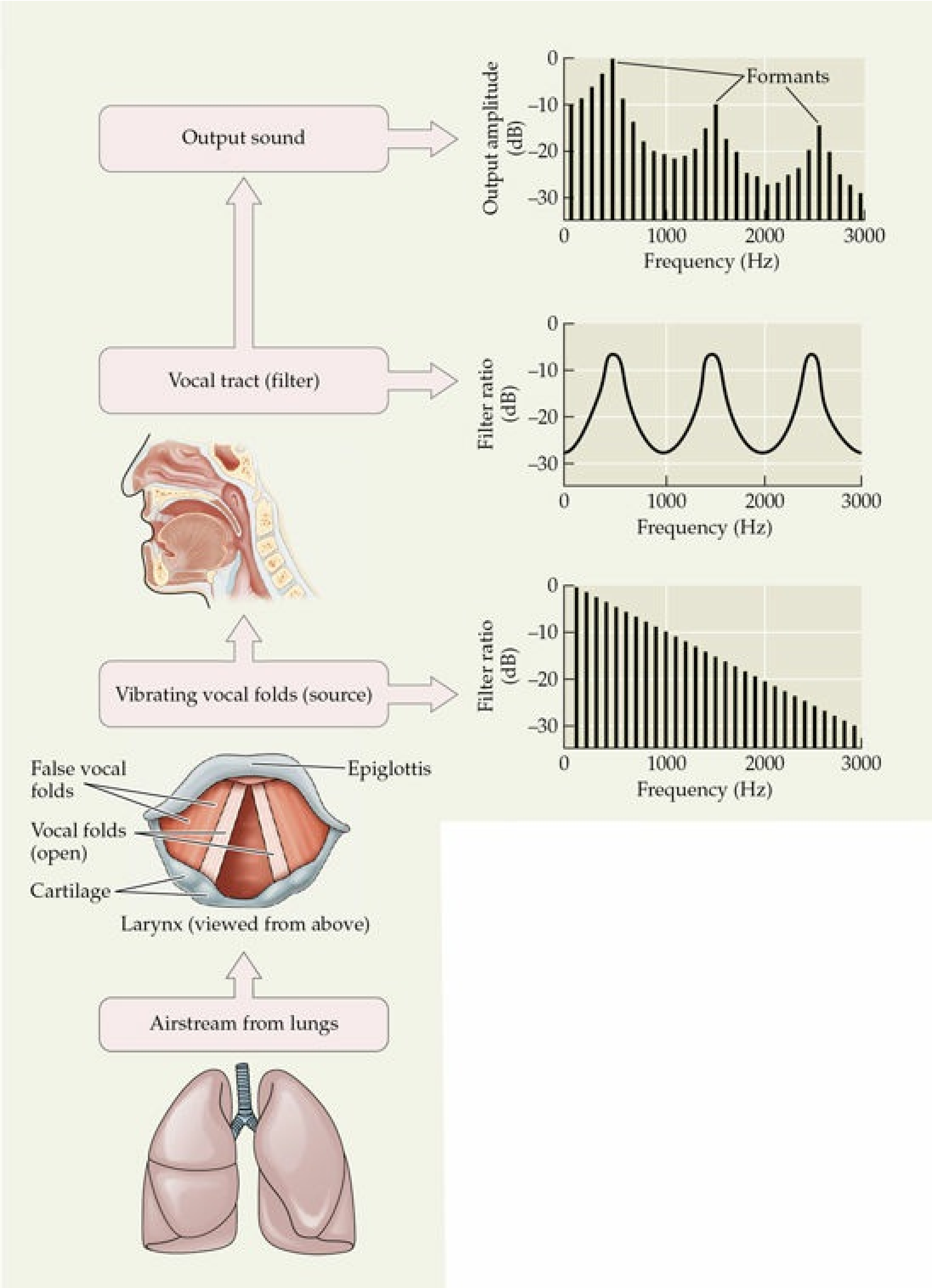


FIGURE 3.1 The human vocal tract includes the vocal apparatus from the larynx to the lips. The structures that form the tract are the larynx, pharynx, soft palate, mouth, and nasal cavities. The musculature of the pharynx, soft palate, tongue, and lips modulates the shape of these cavities, which in turn alters the distribution of energy in the harmonic series generated by the vibrations of the vocal folds in specific ways to create different vowel and other speech sound signals. (From Purves et al., 2012.)

Like the body of a musical instrument, the vocal tract above the vocal folds shapes and filters the amplitudes in the harmonic series produced by vibration of the folds (Figure 3.2). This "source-filter model" of speech was proposed in the nineteenth century by the German anatomist and physiologist Johannes Müller and has been accepted ever since. The lungs serve as a reservoir of air, while the muscles of the diaphragm and chest wall provide the motive force. The vocal folds then generate the vibration that characterizes *voiced* (periodic) sound signals, defined as those made when vowels (and voiced consonants such as "m" or "z") are uttered. *Unvoiced* sound signals are produced when vocal fold vibration is not involved, as in the utterance of most consonants (like "f" or "s"), or vowels when they are whispered. Either way, the shapes of pharyngeal, oral, and nasal cavities continuously modulate the signal. The shapes of these cavities are actively changed by the musculature of the larynx, pharynx, soft palate, tongue, and lips, producing different natural resonances and subsidiary frequency peaks in the harmonic series that originated in the larynx. The relative amplitudes and frequencies of the resulting peaks of intensity in the original harmonic series—called speech *formants*—are the basis of the different vowel sound signals that are produced in any language.


Output sound
Vocal tract (filter)
Vibrating vocal folds (source)
False vocal folds
Epiglottis
Vocal folds (open)
Cartilage
Larynx (viewed from above)
Airstream from lungs
Formants
Output amplitude (dB)
Filter ratio (dB)
Filter ratio (dB)
0
–10
–20
–30
0
1000
2000
3000
Frequency (Hz)

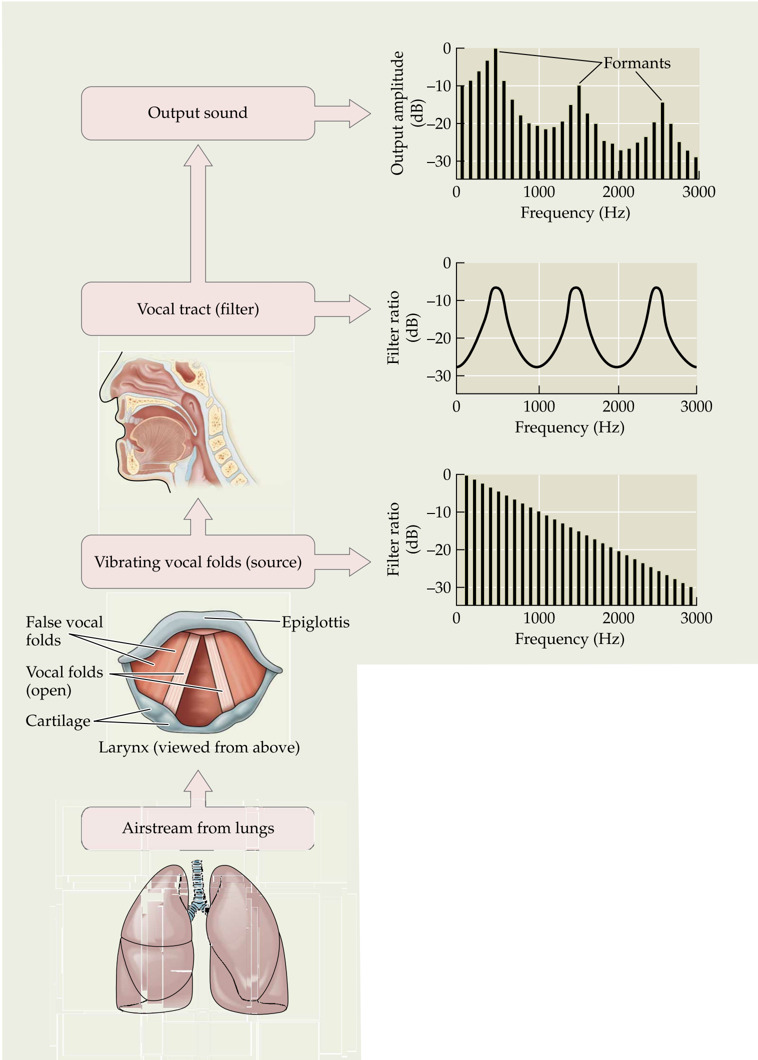


FIGURE 3.2 The source-filter model of vocal sound production. Based on their tension and the force of the air expelled by the lungs, the vocal folds in the larynx initiate vibrations that eventually lead to the tones in "voiced" speech sound signals. The modifiable shape of the rest of the vocal tract—the laryngeal, pharyngeal, oral, and nasal cavities—filters the vocal fold harmonics by the superposition of their own resonances to create the speech sound signals received by listeners. (From Purves et al., 2012.)

The Perception of Vocal Sounds

The different speech sound signals produced in a language are called *phones,* and the auditory perceptions they elicit are called *phonemes*. One or more phones make up syllables in speech, which in turn make up words, which ultimately make sentences. Considering languages worldwide (about 6000, a number that is falling fast as a result of globalization and the intrusion of mining, logging and other industries into previously remote areas), linguists estimate the number of phones to be about 200, of which approximately 30 to 100 are used in any given language. The culturally determined use of different subsets and articulations used in various languages explains why people have trouble learning a new language. In addition to vocabulary, grammar, and syntax they have to produce unfamiliar phones and comprehend new phonemes. Since the brain has a limited ability to change its connectivity after childhood, adult second-language learners retain the accents and grammatical biases characteristic of their native language.

Phones, broadly considered, can be divided into *vowel* and *consonant* speech sound signals. The approximately 40–50 phones that characterize American English (different linguists make different estimates) are about equally divided between vowels and consonants, although other languages vary in this respect, and an emphasis unusual phones such as "clicks" is apparent in some African languages (made by slapping the tongue against the lower teeth and floor of the mouth). Vowel sounds comprise most of the *voiced* components of speech—the elemental speech sounds that are generated by vibration of the vocal folds in any language (see above). Because these oscillations are periodic, vowel sounds have tonality, eliciting a perception of pitch in both speech and song. The variation in the pitch of speech over time is called *speech prosody,* another key term that will figure later in sorting out the relationship between music and speech.

Consonants are phones (or phonemes) that typically begin and / or end syllables, in contrast to vowels, which form the "nucleus" of each syllable and generate the tonality of speech. Consonant sound signals are usually briefer than vowel sounds, involve more rapid changes in sound energy over time, are acoustically more complex, and generally have less power. They are categorized according to the site (or sites) in the vocal tract that determines them (the place of articulation) or the way they are generated (the manner of articulation). Somewhat surprisingly, consonants are the more important carriers of information in speech. Thus, when subjects are asked to repeat spoken sentences from which either the vowel or the consonant sounds have been artificially removed, their understanding of what was said is more complete without vowels than without consonants. This observation jibes with the fact that elderly listeners have particular difficulty understanding speech because of high-frequency hearing loss, which is where many unvoiced consonants have the most power.

Although it seems that speech—whether spoken or heard—is physically divided into words, syllables, and / or phones, this impression is wrong. Recordings of speech show that there is no physical break between strings of syllables or words (Figure 3.3). Moreover, in laboratory studies, listeners generally don't notice short bursts of noise that interrupt speech sounds ten to fifteen times per second. The relevant neural processing and ultimate understanding of speech evidently proceeds in a more holistic way. This difference between signal and percept is further evidence that what we hear is actively created by the auditory system, and not just a representation of physical pressure changes at the ear. The elements of language learned in school are not natural units of speech production or auditory processing.

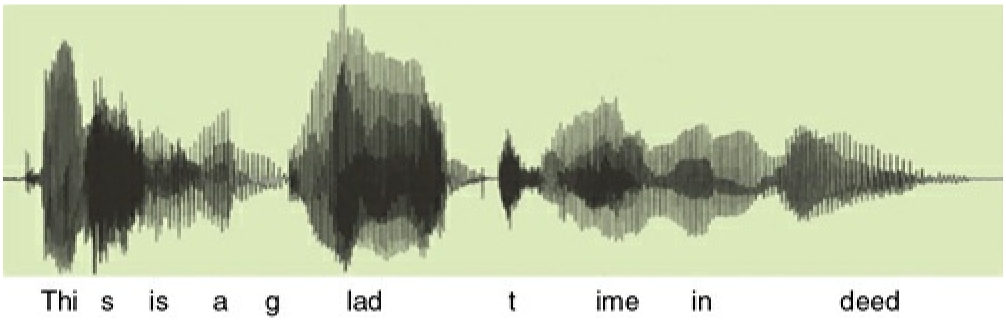


FIGURE 3.3 The discrepancy between speech sound signals and heard speech. The sound signal of the sentence over time (about 2 seconds) does not reflect the breaks between words or the syllables perceived. (From Purves et al., 2008.)

The Perception of Vocal Loudness in Empirical Terms

Although not the first to consider the puzzling relation between the objective intensity of sound signals and subjective loudness (see Chapter 2), psychologist Stanley Stevens working at Harvard in the 1940s through the early 1970s showed that the relationship of physical intensity and loudness was a nonlinear power function, an idea that he extended to other perceptual qualities as well.[1] In the case of loudness, Stevens' "power law" refers to the fact that at low sound pressure levels, loudness is greater than predicted by a linear relationship, whereas at higher pressure levels it is less (see Figure 3.4A, for example). Thus as pointed out in Chapter 2, the link between the physical measure of sound signals and their perceptual effects is not the one-to-one relation between objective and subjective domains that one would expect if sound signals were represented objectively by the auditory system. Stevens demonstrated this relationship by having volunteers make subjective judgments of the relative loudness ("magnitude estimations") of sound signals in the laboratory.

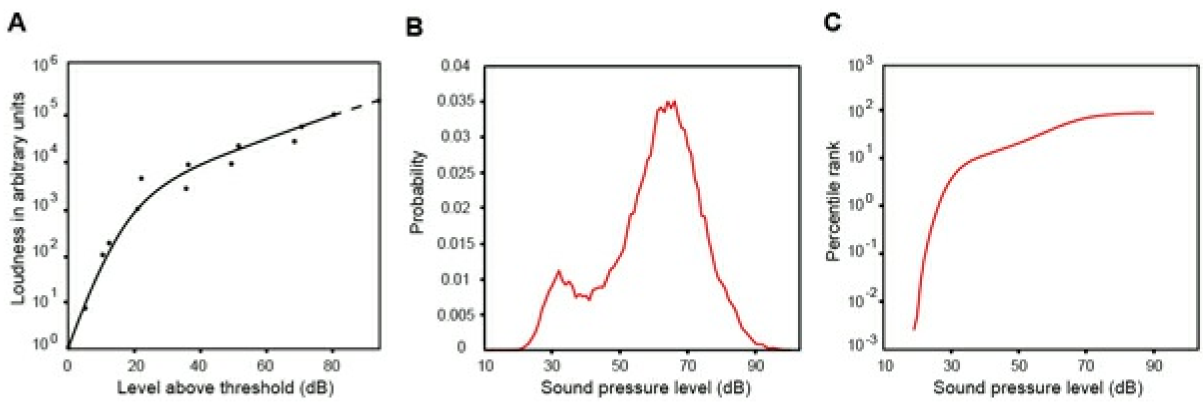


FIGURE 3.4 The relationship between listeners subjective sense of loudness in response to objective intensities in speech sound signals. (A) Loudness as a function of the intensity of the speech sound

signals we routinely hear (redrawn from data from Fletcher and Galt, 1950). (B) The probability distribution of sound pressure levels determined by analysis of a database of American English speech. (C) The cumulative probability distribution of speech sound intensities to which we are routinely exposed (B). The generally similar functions in (A) and (C) imply that loudness accords with the frequency of occurrence of intensities in sound stimuli rather than the intensities of the signal per se. (From Monson, Han, and Purves, 2013. CC BY 3.0.)

Figure 3.4 shows how these observations can be explained by accumulated experience with the intensities of speech and other tonal sounds. Although Stevens and others provided no reason for their results, Richard Warren at the University of Wisconsin suggested in the 1980s that the nonlinear relationship between objective signals and subjective responses might have to do with listeners' experience with the attenuation of sound as a function of distance, and more generally with the "conditions and events responsible for stimulation."[2] This idea also accords with the observation made by Peter Ladefoged and colleagues that when subjects are asked to judge the loudness of speech sounds, their responses correlate better with the measurements of the speaker's vocal effort than with measurements of sound pressure level at the listener's ear.[3] All told, a fair amount of evidence points in the same empirical direction: the loudness we hear is strongly influenced by past experience.

The Perception of Pitch in Empirical Terms

Since the most prevalent sources of tonal sound signals in the auditory environment in which we evolved would have been conspecific vocal sounds, it makes sense that the perception of tonality arose in large part to extract information from human or other animal vocalizations. Like loudness, then, pitch may also be determined empirically by associations made between perceptions and behavioral success. In other words, reproductive success based on ecological meaning of the signal in past experience rather than signal frequencies as such. This strategy would again provide a way to get around the inevitable entanglement of information in auditory stimuli (see Chapter 1).

To explore this possibility, a database of vocal sound signals is needed. A popular option is the Texas Instruments / Massachusetts Institute of Technology Acoustic-Phonetic Continuous Speech Corpus (usually referred to as the TIMIT database), which comprises 6300 utterances of ten brief sentences by several hundred male and female American English speakers (Figure 3.5). Perhaps the most obvious phenomenon to start with in pitch is the fact that we hear the pitch of fundamental frequency of a harmonic series whether there is energy at that frequency or not (see Chapter 2). By the mid-nineteenth century, the German auditory scientist August Seebeck had shown that the frequency of the pitch heard in response to a set of two or more successive harmonics corresponds to the greatest common divisor of the harmonic set, even without corresponding spectral energy at that frequency (the “missing fundamental”; see Figure 2.4).[4] Hearing the fundamental frequency of a harmonic set without power at fundamental frequency is also referred to as hearing a “virtual” pitch.

Figures 3.6A and B show an artificial sound signal that includes only of the third, fourth, and fifth harmonics of a fundamental whose frequency is 150 Hz. Figure 3.6C compares the relative similarity of this stimulus to thousands of speech sound segments sampled from the speech database, plotting the strongest periodicity in each speech segment against the segment’s cross-correlation with the artificial stimulus. Figure 3.6D shows the average correlations as a function of periodicity. The frequency associated with the maximum of this function is 150 Hz, which matches the absent fundamental of the frequency components making up the stimulus in Figure 3.6A. It is also the pitch listeners hear in response to this sound signal. The presumptive reason is that the upper harmonics suffice to trigger a neural association that has been biologically advantageous over eons of evolutionary experience.

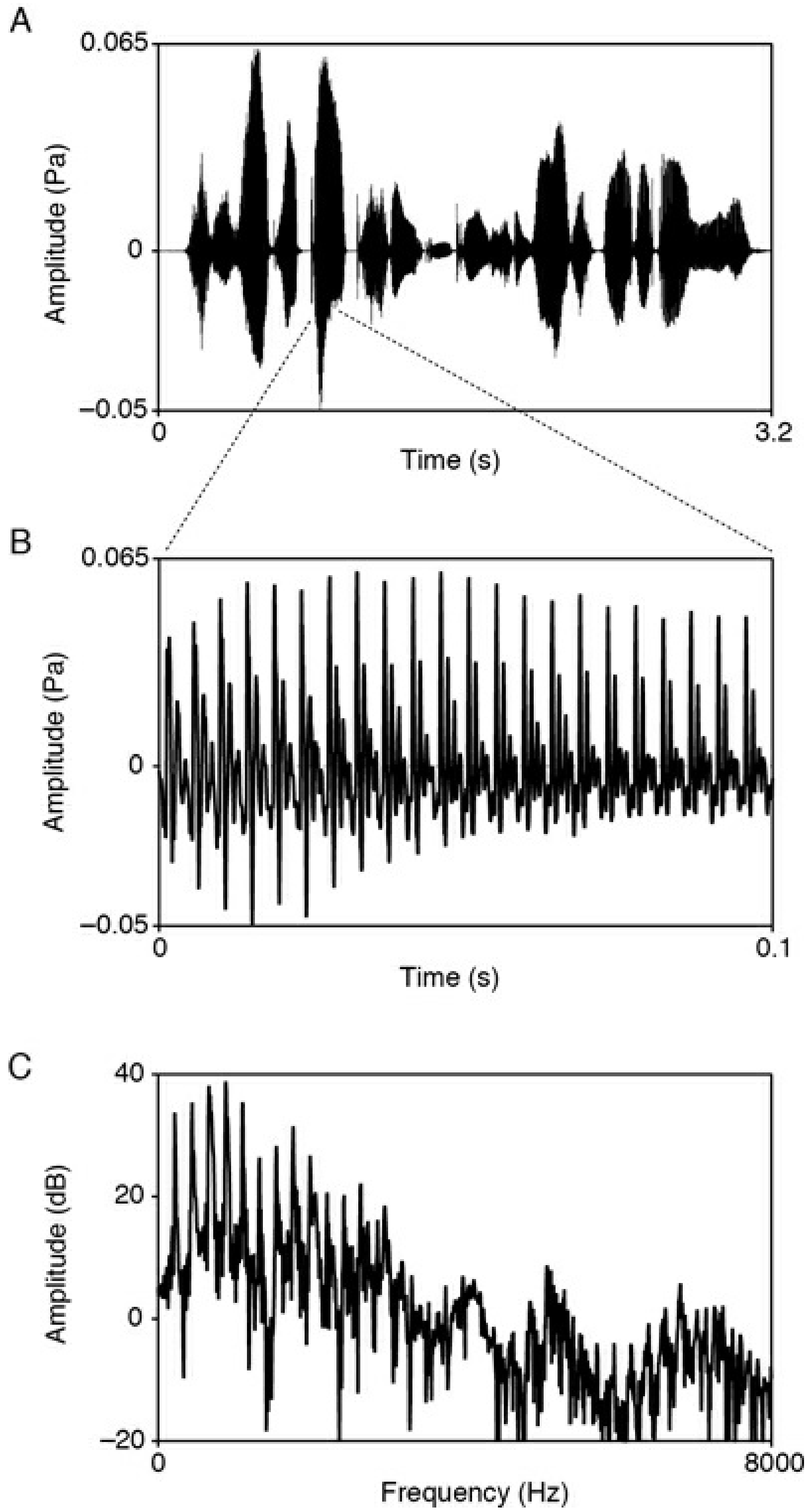


FIGURE 3.5 Using voiced speech signals to explore pitch perception in empirical terms. (A) Variation of the sound pressure level over time in a representative utterance (the sentence in this example is "She had your dark suit in greasy wash water all year"). (B) Blowup of a 0.1-second segment extracted from the utterance (the vowel sound in "dark"). (C) The spectrum of the extracted segment in (B). Incidentally, note the difference between actual voiced speech spectra like this and the ideal version in Figure 3.2. (From Schwartz, Howe, and Purves, 2003.)

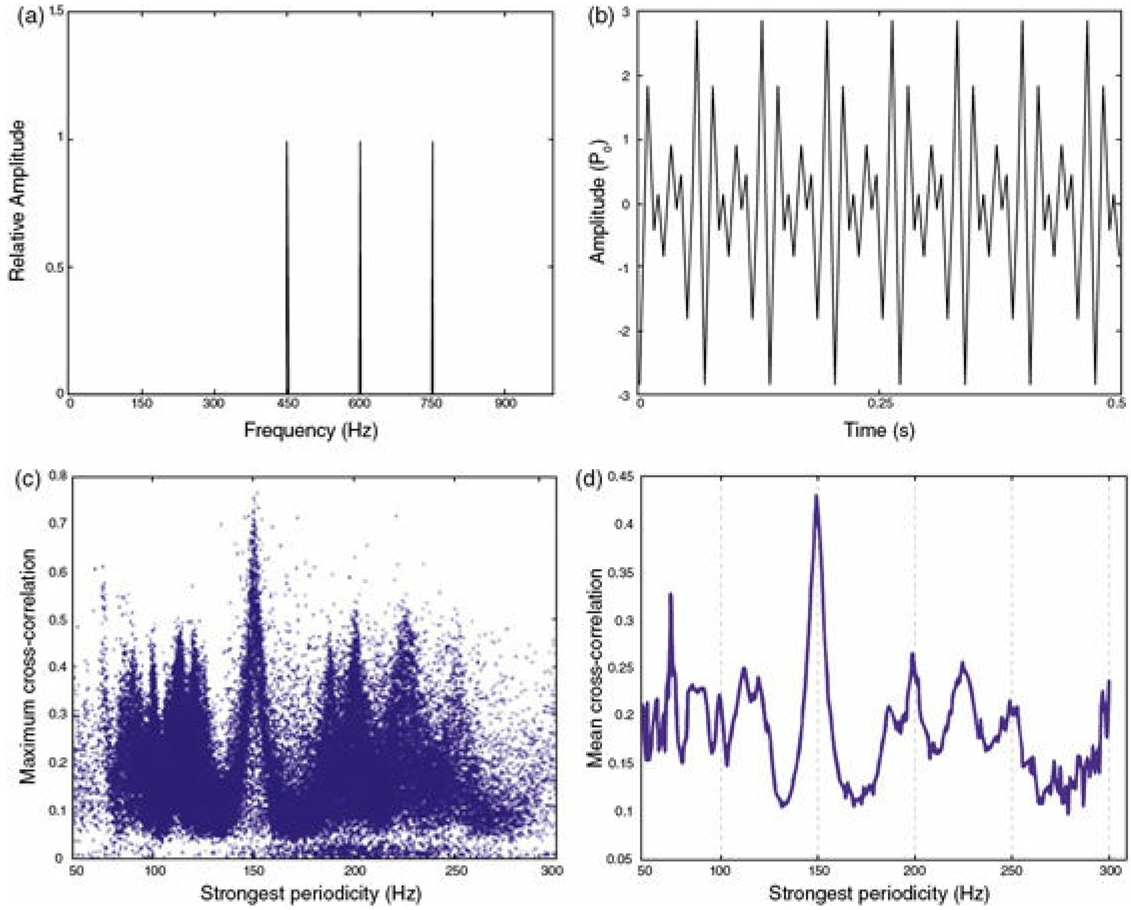


FIGURE 3.6 Explaining the perception of the "missing fundamental" based on experience with vocal sounds. (A) Schematic spectrum of a sound signal comprising the third, fourth, and fifth harmonics of a series whose fundamental is 150 Hz. (B) Representation of the same signal in time. (C) The maximum cross correlation of each speech sound in a large number of speech segments plotted against the strongest periodicity of each segment. (D) Average maximum correlation derived from the data in (C). The periodicity associated with the maximum of the function is 150 Hz. The other maxima at 75 and 225 Hz are consistent with the observation that some listeners identify two or more distinct pitches in response to complex tonal stimuli. (From Schwartz and Purves, 2004. © 2004 Elsevier B. V. Reprinted by permission of Elsevier.)

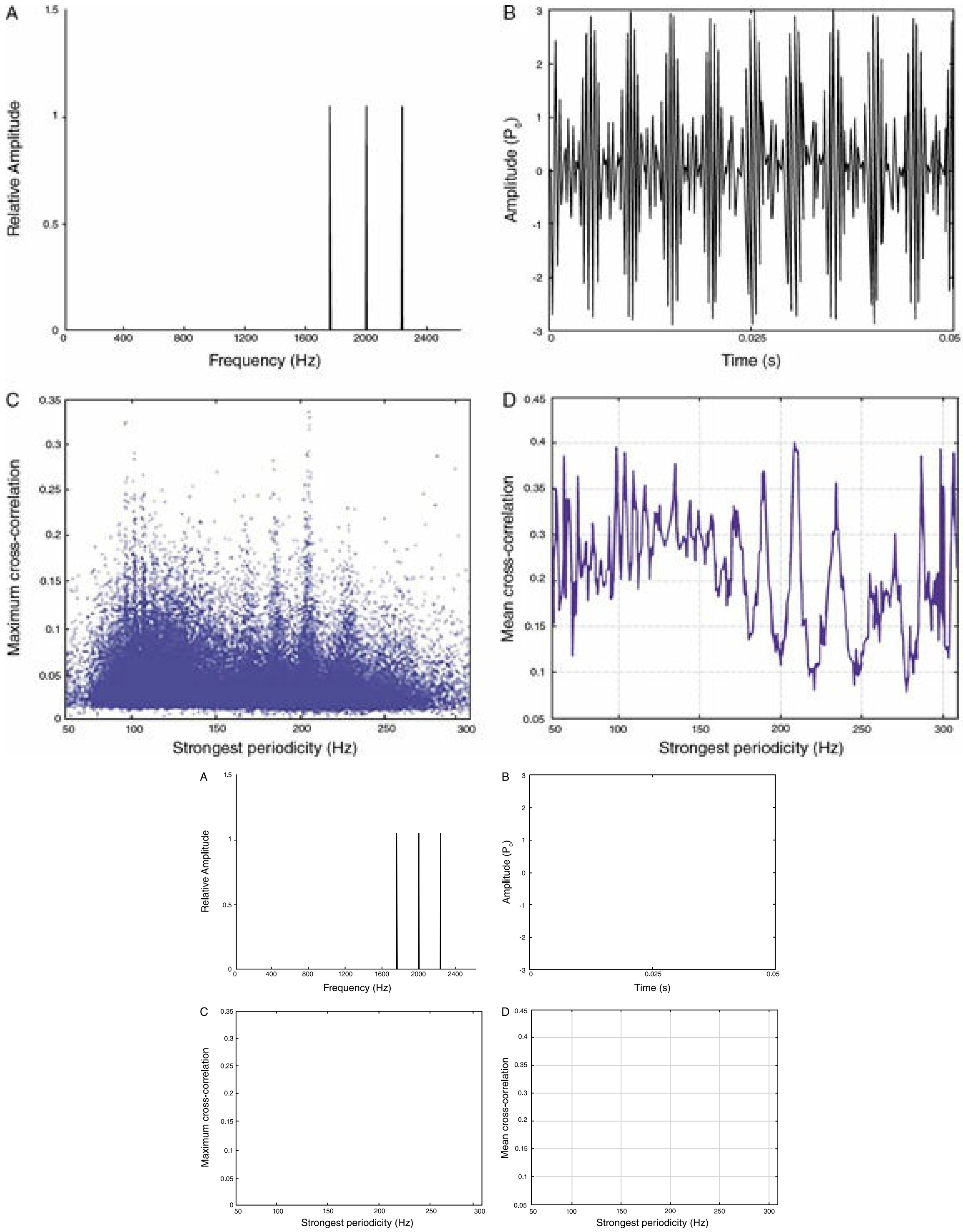


FIGURE 3.7 Explaining the "pitch-shift of the residue" based on experience with vocal sounds. (A) Schematic spectrum of a sound signal comprising the 9th, 10th, and 11th harmonics of a 200-Hz fundamental in which the frequency of each harmonic has been increased by 40 Hz. (B) Representation of the same signal in time. (C) The maximum cross correlation of thousands of speech sounds with the signal in (B) plotted against the strongest periodicity in each speech segment. (D) Average correlation derived from the data in (C). Although the function has many local peaks, the periodicity associated with the maximum of the function is 204 Hz, the pitch heard in psychophysical studies using this stimulus. (From Schwartz and Purves, 2004. © 2004 Elsevier B. V. Reprinted by permission of Elsevier.)

Another phenomenon that supports the idea that pitch is determined empirically comes from the perceptual result of surreptitiously changing frequencies of a harmonic series (hearing the "pitch-shift of the residue," as mentioned in Chapter 2). When the harmonics of a sound signal are altered by adding or subtracting a constant value such that they have no common divisor, the pitch that listeners hear typically shifts in the direction of the frequency change, no longer corresponding to the fundamental frequency of the set. For example, Figures 3.7A and B show frequency and time-domain representations of an artificial stimulus comprising the ninth, tenth, and eleventh harmonics of a series with a fundamental of 200 Hz. The value of each harmonic, however, has been increased by 40 Hz. The harmonics are thus 1840, 2040, and 2240 Hz, instead of 1800, 2000, and 2200 Hz. As a result, the fundamental frequency of the set is 40 Hz, even though the spacing between the harmonics remains 200 Hz. In psychophysical testing, the pitch subjects hear in response to this signal is ~204 Hz. Figures 3.7C and D show the mean of the maximum cross-correlation coefficients for speech, as in Figure 3.6. The periodicity associated with the maximum of the function is 204 Hz, implying again that perceived pitch arises from routine experience with vocal harmonics.

Other phenomena such as "spectral dominance" (hearing ambiguous spectra as biased toward speech spectra), perceived "pitch strength" (hearing pitch more strongly in the midrange of speech fundamentals) mentioned in Chapter 2, although not shown, also accord with predictions made on the basis of species and lifetime experience.

The upshot of these examples is that rather than hearing the frequencies of acoustic signals per se, listeners hear pitches that correspond to the behavioral significance of sound signal frequencies experienced over individual and evolutionary time. The rationale is that perceiving pitch in this way, much as perceiving loudness, is an effective way to generate successful behavior in a world in which unambiguous physical information about sound energy *sources* is not available to biological listeners.

Conclusion

Many aspects of loudness and pitch are difficult to explain in terms of the physical characteristics of sound signals, or the peripheral "transfer functions" generated by the external, middle, and inner ear. It may be that these phenomena are better understood as evidence that what we hear depends on experience with biological success of perceived sensory qualities rather than the physical characteristics of acoustic or other sensory signals. The reason for this strategy is a way of dealing with the fact that hearing and other sensory systems cannot specify physical sources of stimuli. As taken up in the next chapter, this biological interpretation of what we hear and why may also provide a way of unraveling the phenomenology of tonal music.

Additional Reading

Fletcher, H. (1924). The physical criterion for determining the pitch of a tone. Phys Rev 23: 427–437.

A classical study of pitch psychophysics.

Ladefoged, P. (2000). *Vowels and Consonants: An Introduction to the Sounds of Languages,* 2nd ed. Oxford, UK: Wiley-Blackwell.

A complete account of Ladefoged's work in linguistics.

Lewicki, M. S. (2002). Efficient coding of natural sounds. Nat Neurosci 5: 356–363.

A study of speech compared to other natural sound signals in the context of information theory. A tough paper to digest, but one that makes an important point.

Miller, G. A. (1991). *The Science of Words.* New York: Freeman.

A highly readable account of many basic facts about speech and language.

Schouten, J. F., R. J. Ritsma, and B. I. Cardozo (1962). Pitch of the residue. J Acoust Soc Am 34: 1418–1424.

Another account of some relevant psychophysics.

Schwartz, D. A., and D. Purves (2004). Pitch is determined by naturally occurring periodic sounds. Hear Res 194: 31–46.

Paper explaining the details underlying the concepts shown in Figures 3.6 and 3.7.

1. A nice summary of this work is S. S. Stevens (1986).

2. R. M. Warren (1999) presents an overview of Warren's thinking about audition. In keeping with Warren's point about loudness, notes with the same fundamental frequency played in the high register of one instrument (e.g., a trombone) and the low register of another instrument with a higher pitch range (e.g., a piccolo) sometimes sound as if the former is playing a "higher" note than the latter because of the visibly greater effort involved.

3. P. Ladefoged and N. P. McKinney (1963).

4. R. S. Turner (1977).

4

Music and Vocal Similarity

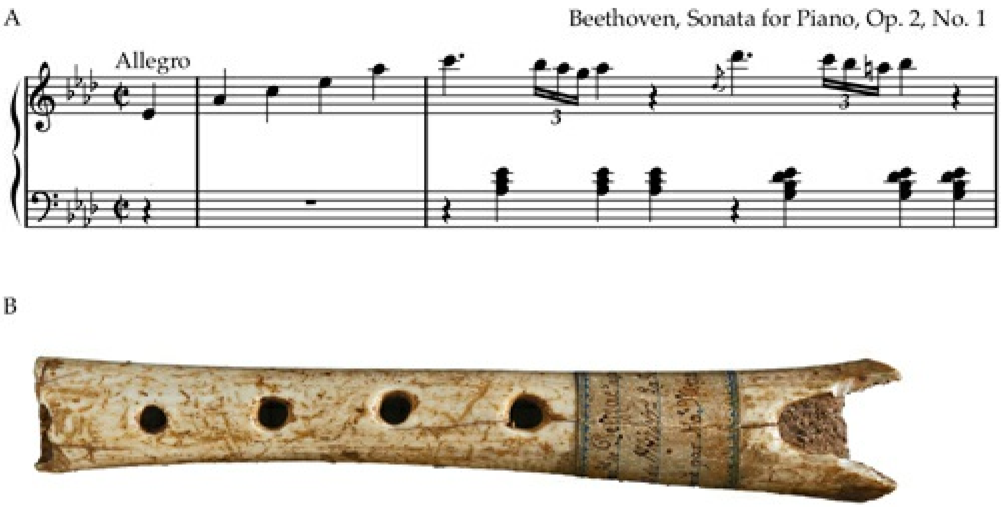
CHAPTER 2 CONSIDERED how otherwise puzzling aspects of loudness and pitch in nonmusical contexts can be rationalized in biological terms, and Chapter 3 reviewed vocalization and evidence that we hear sound signals based on human evolutionary and lifetime experience rather than on their physical parameters. This chapter turns to music and the idea that, based on the biological importance of vocal recognition, the phenomenology of musical tones and tonal relationships can also be understood in empirical terms. Despite a wide range of culturally specific variations, people worldwide tend to use specific tone combinations to create music. Music theory and practice in Western and many Eastern cultures express this shared tendency by the use of octaves divided into the thirteen notes and twelve intervals that define the *chromatic scale*.[1] Subsets of this collection make up the commonly used scales, which typically emphasize only five, six, or seven of these intervals. Despite millennia of study and speculation, the reasons for this parsing are not understood. Equally puzzling is why different scales tend to elicit particular aesthetic and emotional effects. Indeed, why humans are attracted to combinations of tones in the first place remains a mystery. Making some sense of these issues is now the goal, beginning here with observations that link musical tones and their relationships to the harmonic series characteristic of human vocalization.

Defining Music

Although everyone recognizes music, formal definitions are vague. The *Oxford English Dictionary* gives a primary definition as: "The art or science of combining vocal or instrumental sounds to produce beauty of form, harmony, melody, rhythm, expressive content, etc." The definition of music given in the glossary is: "Complex periodic sounds produced by a variety of physical instruments (including the human vocal tract) that are appreciated by humans as pleasing and affective, typically implemented by specific tone combinations in melodies and harmonies." The key concepts in these and other definitions are emotion and aesthetic appreciation, the impetus for all art forms.

The effects of musical tones entail all of the perceptual categories discussed earlier —loudness, pitch, and timbre. In addition, rhythm, tempo, and meter are temporal features of music that link it to motor behavior ranging from toe tapping and clapping to elaborate choreography. Rhythm refers to accentuated beats or subdivisions of beats that correspond to a listener's inclination to tap or clap; tempo is the rate at which beats occur (the number of beats per minute, indicated by a metronome); and meter refers to the organization of the beats (i.e., how many beats there are in each measure, a measure being a "natural" grouping of beats [e.g., in threes, as in a waltz, or fours, as in "straight" or "common" time]).

Although these temporal aspects of music are obviously important, the focus of most music theory over the centuries has been on tones and tonal relationships. A sequence of musical tones (Figure 4.1A) is the basis of *melody,* whereas combinations of tones played more or less simultaneously are the basis of *harmony.* Although the particulars of music have varied greatly over time and in different cultures, the structure of musical instruments dating back tens of thousands of years suggests that music has favored much the same tonal intervals since the dawn of human enthusiasm for this art (Figure 4.1B).



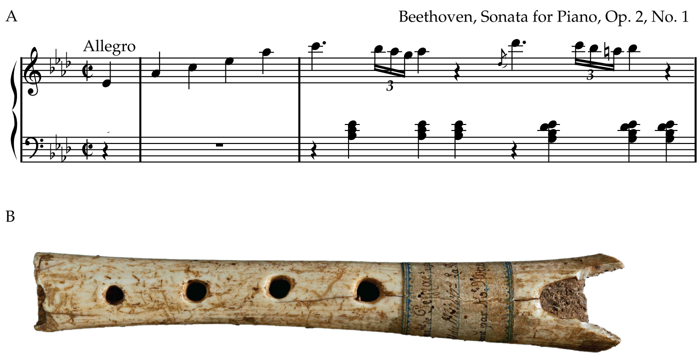


FIGURE 4.1 Some basic characteristics of music. (A) This segment of a Western score indicates several features: a melodic line (tones played sequentially), harmony (tones played simultaneously), the relative duration of notes (e.g., the quarter notes in the first full measure) and the arrangement of the notes within each measure (indicated by vertical lines) that conveys rhythm and meter. "Allegro" alerts the performer to the composer's intention with respect to tempo (i.e., that the section be played "brightly" at a fast tempo). Finally, the segment was composed with a specific *key* in mind, which refers to the low or reference note in the collection of tone combinations (scale) the composer is using. In this example, this note is A-flat, indicated by the key signature at the beginning of the score. (B) This flute was discovered at an archeological site in France and is estimated to be about 32,000 years old. The distances between the holes here and in other more playable ancient flutes suggest that the tone combinations favored today have been preferred for many millennia. Photo © The Trustees of the British Museum. All rights reserved. (After Purves et al., 2013.)

Musical Terminology and the Chromatic Scale

An obstacle to thinking about music for many people (not least the author) is the terminology of Western and other music theory. Music theory is a relatively recent invention and is not essential in the sense that many excellent musicians have no theoretical training and play "by ear." Nonetheless, it is difficult to talk about music without defining and employing the terms that have come into widespread use in recent centuries.

Foremost among these is the *musical interval,* a term that refers to the frequency distance between the low or reference note in a given scale and a higher note, or between any other two notes in a scale (Figure 4.2).[2] An equally critical term is *octave,* defined as the interval between two tones when the fundamental frequency of the first tone is doubled (see Chapter 2). The word comes from the fact that the most commonly used Western scales—the "do, re, mi" major scale for instance—comprises an 8-note subset of the thirteen notes in the *chromatic scale.* [3] A *note* thus refers to a particular frequency and its octave multiples. Since in most circumstances notes separated by one or more octaves are more or less fungible, especially in harmonies, they are given the same letter name (A, B, C, D, E, F, or G). For example, in Figure 4.2, either of the notes at the beginning and end of the octave illustrated could be used without producing a dissonant effect when a C is called for in a combination of tones played together. The two Cs are distinguished only by a qualifier that denotes the ordinal position of the octave in question over the range of a musical instrument (e.g., C_1, C_2, C_3 … C_7 on a modern piano, whose eighty-eight keys span a little more than seven octaves).

The octave illustrated on the piano in Figure 4.2 begins with C, but could have begun on any note, which defines the *musical key* being used. The key is critical because, given the organization of notes on a keyboard or any other instrument, it determines the intervals (frequency distances) between the notes in the scale being used, which may be quite different. For instance, whereas the intervals in any major scale have the same pattern, the pattern of intervals in a minor scale differs from the pattern in a major scale, with perceptual consequences that are critical to effect of a musical piece on listeners (see Chapter 6).

Finally, although the thirteen notes and twelve intervals of the chromatic scale provide a framework for discussing music worldwide, all musical traditions make use of smaller divisions that may be formally specified or not. These *microtones* are specifically notated in traditional Arabic, Persian, and Indian music, giving rise to tonal sets that are larger than the thirteen-note chromatic set in Figure 4.2. But as discussed in Chapters 6 and 7, Western music is also replete with smaller intervals that are for the most part informal embellishments.

Interval Name	Abbr.	Ratio	Increment (cents)	Keyboard
Perfect Unison	P1	1 : 1	0	
Minor Second	m2	16 :15	100	
Major Second	M2	9 : 8	200	
Minor Third	m3	6 : 5	300	
Major Third	M3	5 : 4	400	
Perfect Fourth	P4	4 : 3	500	
Tritone	tt	7 : 5	600	
Perfect Fifth	P5	3 : 2	700	
Minor Sixth	m6	8 : 5	800	
Major Sixth	M6	5 : 3	900	
Minor Seventh	m7	16: 9	1000	
Major Seventh	M7	15 :8	1100	
Perfect Octave	P8	2 : 1	1200	

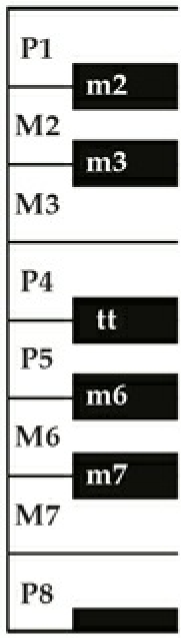


FIGURE 4.2 Western musical intervals over an octave, illustrated on a piano keyboard. Columns indicate the 12 interval names, abbreviations and fundamental frequency ratios in just intonation tuning. The column in *cents* pertains to the logarithmically equal steps over an octave used in the equal temperament system of musical tuning, which is pretty much universal today. Notice that, as decimal values, the just intonation ratios do not always specify the same increments as equal temperament tuning (see text).

If all of this seems complicated, that's because it is! But without these basic concepts discussing tonal music is impossible.

Tuning Systems and Their Problems

The phrase *tuning systems* refers to protocols that spell out exactly how to partition the frequencies over an octave, a critical decision for players and instrument makers alike. When octaves are divided into the twelve ascending intervals of the chromatic scale, each note is incremented by about 6 percent over the previous one, the exact value depending on the system being used. Debates about how to make this twelvefold division in tuning instruments go back thousands of years. Although the history of tuning over the centuries includes many complex variants (estimates by some music historians range as high as 150),[4] only two protocols are routinely considered in contemporary music theory. These are *just intonation tuning* and *equal temperament tuning*.

Justly tuned intervals are determined by the ratio of the fundamental frequency of the note in question to the fundamental frequency of the low or reference note in a scale (see Figure 4.2). The just intonation system is an inheritance of Pythagorean theory, in which the consonance (pleasing quality) of tonal combinations was taken to arise from identical strings whose relative lengths or tensions defined small integer ratios. For reasons of mathematical "purity," Pythagoreans limited these consonant ratios to 1:1, 2:1, 3:2, and 4:3. As described in Chapter 5, the number of intervals in music theory eventually grew to include other small integer ratios, leading to the set of chromatic ratios in Figure 4.2.[5] In this system, the fundamental frequency of each tone in an ascending scale increases by ~4 to 8 percent over the fundamental of preceding note. These proportional increments are called "semitones," even though they differ slightly for some intervals. In the eighteenth century, the French composer and music theorist Jean-Philippe Rameau pointed out that the integer ratios of tone combinations are present in any harmonic series, providing some justification for this system in physical terms.[6]

The problem with just intonation is that when playing in multiple keys, which became common in Western music by the Renaissance, the intervals determined in this way differ slightly, as just mentioned. Because the twelve intervals are not identical, instruments that are in tune in one key (e.g., a scale beginning on A on the piano keyboard) are "out of tune" in another key (e.g., a scale beginning on G). Instruments without frets, such as a violin, or tones sung a cappella, allow skilled performers to make adjustments in real time. But for makers of fretted string instruments, keyboard instruments, flutes, and other wind instruments with fixed holes, this problem presented a nightmare. In the West, the increasing popularity of smaller keyboard instruments like the harpsichord and accordion demanded a solution.

The answer to the problem with just intonation emerged in the Renaissance in Europe (and earlier in China) in the form of equal temperament, a tuning system based on twelve proportionately equal increments over an octave. In this system, the octave is parsed by increasing the frequency of each successive note by exactly $2^{1/12}$, or 5.9463 percent over its predecessor. As a result, the proportionality between the intervals in different keys is maintained. Since this requirement is mathematically unique, equal temperament tuning is universal in most popular and classical music today, and it is here to stay. The only alternative is sticking to a single key in just intonation tuning, and not many musicians or composers want to abide by this limitation.

A problem with equal temperament, however, is that some highly trained musicians find tone combinations on this basis to be less pleasing than combinations over an octave in just intonation tuning. More important from a scientific perspective, whereas the problem of "going out of tune" across keys is resolved by equal temperament tuning, this ad hoc compromise can't explain the tonal phenomena taken up in the following section.

The bottom line is that equal temperament sacrifices a modicum of aesthetic purity for musical flexibility, and few listeners are bothered the difference. From the perspective of understanding music in biological terms, however, the fact that equal temperament works well enough is not much help.

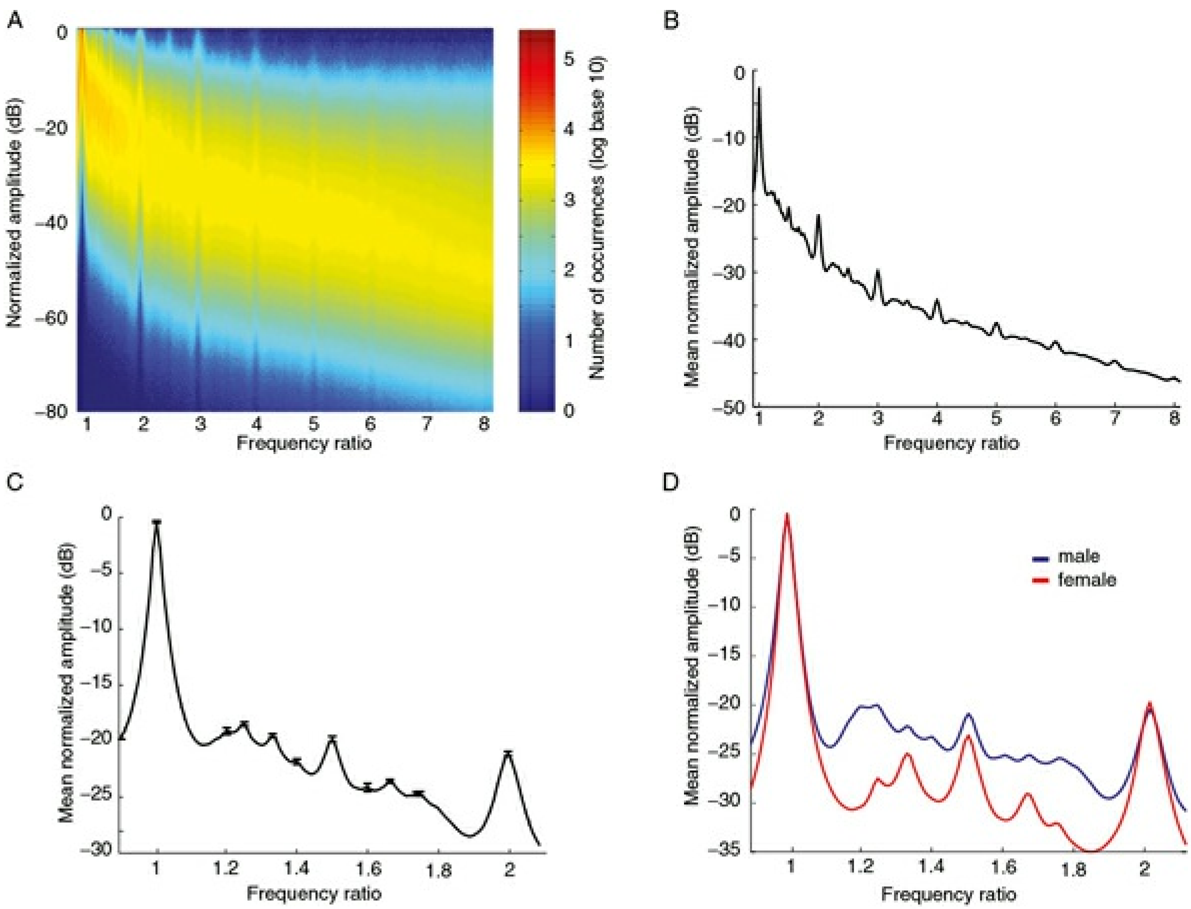
Musical Phenomena That Need to Be Explained

Although the terminology of traditional music theory can be daunting, the bright side is that it provides a means of describing the major scientific challenges in music. Just as loudness and pitch in the absence of music present a series of phenomena that need to be explained (see Chapters 2 and 3), there are puzzles in musical tonality whose solutions would presumably reveal a good deal about music and audition. Why, for instance, are there octaves in music? Why are tones an octave apart more or less equivalent? Why does music tend to parse octaves into twelve divisions? Why does music worldwide tend to use the same subsets of these chromatic scale tones? Why do some tone combinations sound more pleasing than others? Why, despite all the ways that octaves could be divided (billions), are only a few dozen scales widely used in music? Why does the emotional impact of music differ according to the scale used? And why, despite widespread commonalities, does the use of musical tones differ across cultures?

Traditional approaches to these questions have been based on the subjective fact that *pleasing tonal combinations* arise from sources whose fundamental frequencies are related by small integer ratios (see Figure 4.2). A framework based on ratios derived from the pleasing quality of tone combinations is indeed appealing, and it continues to dominate much thinking about these issues. Given the arguments in Chapter 1, however, biology may be a better bet in seeking answers to this long list of largely unexplained musical phenomena. To reiterate, the premise of a biological approach is that the human sense of tonality arose over the course of human evolution to reap the ecological value of recognizing and processing conspecific vocalizations, which are the most important tonal sound signals in our natural environment.

Vocal Sound Signals and Musical Intervals

One way to begin exploring music in biological terms is to ask whether musical intervals are represented in human vocalizations. This may seem an odd way to start, since vocalizations are single tones, not tone combinations. Nonetheless, the harmonics in any series are related by integer ratios, and Figure 4.3 shows that musical intervals are evident in accumulated vocal sound signals. Figure 4.3A graphs the distribution of energy in the compiled spectra of about 100,000 brief segments of voiced (tonal) speech. The mean values are shown in Figure 4.3B, and the blowups in Figures 4.3C and 4.3D show the average concentrations of sound signal energy over a single octave.[7]



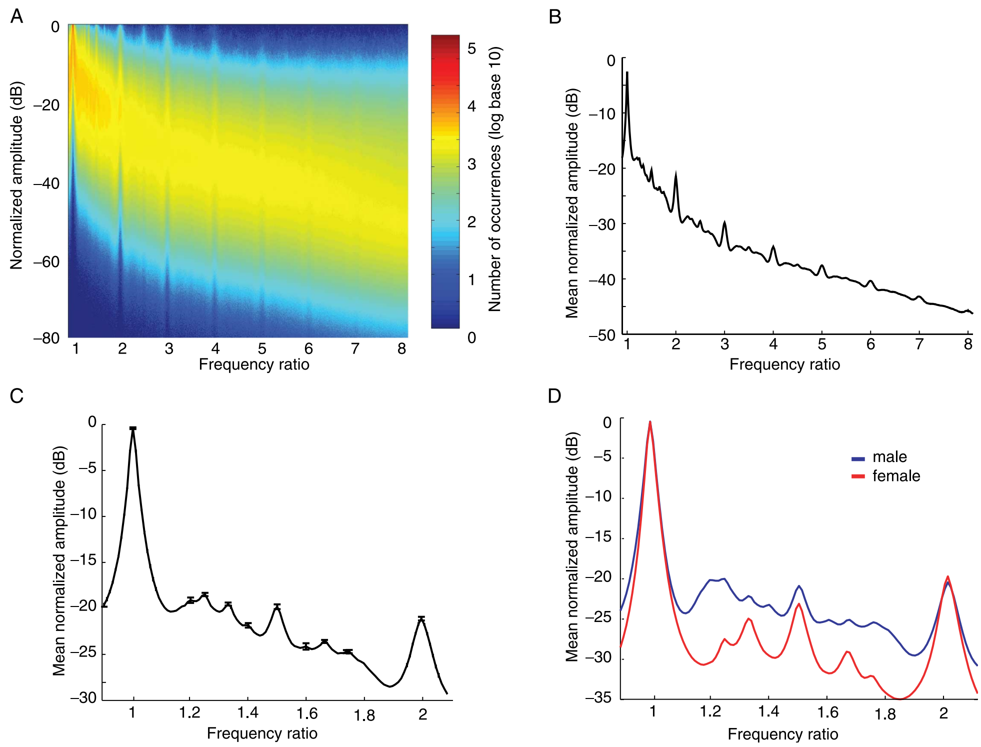


FIGURE 4.3 Spectral characteristics of compiled voiced speech. The analysis is based on ~100,000 American English speech segments drawn from the TIMIT speech database described in Chapter 3. (A) Frequency ratios and their amplitudes, normalized with respect to the frequency of the maximum amplitude in each speech segment. (B) Mean values over the same range. (C) Magnification of the plot in (B) over a single octave. (D) Data in (C) shown separately for male and female speakers. (From Schwartz et al., 2003.)

Figure 4.4 shows in turn that these energy peaks closely correspond to the interval ratios in Figure 4.2. Only the three least consonant intervals in the chromatic scale (the minor second, major seventh, and major second) are missing, presumably because their effects are obscured by the large peaks at unison and the octave. Thus, even though the fundamental frequencies of pitches in human vocalization (vocal prosody) don't represent musical intervals as such (see Chapter 5), musical intervals are apparent in the energy distributions of compiled speech sound spectra.

To make sure that these results are not particular to American English, the same analysis was carried out for speech segments drawn from Farsi, Mandarin, and Tamil. Although the averages differ somewhat across languages, the location and relative prominence of the bumps in the patterns are much the same as those in Figure 4.4. Thus, as might be expected from the anatomical similarity of the larynx and the rest of the vocal tract among humans, these characteristics of voiced speech are largely independent of the language spoken.

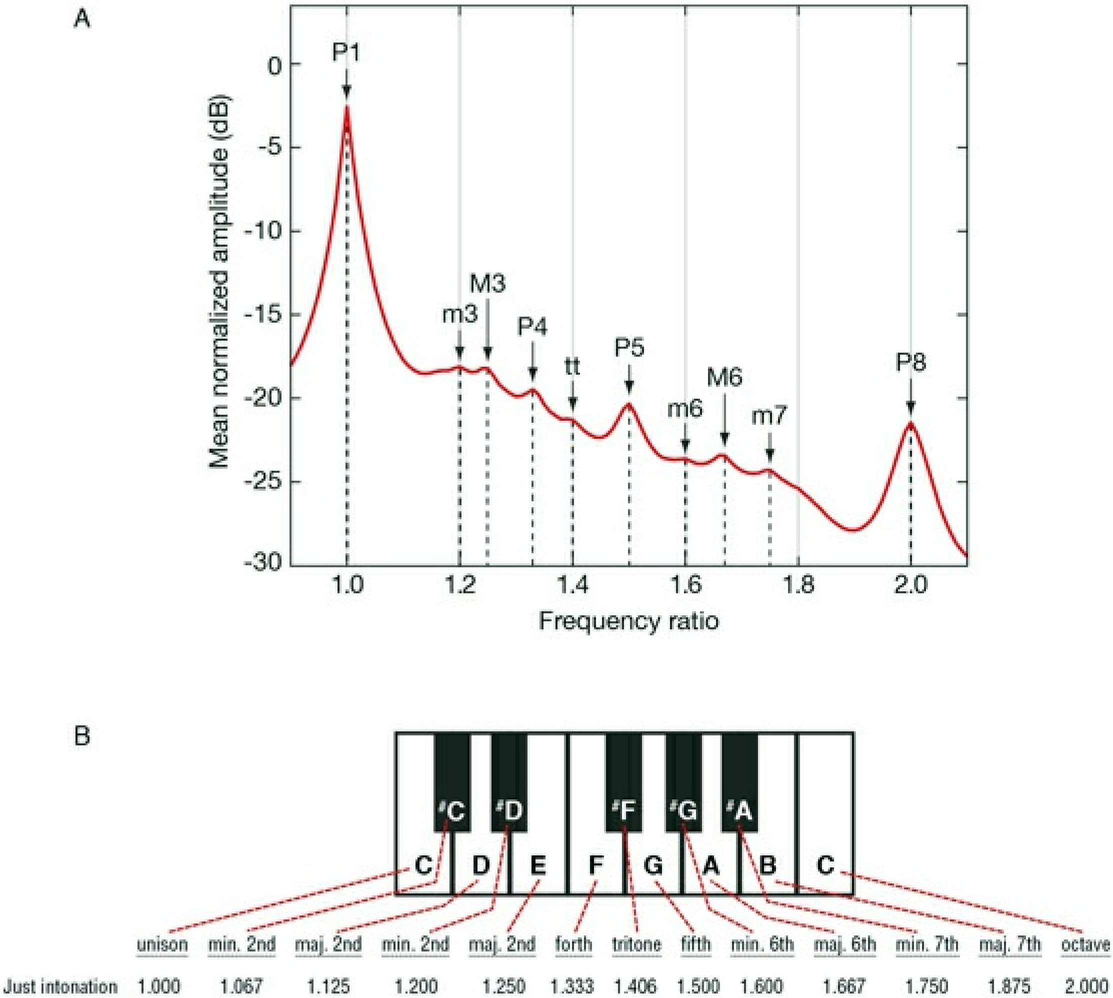


FIGURE 4.4 Comparison of the spectrum of voiced human speech sounds and the intervals of the chromatic scale. (A) The chromatic intervals indicated the arrows correspond to the amplitude peaks in the normalized spectrum of speech sounds over a single octave, redrawn from Figure 4.3C. (B) The names of the musical intervals and the frequency ratios corresponding to each peak in just intonation tuning are shown on a piano keyboard for comparison. (From Schwartz et al., 2003.)

It seems unlikely, however, that the mere presence of harmonic ratios or musical intervals in compiled human vocalization explains the attractiveness (consonance) of particular tone combinations. Something is clearly missing.

Vocal Similarity as a Basis for Attraction to Tone Combinations

In most music theories, the consonance of tone combinations whose fundamental frequencies are related by small integer ratios is taken as axiomatic; that is, no underlying reason is given for this connection between ratios and musical appeal. The argument here is that the missing piece in the puzzle is supplied by tonal music conceived in biological terms. The idea is that how closely the harmonics in tone combinations match the uniform harmonic series that characterizes human vocalization determines the relative consonance of any tone combination. Thus rather than ratios, the metric of consonance is vocal similarity: the greater the resemblance of a tone combination to the harmonic series characteristic of voiced vocalization, and the greater its attraction for the listener.

Considered in this way, the only tone combination that perfectly matches the uniform harmonic series that characterizes voiced vocal sounds is *unison* (a note played with itself) . A note with harmonics at twice the fundamental frequency of the reference note (an octave) corresponds with every other harmonic of the low note, thus duplicating (and emphasizing) half the harmonics in that combination. A note with harmonics spaced at intervals three times the fundamental of the reference note (a perfect fifth) duplicates every third harmonic; a set with harmonics at four times the fundamental of the reference note (a perfect fourth) duplicates every fourth harmonic, and so on. Notice, however, that when the reference note of a scale is played with any note other than itself or the octave, the combination also introduces non-corresponding harmonics that are also important in explaining aspects of tonal phenomenology (see Chapters 5 and 6).

This way of looking at tone combinations provides some insight into the emergence of chromatic intervals in accumulated speech signals (see Figure 4.4). When the harmonic series of a large number of speech sound signals are compiled, the harmonic coincidences would tend form peaks of vocal energy, whereas the non-coincidences would not. A seeming confound to a listener's appreciation of this fact, however, is that voiced speech sounds are single tones, not tone combinations. We don't generally hear compiled speech segments.

Although any voiced utterance is indeed a single harmonic series, most phones last on the order of 100 milliseconds and come in bunches that make up syllables, words and sentences, or the antecedents of grammar and syntax in pre-lingual vocalizations. Because a person's average fundamental frequency varies over an octave or more, it would not have taken ancient humans long—or any of us as toddlers—to recognize from our own utterances that some tone combinations in vocal sequences sound more appealing than others, leading to a crude version of melodic combinations and eventually to vocalizing in consonant tone sequences (singing) just for the fun of it. The underlying reason for the pleasure derived would be the biological advantages that accrue from recognizing and responding to conspecific vocalizations.

Conclusion

The intervals of the chromatic scale are present in harmonic ratios and compiled experience with human vocalizations. Thus, even in the absence of music as we normally think of it, humans have always been exposed to tone combinations. Exposure, however, doesn't explain why we find some tone combinations to be more attractive than others. The argument here is that the degree of harmonic correspondence between any two tones conveys the degree of their combined similarity to a full harmonic series, and thus the relative biological value of a sound signal to a listener. The question explored in the following chapters is whether this attraction to the vocal similarity of tone combinations can account for the full range of phenomenology evident in tonal music.

Additional Reading

Burkholder, J. P., D. J. Grout, and C. V. Palisca (2014). *A History of Western Music,* 9th ed. New York: Norton.

A detailed account of the development of music in the West.

Cook, P. R., ed. (2001). *Music, Cognition, and Computerized Sound: an Introduction to Psychoacoustics.* Cambridge, MA: MIT Press.

A collection of generally fine essays on music and psychoacoustics.

Duffin, R. W. (2007). *How Equal Temperament Ruined Harmony (and Why You Should Care).* New York: Norton.

An engaging account of the history of tuning systems, emphasizing ongoing concerns with use of equal temperament tuning.

Hillenbrand, J., L. A. Getty, M. J. Clark, and K. Wheeler (1995). Acoustic characteristics of American English vowels. J Acoust Soc Am 97: 3099–3111.

An account of the intricacies of voiced speech.

Isacoff, S. (2001). *Temperament: The Idea That Solved Music's Greatest Riddle.* New York: Knopf.

A perspective opposite that of the Duffin book; Isacoff argues that equal temperament tuning is invaluable (see the Mark Allen Group online review in *Gramophone* ("The Tuning Wars") for an interesting assessment of these two books).

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch.* New York: Oxford University Press.

A traditional overview of the phenomenology of musical pitch.

Patel, A. D. (2008). *Music, Language, and the Brain.* New York: Oxford University Press.

A broad account of how language and music are related; the book also reviews the importance of rhythm and other issues not covered here.

1. As mentioned in Chapter 3, there are exceptions to this generalization, as in Javanese and African music played on metallophone or lamelleophone instruments that don't produce the same harmonic series generated by

string and wind instruments.

2. The related term *scale degree* is represented by a number adorned with a caret for each sequentially higher note.

3. The origin of term “chromatic” is unclear, but it dates back many centuries in Western music theory and refers to the perceived quality or “color” of these thirteen different notes in music.

4. J. M. Barbour (1951).

5. The stipulations of music philosophy and theory are of course not the same as musical practice. Folk musicians in ancient Greece or anywhere else presumably played and sang whatever pleased them.

6. J.-P. Rameau (1722). Rameau was one of the first theorists to suggest that musical consonance might be based on vocal similarity.

7. For details see D. A. Schwartz, C. Q. Howe, and D. Purves (2003).

5

Consonance and Dissonance

A FUNDAMENTAL QUESTION about the perception of music—and arguably the question at the root of all tonal phenomenology—is why, in all cultures, some tone combinations are perceived as relatively consonant or "harmonious," and others as relatively dissonant or "inharmonious." These perceived differences are the basis for harmony when tones are played simultaneously and for melody when they are played sequentially. In Western and much other music, the most agreeable combinations are often used to convey a sense of tonal "home base" at the beginning of a musical piece or a section of it, and a sense resolution at the end of a piece or section. In contrast, less harmonious combinations are typically used to evoke a sense of tension or of transition in a chord or a melodic sequence. These phenomena, like the tonal preferences in the context of scales (Chapter 6), the emotions elicited by different scales (Chapter 7), and cultural differences in tonal usage (Chapter 8) have to do with *aesthetics.* Although sometimes obscured by experts expounding on unique qualities of an antique violin or a fashionable work of art,[1] in scientific terms aesthetics boils down to either liking or disliking sensory inputs, based ultimately on their relevance to biological (reproductive) success. We tend to be attracted and attentive to stimuli that contributed to our success as a species and have an aversion to those that were less helpful or irrelevant. Thus the idea of biologically determined consonance introduced in Chapter 4 and the alternatives warrant a closer look.

Assessing Consonance and Dissonance

Although definitions of auditory consonance vary, the most widely accepted one in the context of tonal music is a pleasing combination of periodically repeating (tonal) sound signals. To bring science to bear on the issue, a number of studies in the late nineteenth and early twentieth centuries focused on ranking the relative pleasantness of musical tone combinations heard by listeners.[2] Since combinations of the sine tones described in Chapter 1 are inherently somewhat unpleasant, these studies of consonance tested two-note combinations (dyads) played on a piano, violin, or other musical instrument.

The relative consonance of the tone combinations in the chromatic scale determined in this way is shown in Figure 5.1. The reason why humans find some of these combinations in harmonies and melodies more pleasing than others has been debated for millennia, but despite ongoing interest, there has been no consensus about the basis of consonance.

Consonance Based on Mathematics

Discussions of consonance usually begin with the ideas of the Greek mathematician and philosopher Pythagoras in the sixth century BCE. As mentioned in Chapter 4, Pythagoras is said to have demonstrated that tone combinations generated by two plucked strings whose lengths and tensions defined small integer ratios are especially pleasing. (The validity of this story is uncertain, since the accounts of it are secondhand, primarily from the writings of the Greek mathematician Nichomachus, who lived several centuries later). For both mathematical and philosophical reasons, Pythagoras and his school limited their concept of pleasing tone combinations to unison (1:1), the octave (2:1), perfect fifth (3:2), and perfect fourth (4:3), ratios that had spiritual and cosmological significance in the Pythagorean worldview.

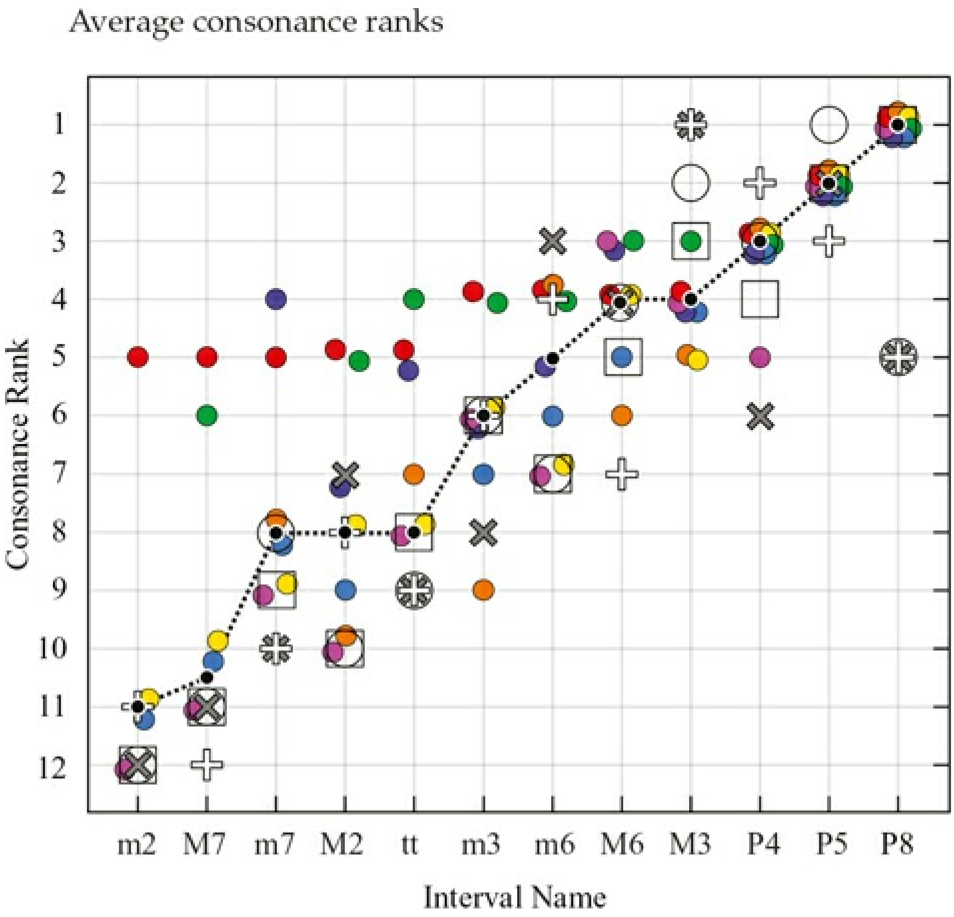


FIGURE 5.1 The relative consonance of two-tone combinations (dyads) assessed in psychophysical studies. The graph shows the relative consonance assigned by listeners to 12 of the 13 notes in the chromatic scale played in combination with the low (reference) note of the scale (unison was not considered in these studies, but would presumably have been the most consonant combination). The black dots connected by the dashed line show the median values of the 8 studies compiled here, each indicated by a different symbol. (From Bowling and Purves, 2015; see also Malmberg, 1918.)

The list of Pythagorean ratios was extended in the early Renaissance by the Italian composer Gioseffo Zarlino. Zarlino eased the Pythagorean constraints in order to include the numbers 5 and 6, thus accommodating the major third (5:4), minor third (6:5), and major sixth (5:3), which had become popular in the polyphonic music of the late Middle Ages. Echoing the Pythagorean outlook, however, Zarlino's rationale included the numerological significance of 6, which is the first integer that equals the sum of all the numbers of which it is a multiple ($1 + 2 + 3 = 1 \times 2 \times 3 = 6$).[3] Additional reasons included the natural world as it was then understood (six planets could be seen in the sky) and Christian theology (the world was created in six days).

Although the Pythagorean take on consonance has long been derided as numerological mysticism, the coincidence of numerical simplicity and pleasing perceptual effects continues to influence concepts of consonance and is hard to ignore. The idea that tone combinations are pleasing because they are mathematically simple, however, raises the question of why simple should be pleasing. And theories of consonance based on mathematical simplicity have no good answer.

Consonance Based on Physics

As a result, more recent explanations of consonance have been based on the physics of sound signals at the ear. Although a physical basis for consonance was suggested by Jean-Phillipe Rameau in the eighteenth century (see Chapter 4), the dominant figure who promoted a physical interpretation of consonance was the nineteenth-century polymath Hermann von Helmholtz. Helmholtz argued that the dissonance of a tone combination depends on the degree to which it exhibits what he called "beating and roughness." He pointed out that the physical bumpiness of a sound signal produced by the interactions among the harmonics of tone combinations is disagreeable and concluded that this effect is the "true and sufficient cause of consonance and dissonance in music."[4] Thus, in Helmholtz's view, consonance is simply the relative absence of the "rough" or beating quality tone combinations.

Helmholtz was not the first to come up with this general idea. Even before Rameau, a link between auditory roughness and dissonance was suggested by Isaac Beeckman, a seventeenth-century Dutch philosopher who was, among other things, a teacher of Descartes. Like Helmholtz, Beeckman concluded that a tone combination sounds less and less pleasant as beating and roughness become increasingly salient. Similarly, the early-eighteenth-century French mathematician Joseph Sauveur also thought that when roughness cannot be heard, musicians take tone combinations to be consonant.[5]

It was not until the nineteenth century, however, that Helmholtz provided a full analysis of beating and the perception of roughness as the basis of consonance and dissonance. He reasoned that when two or more musical tones are combined, beating and roughness arise not only from the interaction of their fundamental frequencies, but from interactions among their harmonics as well. Helmholtz knew that when combined two harmonic series would have a characteristic "beating" pattern due to constructive and destructive interference (Figure 5.2). In accord with his hypothesis, the less beating and roughness the greater the consonance heard by listeners (see Figure 5.1).[6]

In this understanding of consonance and dissonance the degree of perceived dyadic roughness depends mainly on the frequency of the periodic fluctuations in Figure 5.2. When the repetition rate of the bumps is relatively low (about one to six per second) the fluctuations are, in Helmholtz's words, "by no means disagreeable to the ear." But when the rate increases to about fifteen to thirty per second, "the collective impression … [is] jarring and rough."[7] As the rate of fluctuation increases still further, the sense of roughness decreases and the impression of the sound signal becomes smooth and “agreeable” again.

The physical reason for these phenomena is straightforward. Since consonant dyads have fundamental frequencies related by smaller integer ratios, this fact means that consonant dyads have relatively high repetition rates compared to dissonant dyads. For example, the waveform of a dyad arises from combined tones an octave apart (a frequency ratio of 2:1) starting on middle C (262 Hz) repeats ~262 times a second; in contrast, when the combined waveform is that of a minor second (16:15 in just intonation tuning) the rate of repetition is only ~16 times a second (see Figure 5.2). Thus, the faster repetition rates of consonant dyads fall in a range perceived as “smooth,” whereas the repetition rates of dissonant dyads fall in a range perceived as “rough.”

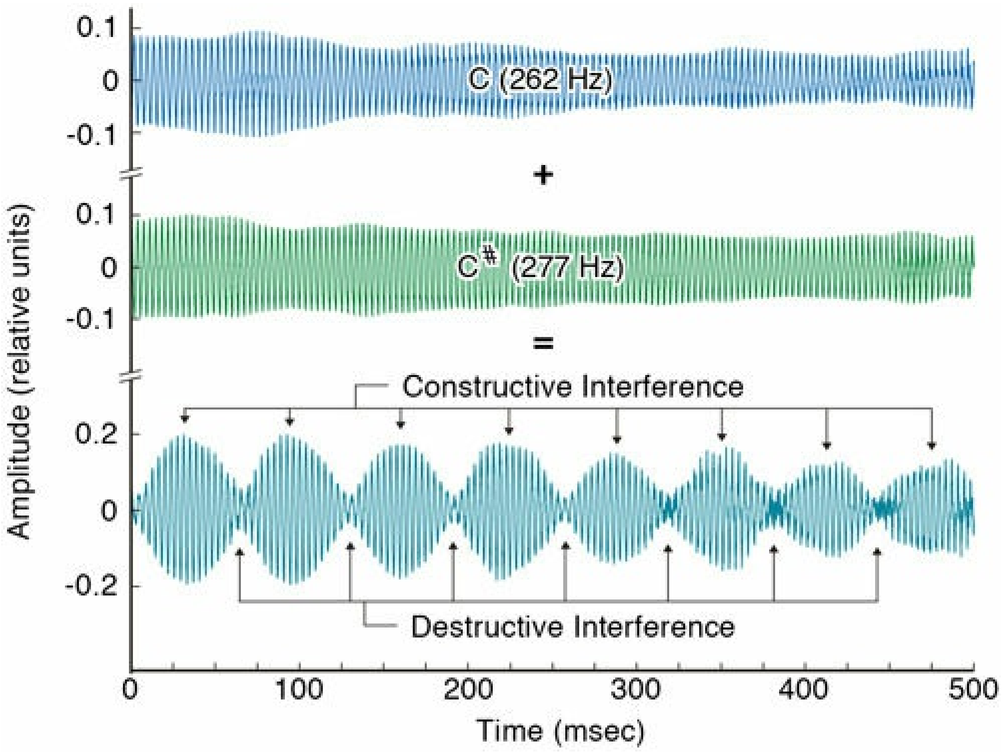


FIGURE 5.2 The physical basis of “beating and roughness” in tonal dyads. The waveforms of two notes—middle C (262 Hz; blue) and C# (277 Hz; green)—played on an organ; this tone combination is a minor second, the most dissonant dyad in the chromatic scale, as indicated in Figure 5.1. When these two tones are combined, alternating periods of constructive and destructive interference result in a periodic fluctuation in the sound signal’s amplitude. In Helmholtz’s interpretation this bumpiness is the cause of perceived dissonance, and its absence the basis of perceived consonance. (After Bowling et al., 2012.)

The relative simplicity of this hypothesis (and no doubt Helmholtz’s scientific stature) led to further studies that generally endorsed his claims using newer methods and observations. With respect to auditory mechanisms, Helmholtz had suggested that the perception of roughness arises from interactions among the vibrations at different frequencies along the basilar membrane of the inner ear. In this conception, roughness is perceived when two frequencies are close enough to stimulate overlapping regions of the membrane. Georg von Békésy's work on this issue in the 1940s and 1950s (see Chapter 2) made it possible to compare frequency resolution with the measured effects of different stimuli on the basilar membrane. At the same time, more detailed psychophysical studies gave rise to the complementary idea of a "critical bandwidth," which refers to the frequency distance within which two tones begin to interact when assessed psychophysically (which corresponds to about 1 mm along the basilar membrane, or about ~3 percent of its overall length; see Chapter 2). These observations further supported Helmholtz's roughness theory.

Other studies later in the twentieth century put physical explanations of consonance on a somewhat different footing. Pattern recognition theories, for instance, supposed that judgments of consonance involve higher-level processes in the auditory system. The idea is that both the physical roughness of sound signals at the ear and stored information derived from past experience play a role in perceived consonance and dissonance, a concept that accords with obvious differences in musical tastes across cultures. The most prominent exponent of this perspective was the German acoustician and engineer Ernst Terhardt, who suggested in the 1970s that the perception of musical intervals derives from familiarity with "specific pitch relations" among the frequencies of the lower harmonics of complex tones.[8] Like accounts based on physical interactions along the basilar membrane, however, Terhardt's idea of "tonal meanings" remains focused on the physical characteristics of sound signals.

In the end, however, proposals that consonance depends on higher order processing, or more subtle aspects of tonal sound signals such as the distribution of amplitudes or frequencies, don't do much to change Helmholtz's basic idea that consonance is first and foremost the absence of physical roughness.

Problems with the Roughness Theory

Despite ongoing enthusiasm for some version of the theory that consonance is the absence of roughness, problems with this perspective have become increasingly apparent. One confound is the awkward fact that dissonance is still perceived when the two tones of a dyad are presented independently to the two ears (e.g., the low or reference note of a scale to one ear and a note a semitone higher to the other, forming the dissonant minor second interval). In this circumstance, there is no physical beating at the level of basilar membranes of either ear.

Another problem is that most studies of roughness examined dyads, but not more complex tone combinations. When chords with more than two tones are used, additional harmonics that increase physical bumpiness don't necessarily reduce consonance. An example is a major seventh dyad (any note and a note a major seventh above it) and a major seventh tetrad (the same note together with a major third, perfect fifth, and major seventh above it). Although the greater number of interacting components in the tetrad increases its physical roughness at the ear, most listeners rate the tetrad as more consonant than the dyad.

Yet another issue concerns the waveforms of a major triad or a major seventh tetrad. These chords are both perceived as consonant, although they repeat about fifty-two and about ten times a second, respectively. Both these repetition frequencies are relatively rough, but don't generate dissonance. A related observation is that removing the harmonics responsible for the roughness of a chord does not necessarily increase its consonance. Finally, timbre and intensity (the latter called the *dynamic* in the context of music) affect consonance and dissonance without affecting roughness. Composers concerned with the orchestration of a piece know that a dissonant interval played slowly by violins in a low register may sound more consonant than a theoretically more consonant major triad played loudly by piccolos in a high register.

Perhaps the most damning evidence against the roughness theory of consonance and dissonance has come from recent studies that distinguish the perceptions of roughness from consonance, confirming that whatever its cause, consonance is not simply based on absence of physically interfering tones. Psychologist Andrew Oxenham and colleagues examined the relationship between consonance and roughness by asking subjects to rate the "pleasantness" of consonant and dissonant chords, using the difference between their ratings as a measure of their preference for consonance.[9] Participants with a strong consonance preference rated consonant chords as more pleasing than dissonant chords, whereas participants with a weak consonance preference rated consonant and dissonant chords as being more or less similar. These and other results showed that consonance preferences are only weakly related to an aversion to roughness, suggesting that these two aspects of tone perception are independent sensory qualities. In another study, the same stimuli were presented to participants with *congenital amusia,* an extreme form of what is referred as "tone deafness." In contrast to a control group, the tone-deaf individuals showed smaller differences between ratings of consonant and dissonant chords, although they didn't differ from the control group with respect to the perception of acoustic roughness.[10] The fact that these subjects exhibited *abnormal* consonance perception but *normal* roughness perception further weakens the idea that the absence of roughness is the basis of consonance. Indeed, an aesthetic effect based simply on the absence of "jarring," as Helmholtz put it, seems dubious in principle.

All told then, a physical explanation of consonance as the absence of roughness is untenable. Nevertheless, roughness *does* track the perceived dyadic dissonance, raising the question of why this should be. The answer may be that roughness in any sound signal is indeed "annoying" and in that sense unpleasant, whether it is heard in response to tonal music or the jackhammer in a nearby construction project. It doesn't follow, however, that the absence of unpleasantness should define pleasantness. This proposition would be analogous to claiming that sugar tastes sweet because it is not sour. Arguments promoting consonance as the absence of roughness neglect the fact that consonance (or sweetness) can be appreciated quite apart from roughness (or sourness).

A Biological Rationale for Consonance and Dissonance

From a biological perspective, the lack of attraction to "rough" sound signals in music arises because amplitude fluctuations below ~50 Hz (the lowest fundamental frequency produced by a large human male) are increasingly inconsistent with human vocalization and imply a diminished reward. Thus consonance seems better explained empirically by past experience with the relative similarity of auditory stimuli to human vocalizations and their biological payoff than by the physical characteristics of sound signals.

Of course a signal that is simply rough (e.g., the growl of a predator) may be just as important biologically or more so than a tone combination. But such signals require a different sort of response, just as sweet and sour tastes call for different behavioral reactions (swallowing something versus spitting it out). In both cases, one class of stimuli is more attractive than others for reasons of basic biology (e.g., a rewarding source of energy if ingested, or a potentially rewarding source of social information if attended)

Other lines of evidence also support a biologically determined interpretation of tonal preferences. First, humans and many other animals (including nonhuman primates) are specifically attracted to conspecific vocalizations compared to other sound signals. Second, the human external ear, ear canal, inner ear, and auditory processing circuitry are all optimized to transmit signals in the range of human vocalizations (see Chapters 1 and 2). Third, as described in Chapter 3, a number of nonmusical pitch phenomena can be explained in terms of the specialization of the human auditory systems for processing vocal sounds. Finally, the frequency range of voiced human vocalizations more or less overlaps the frequency range of tonal music (roughly 50 to 5000 Hz).

If an appreciation of tonal sounds has indeed arisen from the benefits that accrue from attending and responding to conspecific vocalizations, the implication for musical consonance is simply that the more voice-like a musical stimulus is, the more we should be drawn to it. Further evidence for this explanation of consonance comes from the issues discussed in the following chapters. These include the musical scales that humans like, the different emotional effects of major and minor music, cultural preferences for specific tone collections, and core musical phenomena such as the octave. All are arguably determined by the harmonic characteristics of human vocalization and their significance for biological well-being.

Conclusion

The reasons for the relative consonance and dissonance of tone combinations have been debated for centuries without resolution. The focus over most of the past 150 years has been on some version of Helmholtz's theory that these perceptual qualities are determined by the presence or absence of the physical roughness that can occur when the harmonics of two or more tones interact. Given the sum of evidence, however, it seems unlikely that musical consonance is due to the absence of roughness, or any other explanation based on the physical nature of sound signals or their peripheral processing. A more plausible interpretation is that humans are especially attracted to sound signals whose harmonic characteristics imply human vocalization as the source.

Additional Reading

Bilsen, F. A. (1995). What do dichotic pitch phenomena tell us about binaural hearing? In: *Advances in Hearing Research* (G. A. Manley, G.M. Klump, C. Köppl, and H. Fastl, eds.). Singapore: World Scientific, pp. 334–341.

A review of experiments in which tone combinations are presented separately to the two ears.

Bowling, D. L., and D. Purves (2015). A biological rationale for musical consonance. Proc Natl Acad Sci USA 112: 11155–11160. doi: 10.1073 / pnas.1505768112

A fuller account of consonance and dissonance theories discussed here.

Christensen, T. (1993). *Rameau and Musical Thought in the Enlightenment.* Cambridge, UK: Cambridge University Press.

A history of Rameau's ideas about consonance.

Deutsch, D. (ed.) (2013). *The Psychology of Music,* 3rd ed. New York: Academic Press.

A compilation of good reviews, several of which are pertinent to the issues covered here.

Forster, C. M. L. (2010). *Musical Mathematics.* San Francisco: Chronicle Books.

A detailed account of Pythagorean history and many other mathematical issues in music.

Helmholtz, H. L. F. (1877 / 1954). *On the Sensations of Tone as a Physiological Basis for the*

Theory of Music (A. J. Ellis, trans.). New York: Dover.

Helmholtz's ideas on consonance and dissonance, and many other auditory issues.

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch.* New York: Oxford University Press.

A review of consonance theories in cognitive terms.

1. C. Fritz, J. Curtin, J. Poitevineau, H. Borsarello, I. Wollman, F.-H. Tao, and T. Ghasarossian (2014). This article shows that expert opinions on aesthetic qualities may have less value than we imagine.
2. C. F. Malmberg (1918) provides an account of most of the psychophysical studies illustrated in Figure 5.1.
3. C. V. Palisca (1961).
4. H. L. F. Helmholtz (1877 / 1954).
5. Palisca (1961).
6. Given the variance of consonance judgments in Figure 5.1, this relationship is only approximate.
7. Helmholtz (1877/1954).
8. E. Terhardt (1984).
9. J. H. McDermott, A. J. Lehr, and A. J. Oxenham (2010).
10. M. Cousineau, J. H. McDermott, and I. Peretz (2012).

6

Musical Scales

SCALES (OR "MODES" IN SOME CONTEXTS) are collections of tones (notes) separated by specific intervals that various cultures have used, knowingly or not, to make music for thousands of years. Whereas the chromatic scale discussed in Chapter 4 defines the superset of intervals in much Western music, the scales routinely used are subsets of this or a similar palette in other cultures. Compositions in Western classical, folk, and popular music, as well as in many Eastern, African, and other musical traditions, are based on relatively few of these more specific tone collections, typically emphasizing the same six- to eight-note sets (i.e., five- to seven intervals). Thus, another fundamental question is why humans across most musical traditions prefer these tonal groupings. This self-imposed limitation is particularly puzzling, since the human auditory system can distinguish hundreds of pitches over an octave, meaning that in principle billions of scales could be used in making music. Why musical practice includes only a few dozen of the enormous number of possible tone collections humans might have used is not known. Based on the arguments in Chapters 4 and 5, however, a good guess would be that the intervals that define musical scales are those that provide the greatest collective similarity to human vocalization. This chapter examines the evidence for this idea.

Defining Musical Scales

Although music theory is not essential to performing music, some additional nomenclature is needed for the discussion here. Basic to this descriptive canon is the *octave* and its division into musical *scales* (*scala* in Greek means "ladder"). Scales are sets of frequency distances (intervals) over an octave that the composers, musicians and listeners have preferred. As introduced in Chapter 4, the overall set of intervals over an octave used in Western and much other music worldwide is the thirteen-note, twelve-interval chromatic scale (see Figure 4.2). The frequency increment of each tone with respect to the preceding one going up the chromatic scale is called a *semitone.* As already mentioned, a semitone is an approximately 6 percent increase over each previous step on the chromatic ladder, the exact value or values depending on the tuning system being used. In practice, however, composers and musicians generally emphasize a subset of the chromatic scale tones in any particular piece. The chromatic intervals formally excluded from a scalar subset are not prohibited, but are used more sparingly and in special ways (e.g., as "passing tones").[1]

The determinant of any scale is the frequency distance separating the notes when arranged in ascending order, beginning with the low or reference note of the set. A familiar example in Western music is the *major scale* illustrated on the piano keyboard in Figure 6.1. This scale is the "do, re, mi" series learned in elementary school, and in this example the low "do" note is a C. The five intervening notes (the black keys in this particular case) complete the chromatic scale by generating the semitone intervals between the adjacent white keys, which are a *whole tone* (two semitones) apart. These intervals are called flats or sharps, depending on the musical context; in each case, the sharps are a semitone higher and flats are a semitone lower than the adjacent note in a scale, and as a group are called "accidentals." As can be heard when playing or singing this scale, the frequency distance from "do" to "re" is two semitones (a *whole tone*), as is the distance from "re" to "mi," "fa" to "sol," "sol" to "la," and "la" to "ti"; the other distances—"mi" to "fa" and "ti" back to "do"—are semitone intervals.

The low or reference note of this or any scale has special importance. Even musical beginners quickly appreciate that this note is the "home base" of any simple piece and that coming back to this note (or a chord based on it) provides a sense of resolution when it appears at the end of a composition or a section within it (called the *cadence* of a musical piece). Why the low note plays this role is yet another musical puzzle to be explained.

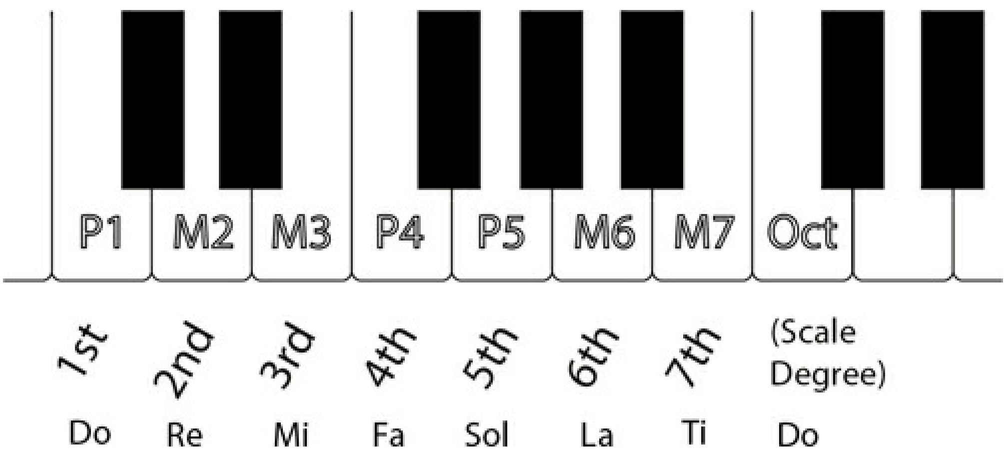


FIGURE 6.1 The major scale as an example. The 7 white notes over an octave that begin with C on a piano keyboard define a C major scale; the 5 black keys are the other notes of the 12-interval / 13-note chromatic scale over an octave. The organization of the keyboard, however, was not derived a priori, but reflects the musical intervals that people like, arranged empirically over the last few centuries in ways that facilitate playing. The ascending notes are given numbers (called scale degrees), and named "do, re, mi …" in the "solfeggio" system used in Western musical teaching. In the formal scheme indicated on the keyboard, M stands for major, P for perfect (meaning a terminating fraction in Pythagorean or just intonation tuning), and Oct for the octave (or P8); thus, M2 is a major second, M3 a major third, P4 a perfect fourth, and so on. The reference note is the low note in the scale to which the higher notes are referred; the same note is called the root note or "tonic" in a chord.

The C major scale in Figure 6.1 is often used as an example because only the white keys are entailed. But this fact can also be confusing, and should not be taken to imply that major scales on a keyboard use only white keys. Were the major scale to begin on a different low note, black keys would be in play. What defines any scale and its musical impact is the sequence of whole tone and semitone intervals it comprises, and thus the intervals drawn from the chromatic superset that are going to be emphasized in a musical piece. Why the intervals emphasized should make such a difference presents another question to be answered.

The Number of Scales

Perhaps as many as 100 scales with different arrangements of semitone and whole-tone intervals over an octave are used in music worldwide today—or more depending on one's taste for detail. The wiggle room in specifying the number of scales in use arises from the fact mentioned in Chapter 5 that all traditions make use of smaller intervals, either formally by specifically notated quarter-tones (e.g., in traditional Persian music), or informally (as in classical Indian music, American blues music, or jazz improvisation). The basic scales used in Western music over the past few centuries are the major and minor *pentatonic* and *heptatonic* scales illustrated in Figure 6.2, which are also prevalent in traditional Indian, Chinese, and Arabic music. The other scales shown are less common, but are used in early liturgical music, folk music, modern jazz, and some classical compositions. In general, pentatonic scales tend to be used in simpler popular ("folk") music, and heptatonic scales in more formal ("classical") compositions.

These cultural and historical facts present another puzzle. Given that listeners can distinguish about 240 frequency distances over an octave in the middle range of hearing, there are an enormous number of ways to divide octaves into five or seven tonal intervals. Using this value of discriminable tones over an octave, the number of possible seven-interval combinations is more than 10^{11}. Even if the number of discriminable intervals over an octave were just twelve, the number of possible scales would still be overwhelming (millions).

Not surprisingly, then, lots of people have grappled with the question of what makes a small number of six- to eight-note (five to seven intervals) scales special. One idea dating back to Pythagoras is that these scalar sets are prevalent because, when a higher note in the scale played with the reference note, the fundamental frequencies of the two tones define relatively small integer ratios (see Figure 4.2). This idea, however, breaks down because only some of the harmonic ratios of the seven notes in the major scale meet this criterion. Whereas unison at 1:1, the octave at 2:1, the perfect fifth at 3:2, and the perfect fourth at 4:3 follow the rule in Pythagorean tuning (tuning according to a series of perfect fifths), the ratio of other intervals do not. For instance, the major seventh in Pythagorean tuning is an ungainly 243 / 128, and the minor second 256 / 243. Moreover, even though their consonance is not detectably different for most listeners, with the exception of unison and the octave none of the ratios in equal temperament tuning used today entail small whole numbers (see Chapter 5).

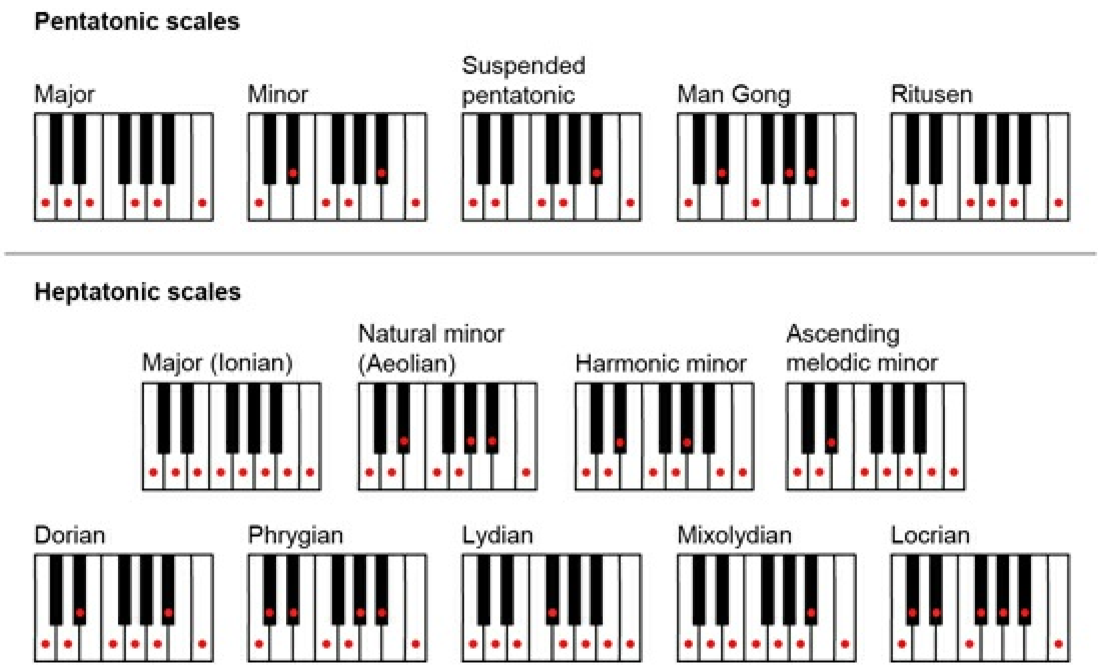


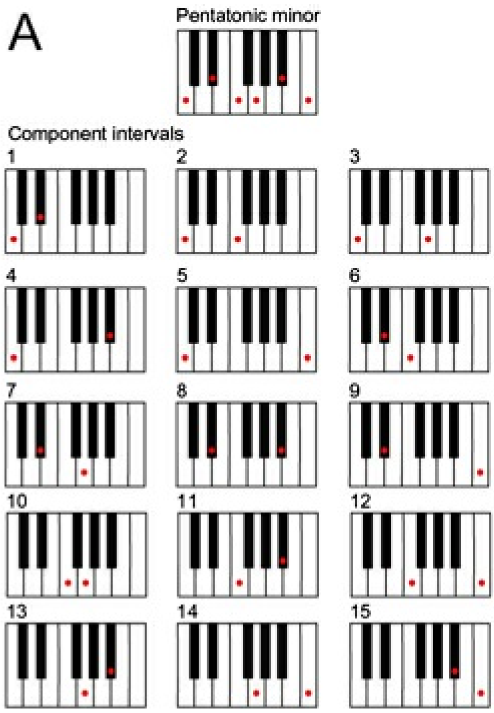
FIGURE 6.2 Pentatonic and heptatonic scales (the notes and intervals making up each scale are indicated by red dots). The differences between various scales depend the specific intervals of the chromatic scale included. Although all the scales shown begin on C and end on C an octave higher, they could begin on any note and would retain their musical identity as long as the intervals between the notes remained the same. The names of some of the pentatonic scales are from the Chinese or Indian nomenclature. The heptatonic sets with Greek names are also called "modes" based on historical usage and the affective impact of the intervals used (see Chapter 7). (From Gill and Purves, 2009. CC BY 3.0.)

The most widely accepted approach to resolving the issue of why the scales in Figure 6.2 are special is based on the integer relationships in a harmonic series. Helmholtz, and others before and since, pointed out that the frequency ratios in a harmonic series define most of the intervals of the chromatic scale in just intonation tuning.[2] However, since the amplitudes of harmonics fall off as the reciprocal of their harmonic number (see Figure 1.3), only the first few harmonics would have much influence on what we hear. Psychophysical studies have shown that people can distinguish only four or five harmonics, and then only when the testing is done under "forced choice" conditions (i.e., an answer is demanded even if the subject is only guessing). Finally, this concept of musical intervals doesn't address the question of *why* we should like small integer ratios in the first place: the ratios derive simply from the tone combinations that listeners perceive as relatively consonant.

A Biological Interpretation of Preferred Scales

Biology may again offer a more compelling explanation of preferred scales than mathematics or physics.[3] The idea introduced in Chapters 4 and 5 is that the common denominator of the two-tone combinations (dyads) preferred in music is their combined similarity to human vocalization, and this concept applies equally well to scales. One way to measure the similarity of any two-tone combination to a full harmonic series (i.e., the similarity of a dyad to this characteristic of voiced speech) is by determining the percentage of harmonics the two tones of a dyad have in common. The resulting values can then be used to give any dyad a *vocal similarity rank*, as in the example in Figure 6.3 (notice that these ranks are in line with the psychophysical ranks of dyadic consonance in Figure 5.1). By calculating the mean percentage similarity of *all* the dyads in a given scale, the overall conformance of any given scale to a harmonic series can be specified (Figure 6.3B). The rationale is based on the *summed* similarity of the intervals in the scale to a harmonic series. The greater the similarity, the greater the appeal based on the biological value of recognizing and responding to this salient characteristic of human vocalization. In short, what applies to dyadic preferences applies equally to scale preferences.

A good deal of evidence accords with this interpretation of why some scales are preferred over others. Table 6.1A shows the 10 pentatonic scales out of ~400,000 examined that have the greatest vocal similarity. The scale topping the list is the minor pentatonic scale, one of the most widely used five-note scales. The second-highest ranked is the *Ritusen* scale, a pentatonic scale used in traditional Indian music. The third and fourth ranked pentatonic scales are the ascending forms of two *ragas* used in classical Indian music (as explained in Chapter 8 *ragas* are the tonal sets used in that tradition). The fifth ranked pentatonic scale is the same as the *Ritusen* scale except that the fifth scale degree (17:10) is slightly higher in frequency compared to the 5:3 major sixth in the *Ritusen* scale. The sixth through eighth ranked five-note scales are the remaining modes of the major / minor pentatonic scales in Figure 6.2, while the ninth ranked scale is another Indian *raga*.



B

Interval	Name	Ratio	Percentage similarity
1	Minor third	6:5	33.33
2	Perfect fourth	4:3	50.00
3	Perfect fifth	3:2	66.67
4	Minor seventh	9:5	28.89
5	Octave	2:1	100.00
6	Major second	10:9	20.00
7	Major third	5:4	40.00
8	Perfect fifth	3:2	66.67
9	Major sixth	5:3	46.67
10	Major second	9:8	22.22
11	Perfect fourth	4:3	50.00
12	Perfect fifth	3:2	66.67
13	Minor third	6:5	33.33
14	Perfect fourth	4:3	50.00
15	Major second	10:9	22.22
Mean percentage similarity:			46.44

FIGURE 6.3 Measuring the vocal similarity of a scale, using the pentatonic minor scale as an example. (A) The 15 possible interval relationships among the tones in this scale in just intonation tuning. (B) The percentage similarity of each interval based on corresponding harmonics in the combined harmonic series specified by the two notes. The mean percentage similarity of the full scale can then be used to predict human preferences for different scales on the basis of their overall similarity to voiced vocalization. (From Gill and Purves, 2009. CC BY 3.0.)

Table 6.1B shows the top 10 heptatonic scales with the highest vocal similarity in sampling a large number (more than 40 million) of the approximately 10^{11} possible seven-note scales. Three of the seven heptatonic scales in Figure 6.2 top the list.[4] The Phrygian mode holds the highest rank, followed by the Dorian mode and the Ionian mode (the latter being the major scale discussed earlier; see Figure 6.1). The fourth ranked scale is similar to the Phrygian mode but contains a neutral second (12:11 in just intonation tuning) instead of a minor second; this collection is a scale used in Arabic music. The Aeolian mode (the natural minor scale) and Lydian mode are the fifth and sixth ranked scales. The next three scales are similar to the Dorian mode but with slight variations in one or two scale degrees. The seventh ranked collection is a scale in classical Indian music with an alternative sharp sixth-scale degree; the eighth ranked scale is another scale used in Arabian music. Although the ninth ranked scale does not represent any well-known musical tone collection, the Mixolydian mode is ranked tenth. The lowest ranked scale is the Locrian mode at fiftieth. The Locrian is recognized in Western music theory but rarely used. In a biological framework, its rank is low because it conforms less well to a harmonic series and thus to a primary signature of human vocal sound signals.

Table 6.1 Top ranked scales assessed according to vocal similarity.

When ranked in this way, the top 10 pentatonic (A) and heptatonic (B) scales correspond to the scales actually used in various musical traditions. None of scales used in music worldwide ranked less than 50th out of the millions examined. Although the numerical differences among ranks are small, given the large number of possible scales, they are all highly significant. (After Gill and Purves, 2009.)

Rank	Scale name	Mean percent similarity
A. Top 10 five-note scales		
1.	Minor	46.44
2.	Ritusen	46.44
3.	Candrika todi	44.28
4.	Asa-gaudi	44.09
5.	—	44.02
6.	Major	44.00
7.	Suspended	43.95
8.	Man Gong	43.85
9.	Catam	43.38
10.	—	43.33
B. Top 10 seven-note scales		
1.	Phrygian	40.39
2.	Dorian	39.99
3.	Major	39.61
4.	Husayni	39.39
5.	Natural minor	39.34
6.	Lydian	38.95
7.	—	38.83
8.	Kardaniya	38.76
9.	—	38.69
10.	Mixolydian	38.59

Although the inclusion of many of the scales used worldwide Table 6.1 is impressive, the ranks do not accord with the preeminence of the Ionian (major) and Aeolian (natural minor) in the Western classical music of the past few centuries. While there is no clear reason for this, the ranking of any scale is bound to be influenced by cultural experience, cultural differences in vocalization in particular, an issue taken up in Chapter 8. This uncertainty should not obscure the fact that out of enormous numbers of possible scales, by the metric of vocal similarity the relative handful in common use come out on top.

The Chromatic Scale

What then about the full twelve-interval, thirteen-note chromatic scale, which includes *all* the semitone intervals over an octave? When the chromatic scale is compared to a random sample of 10 million possible twelve-interval scales (the full number of possibilities is much, much greater), about 1.5 million had a higher mean percentage similarity to a harmonic series than the chromatic scale, and none of these defined a known musical palette. Thus, the chromatic scale, as a collection, bears little or no similarity to voiced speech.[5] This outcome is consistent with the observation that the *full* set of twelve chromatic tones is not emphasized in the vast majority of music. Although Arnold Schoenberg, Anton Webern, Alban Berg, and other modernists have composed interesting music that emphasizes the full chromatic scale, these "twelve-tone" or "atonal" compositions are often unappealing to listeners, especially at first. A much-cited example is Igor Stravinsky's "The Rite of Spring," in which he used a wide range of chromatic tonalities as well as unconventional rhythms and meters. The piece was widely rejected when introduced in 1913, but is now standard classical fare. In any event, the chromatic scale expresses a set of tonal possibilities rather than a scale that is used (or appreciated) as such.

In terms of vocal similarity, the implied limit of the thirteen notes and twelve intervals in the chromatic scale, and the concept of the semitone as a basic musical unit, are also unfounded. Since dividing an octave while maximizing the number of the coincident harmonics and minimizing the other harmonics in tonal dyads has no end point, there is no reason to limit musical scales to thirteen notes. In the midrange of human hearing, the just noticeable difference of pitch is ~0.3 to 0.5 percent, which in principle would allow as many as 200 notes over an octave. In fact, larger scales have been used by composers Harry Partch (a forty-three-note scale) and Karlheinz Stockhausen (an eighty-one-note scale), although to mixed reviews.

In sum, the implied limits in music theory specified by the thirteen notes in the chromatic scale, and the concept of semitone as a musical unit, have no basis in a biological framework. They are simply conventions that have been useful in teaching and notation. In biological terms, tonal relationships are determined by how well the harmonics of two tones correspond to a single harmonic series, which measures the vocal similarity of a particular dyad, or all the dyads in a scale. Rather than taking subjective consonance as a starting point, this approach *predicts* consonance based on an objective metric. In this framework, any appealing musical system orders notes over an octave according to their relative ability to indicate voiced speech in explicit (harmonic) or implicit (melodic) combinations.

Why Octaves?

The existence of octaves and the approximate musical equivalence of tones an octave apart are central features of tonal music. But the reason underlying the importance of the octave and its functions in music remains unclear. The term *octave* in any usage (e.g., electronics) is historical, and derives from the empirical fact that many musical scales comprise *eight* notes. Conventional music theory, however, does not explain why octaves are special, as they clearly are. Their existence and the fungibility of notes separated by one or more octaves are accepted as musical axioms. Of course it has long been noted that 50 percent of the harmonics in two notes separated by an octave are aligned and this fact is sometimes taken as an explanation. But other dyads such as a perfect fifth also have a relatively high degree of harmonic correspondence (33 percent) and don't play a similar role in music.

The rationale for the importance of octaves in biological terms seems straightforward. The frequency distances of octave intervals are *integer* multiples of the higher note of a dyadic pair. While this relationship would be of no consequence in itself, only integer multiplication maintains the harmonic correspondences that determine the degree of vocal similarity, and thus the relative biological value of any specific tone combination. Consider, for example, the harmonic series with a fundamental frequency of middle C (C4) on a piano keyboard. When played with itself (i.e., unison) *every* harmonic in one tone is matched to a harmonic in the other. Multiplying the fundamental of one of the tones by a factor of two creates the octave interval C4 to C5, which maintains the precise correspondence the harmonics in the two series, but with half the harmonics in the lower note lacking a corresponding harmonic in the series of higher note. The result, in addition to a higher pitch, is a difference in the timbre of the two notes, explaining why notes an octave apart are not simply identical in musical practice. Multiplying C4 by a larger integer (e.g., three or four) also maintains precise matches for each of the higher note harmonics, but with progressively more *non-corresponding* harmonics in the lower note of the dyad, introducing a greater difference in the timbre of notes two or more octaves apart. The upshot is that multiplying one note in a musical dyad by a factor of two maximizes the harmonic correspondence of the pair and thus its relative similarity to a complete harmonic series. Since the degree of vocal similarity is unchanged in any octave relationship, the attraction of *any* notes one or more octaves apart should be heard as similar.

Multiplication of the higher note of a dyad by any *non-integer* value, however, fails to achieve the same degree of vocal similarity because at least some of the harmonics in the series of the higher note no longer have a match in the lower note series, thus introducing some degree of dissonance (i.e., movement away from the harmonic uniformity that characterizes voiced speech). Indeed, this difference is what defines the eleven other frequency intervals in the chromatic scale and their different roles in music.

In biological terms, the observation that there are eight notes in a diatonic scale is irrelevant to understanding the preeminence of octaves.

Reasons for the Small Numbers of Notes in Scales

Finally, the scales used in most music worldwide emphasize six to eight tones as in the eight-note, seven-interval (diatonic) scales used in much Western classical music; the six-note, five-interval (pentatonic) scale characteristic of classical Chinese music, folk and pop music; and the seven-note, six-interval "blues scale" that is especially popular today. Why should this be? As already indicated, there is, in principle, no limit to dividing octaves into finer divisions.

In biological terms, the reason is that as more tones are added to a scale, the number of harmonic correspondences that convey vocal similarity summed over an octave (see Table 6.1) decreases progressively relative to the number of non-corresponding harmonics. Thus, as more notes are added to any scale, summed vocal similarity decreases because of the inclusion of dyads with less and less correspondence between their harmonics. Indeed, this effect is already apparent in the vocal similarity ranks of the five note versus seven note scales in Table 6.1, and explains why pentatonic scales are favored in pop and folk music. Empirically, when more than about eight notes are emphasized in a piece the overall decrease of vocal similarity causes a progressive loss of overall musical appeal.

How Musical Tone Combinations Might Have Arisen

Musical intervals are explicit or implicit *combinations* of two tones; single notes played on an instrument or sung are more or less equal in terms of their attraction and, in this sense, musically meaningless.[6] Since a speaker or singer can't produce two notes at once (with exceptions such as Tibetan "throat singers," or individuals who have learned how to whistle a fugue), isolated vocalizations, like isolated notes on a piano, do not make music. Thus, origins of music must have entailed individuals creating tones in sufficient temporal proximity to be heard as the tonal relationships in melodies, or as multiple voices creating explicit harmonies.

The implication is that music as defined in Chapter 4 would not have appeared until humans began to vocalize in fairly sophisticated ways, stringing together the voiced sounds of individuals or as combined vocalization in duets or chants.[7] Ancient humans would presumably have noticed that sequential voiced sounds or communal chants that maximized harmonic correspondence are more appealing and project better than those that do not. As vocalizing together in human cultures became more common, consonant intervals based on their degree of harmonic correspondence would have been discovered, leading eventually the subsets of full twelve-interval / thirteen-note superset used in much music today. If the evidence from ancient instruments is to be believed (see Figure 4.1B), much of this progress happened thousands of years before Pythagoras and other music theorists got into the act.

Conclusion

The observations described in this chapter point to a biological basis for the scales that humans have preferred since the dawn of history. Since we can distinguish more than 200 different pitches over an octave in the midrange of hearing, very large numbers of tone combinations could have been used to divide this frequency distance. Nonetheless, music in many traditions is based on a surprisingly small number of scales. The fact that the intervals found in the most widely used scales throughout history and across cultures are those with the greatest aggregate similarity to a harmonic series supports the idea that we humans prefer tone combinations, whether as dyads or scales, that signify conspecific vocalization. The advantage of this biological perspective is its ability to rationalize a variety of scalar phenomena—including the small number of scales used in music, the relative attractiveness of different scales, why the chromatic scale is not used as such, why octaves are special, and the limited number of notes in scales—all of which are otherwise hard to explain.

Additional Reading

Burkholder, J. P., D. Grout, and C. V. Palisca (2005). *A History of Western Music,* 9th ed. New York: Norton.

As mentioned earlier, a detailed account of the evolution of Western music that is especially pertinent here.

Burns, E. M. (1999). Intervals, scales and tuning. In: *The Psychology of Music* (D. Deutsch, ed.). New York: Academic Press, pp. 215–264.

A good general review of these issues in standard (i.e., nonbiological) terms.

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

A study of scale preferences based on biology with a more detailed account of many of the issues summarized here.

Huron, D. (1994). Interval-class content in equally-tempered pitch-class sets: Common scales exhibit optimum tonal consonance. Music Percept 11: 289–305.

A different understanding of musical scales.

Krumhansl, C.L., and R. N. Shepard (1979). Quantification of the hierarchy of tonal functions within a diatonic context. J Exp Psychol 5: 579–594.

Another rationale for scales grounded in psychology.

1. Recall that smaller divisions of these intervals are virtually universal in music around the world, either as formally notated "microtones" or as informal embellishments.

2. As discussed in Chapter 5 this tuning system and its interval ratios are based on what listeners hear as "pleasing."

3. Some might argue that auditory biology is predicated on engineering principles. The nature of Darwinian evolution by natural selection, however, defeats this idea. Evolution is not driven by principles, but by what generates successful behavior and reproduction in the niche an organism occupies.

4. The analysis is based on small integer ratios (just intonation) not equal temperament (see Chapter 4). Although the latter system alters the size of some chromatic intervals in just intonation in order to make all twelve semitones identical, the scales would have the same relative ranks.

5. This statement may seem to contradict the evidence described in Chapter 3 showing that accumulated samples of voiced speech emphasize the chromatic intervals. The two observations, however, are not comparable.

6. Tones in isolation *are* more appealing than nontonal sound signals, as might be expected, but by definition music entails relationships between two or more tones.

7. Many primates vocalize in a call and response format for social purposes. The nonhuman primate best known for duetting is the Gibbon, and impressive examples can be heard on YouTube (https://www.youtube.com/watch?v=o-c3TF6ymsM). Arguably these interactions are harbingers of the human discovery of melody and harmony in prelingual vocalization.

7

Music and Emotion

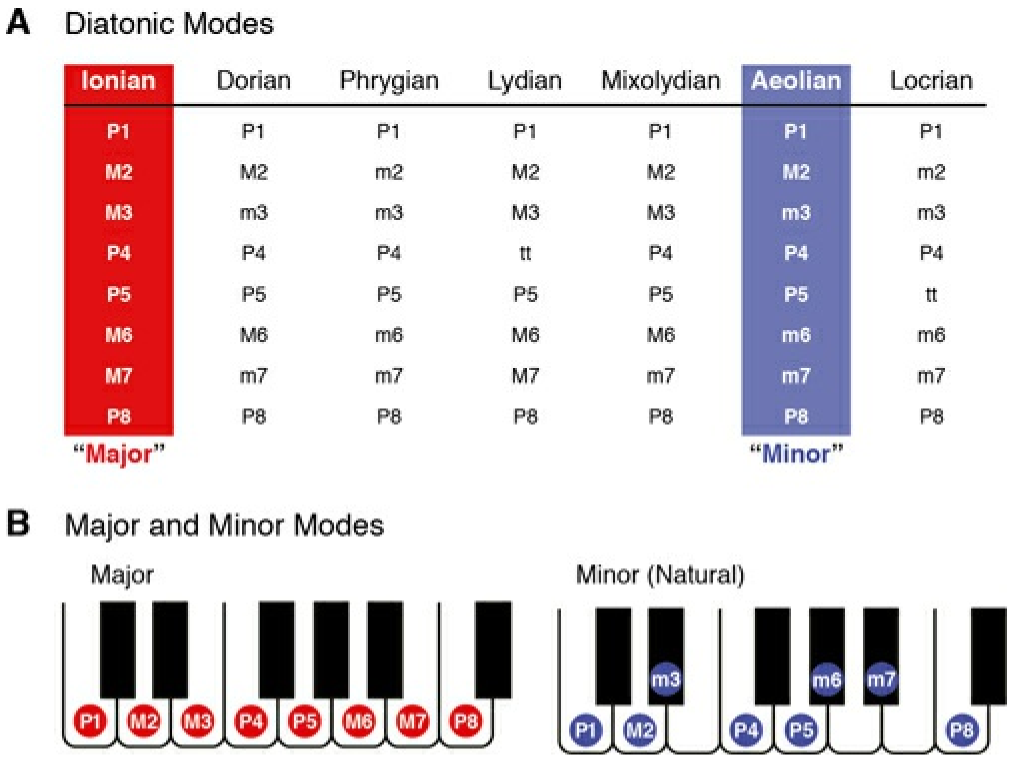
Music elicits emotional responses in listeners, often quite powerfully, an effect that is arguably the goal of music and the reason we are so strongly drawn to it. The affective impact of music depends on a variety of factors, including intensity and its variation over the course of a piece (the "dynamic") as well as tempo, rhythm, timbre, and the tonal intervals used. The way that emotion is conveyed by most of these factors seems clear enough: imitation of one or more aspects of an emotional state. If, for instance, a composer wants to imbue a composition with excitement and produce that sense in listeners, the dynamic tends to be *forte* (loud), the tempo fast and the rhythm animated. Conversely, if a subdued emotion is the aim, the dynamic is typically *piano* (soft), the tempo slow, and the rhythm staid. These effects on the listener presumably arise because the characteristics of the music accord with the ways the corresponding emotional state is expressed in human physiology and behavior, eliciting an association on this basis. The reason for the emotional effect of the tonal intervals used in music, however, is not obvious at all. Given the gist of the previous chapters, however, a plausible hypothesis is that emotional reactions to musical tones are likewise generated by imitation and association. In this case, the association in listeners would arise because the intervals that distinguish different tone collections (scales) imitate the tonal characteristics of voiced speech and nonlinguistic vocalizations uttered in one or another emotional state. This chapter examines the basis for this idea, focusing on the tonal characteristics of major and minor music and their emotional impact.

Major and Minor Scales

Musicians and listeners must have been aware long before the abstract conception of scales came into use that different tone collections tend to elicit different emotions. For example, early church music was limited to stipulated arrangements of the three intervals between the tonic note of a chord (in that era called the "final tone") and a perfect fourth (a *tetrachord*), as in the Gregorian chants developed by the Catholic Church in the ninth and tenth centuries.[1] The reason for this liturgical restriction was presumably to maintain a mood of subdued reverence that differed from popular ("folk") music that then, as now, elicits more carnal emotions pertinent to the needs, desires, and disappointments of daily life.

In the Middle Ages musical tone collections were generally called "modes" in the West (Figure 7.1). Although the difference between a "mode" and a "scale" is largely historical, the distinction persists in music theory today and can be confusing. *Mode* refers to collections of tones thought to elicit a particular feeling or state, much like the Indian *ragas* discussed in the next chapter. *Scales* is a more modern term for musical tone collections, as described in Chapter 6. The term used depends on the musical context and period under consideration. In any event, the focus in recent centuries has been on the seven-interval, eight-note collections in Figure 7.1.

As mentioned earlier, perhaps a hundred different types of scales are used in music today. In the Middle Ages, however, the focus was on formal permutations of the ascending sequence of whole tone and semitone intervals in seven-interval, eight-note collections that led to these seven scales or modes. Each was named retrospectively after regions of ancient Greece for reasons that may have been related to the music of those areas (there seems to be no clear evidence on this point). Since the Renaissance, these seven modes have been collectively referred to as *diatonic* scales, which is just a rubric for these particular tone collections.[2]



Ionian	Dorian	Phrygian	Lydian	Mixolydian	Aeolian	Locrian
P1	P1	P1	P1	P1	P1	P1
M2	M2	m2	M2	M2	M2	m2
M3	m3	m3	M3	M3	m3	m3
P4	P4	P4	tt	P4	P4	P4
P5	P5	P5	P5	P5	P5	tt
M6	M6	m6	M6	M6	m6	m6
M7	m7	m7	M7	m7	m7	m7
P8	P8	P8	P8	P8	P8	P8

FIGURE 7.1 Major and minor modes (or scales). (A) The seven diatonic scales, or "modes," highlighting the Ionian (major) and Aeolian (minor) scales. Although the Ionian and Aeolian modes have been preeminent in Western music since the sixteenth century, most of the other modes continue to be used today. For example, the Dorian mode is used in plainchant and some folk music, the Phrygian mode is used in flamenco music, and the Mixolydian is used in some modern jazz. The Lydian and Locrian modes are rarely used because the less consonant tritone takes the place of the perfect fourth between the first and fourth scale degrees in the Lydian mode, and of the perfect fifth between the first and fifth scale degrees in the Locrian mode. (B) The major and minor scales illustrated on a piano keyboard. The difference between the two is that three notes in a major scale—the third, sixth, and seventh scale degrees—are shifted down a semitone in a minor scale. (Although the semitone, as pointed out in Chapter 6, has no scientific standing, it remains a useful term in discussing Western and other musical traditions.)

Of the seven scale classes illustrated in Figure 7.1, the Ionian and the Aeolian are the most widely used today and are referred to simply as the *major* and *minor* scales. One reason for the prevalence of these two scales—perhaps the main reason—is their distinctly different affective impacts. Other factors being equal (i.e., if played with the same intensity, at the same tempo and with the same rhythm and timbre), music emphasizing the tones of a major scale tends to be perceived as relatively excited, happy, bright, or martial, whereas music emphasizing the tones of a minor scale tends to be perceived as subdued, sad, dark, or wistful. There has been no agreement, however, about how and why these scales and the different sequences of intervals that distinguish them have distinct emotional effects.

An Empirical Analysis of Major and Minor Music

The empirical differences between major and minor music—as opposed to the formal distinctions made in music theory—can be evaluated using databases of classical and folk melodies composed in major and minor keys (Figure 7.2). In both classical and folk genres, major thirds make up about 16 or 17 percent of the notes in major melodies, but less than 1 percent of the notes in minor melodies. This pattern is reversed for minor thirds, which comprise less than 1 percent of the notes in major melodies, but about 14 percent of the notes in minor melodies. The occurrence of major and minor sixths and sevenths also distinguishes major and minor music by virtue of these same biases, but less robustly.

Sorting out the necessary but somewhat intrusive terminology tends to obscure the obvious difference between major and minor scales, and the consequences for music composed in one or the other of these scale classes. Whether considered in relation the tonic or root tone in harmonies, or to the preceding note in the melody line of a piece, major music employs *larger intervals* more frequently than minor music. With respect to the melody, the salient distinction between major and minor music is the prevalence of whole tone versus semitone intervals. Major music is characterized by increased numbers of major seconds (whole-tone intervals), and minor music by an increased number of minor seconds (semitone intervals). With respect to harmonies, major music emphasizes intervals that are further from the tonic compared to minor music. The upshot is that whether in terms of harmony or melody, major music tends to emphasize larger tonal distances and minor music smaller ones.

A Major Melodies

Intervals	Classical (%)	Folk (%)
Unison/Octave	23.1	22.5
Minor Second	0.4	0.1
Major Second	11.6	15.4
Minor Third	0.9	0.0
Major Third	17.2	16.2
Perfect Fourth	8.7	8.5
Tritone	1.0	0.4
Perfect Fifth	21.6	21.3
Minor Sixth	0.7	0.0
Major Sixth	8.0	8.4
Minor Seventh	0.7	0.2
Major Seventh	6.2	6.9

B Minor Melodies

Intervals	Classical (%)	Folk (%)
Unison/Octave	22.8	20.5
Minor Second	0.6	0.1
Major Second	10.7	19.2
Minor Third	14.4	14.1
Major Third	0.7	0.2
Perfect Fourth	8.7	9.1
Tritone	1.4	0.2
Perfect Fifth	23.3	22.6
Minor Sixth	7.8	1.4
Major Sixth	1.2	2.5
Minor Seventh	3.6	7.2
Major Seventh	4.9	3.1

FIGURE 7.2 Frequency of occurrence of different intervals in major (A) and minor (B) Western classical and folk music. The biased use of chromatic intervals that distinguish major and minor music are highlighted, with red signifying major intervals and blue minor ones; the lighter colors indicate intervals that make less salient contributions. As indicated in Chapter 6, none of the 12 chromatic intervals are prohibited in musical compositions; the notes not formally included in a given scale are simply used sparingly and in special circumstances. (Data are from Bowling et al., 2010.)

Comparison of Major and Minor Music with Speech

Given the implication in Chapter 6 that the tonal intervals used in dyads or musical scales have arisen because of their similarity to human vocalization, how then, if at all, do these empirical differences between major and minor compositions reflect the differences of vocalization in particular emotional states?

In addressing this question, the simplest comparison to make between music and vocalization is of *prosody,* which in vocalization refers primarily to the up and down variations in pitch that occur in normal speech, and is one of the ways that emotion is conveyed by the tones of any utterance (intensity, tempo, rhythm, timbre, and semantic content being others). Individuals who by virtue of brain damage lose the ability to express emotion vocally tend to speak in a monotone. Such syndromes are typically due to lesions of the right hemisphere in areas that, with respect to brain anatomy, roughly correspond to the better (but incompletely) understood language areas in the left hemisphere (see Appendix). The loss of prosody is a serious deficit, as is apparent if you say "I love you!" in a monotone.

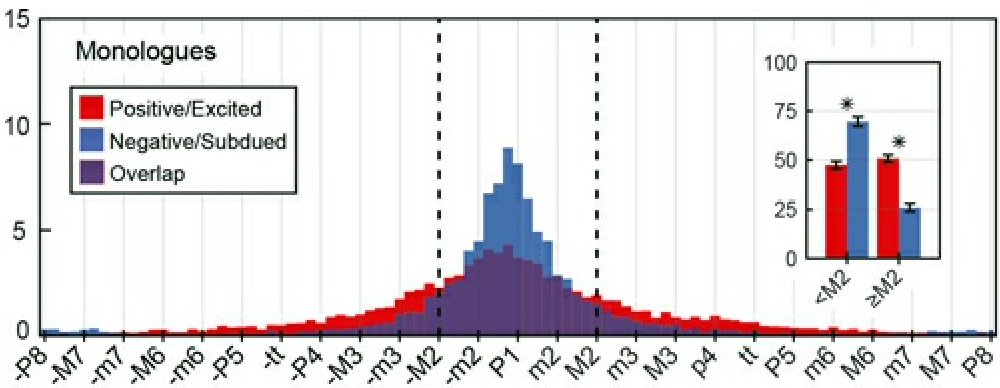


FIGURE 7.3 The distribution of tonal intervals in monologues read by American English speakers in different emotional states. When uttered in an excited manner the prosodic intervals in speech are generally larger than when the same monologues are spoken in a subdued manner. The inset shows the results in terms of intervals greater or smaller than a major second, indicated by the dashed lines in the main graph. (Data from Han et al., 2011.)

A relevant comparison of the two domains—music and speech—is thus between variations in the tonal intervals that characterize major and minor music, and variations in the prosody of speech uttered in an excited, happy state versus a subdued, sad state (Figure 7.3). Although speech does not exhibit musical intervals as such in any emotional state (i.e., there are no peaks in the distributions of frequency intervals in Figure 7.3 at the musical intervals indicated along the abscissa of the graph), excited speech entails larger intervals on average than subdued speech (and presumably in nonlinguistic vocalizations as well). Thus, speech prosody in excited and subdued emotional states and the tonal variations that characterize major and minor music tend to track each other, forming one basis for associations made by listeners.

Comparison of Music and Speech Spectra

Another way to compare musical and vocal sound signals with respect to a rationale for the emotional effects of major and minor scales is to assess the spectra of major and minor scale tones in relation to the spectra of speech uttered in an excited or a subdued state. As outlined in Chapter 1, spectra indicate the distribution of energy at different frequencies over a brief window of time, an analysis that provides a different way of measuring the characteristics of musical intervals and voiced speech that could lead to associations made by listeners.

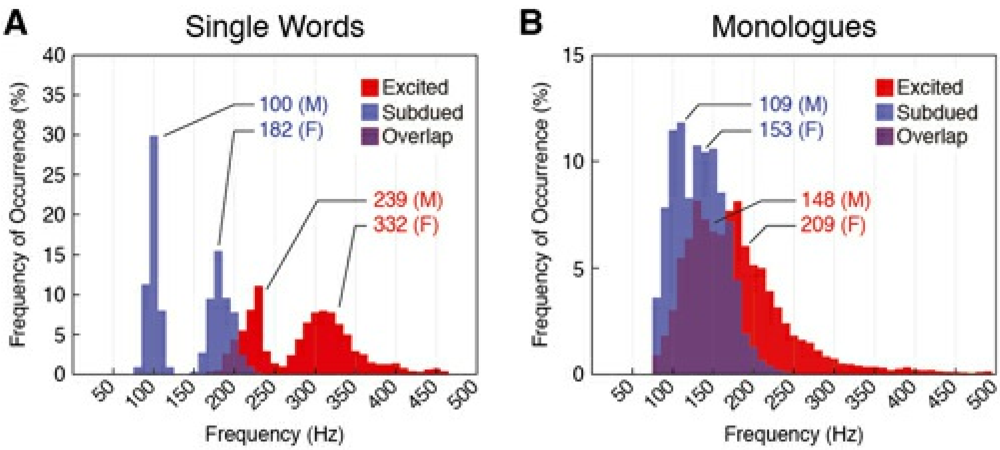


FIGURE 7.4 The fundamental frequencies of excited (red) and subdued (blue) voiced speech segments for male and female speakers. The participants were instructed to utter single words (A) or monologues (B) as if they were excited and happy (red), or conversely as if they were subdued and sad (blue). The differences between the mean fundamentals of excited and subdued voiced speech segments are indicated. (After Bowling et al., 2010.)

In music, the ratio of the fundamental frequency in the spectrum of the higher note to the fundamental of the lower note defines each interval played, while the relative prevalence of tones in a piece distinguishes whether the composition is major or minor (see Figure 7.2). In voiced speech, the fundamental frequencies of vocal pitch are equally important, conveying the emotional state of a speaker, as well as information about the speaker's age, size, gender, and identity (see earlier).

Figure 7.4 shows the distributions of the fundamental frequencies derived from the spectra of voiced speech segments for individual speakers uttering speech in an excited versus a subdued manner. The comparison shows, as one would expect, that the fundamental frequencies in excited speech are a good deal higher than those of subdued speech.

Assessing the relevant fundamentals in music is more complicated in that musical intervals are defined by the ratio of the fundamental frequencies of *two* tones (i.e., a higher tone with respect to the lower one, which in a scale is the low note or reference tone). Thus, there is no single fundamental as there is for a voiced speech sound. As described in Chapter 6, a way around this difference is to consider the combined harmonics of the notes as a single harmonic series, the fundamental of the combination being the greatest common divisor. As shown in Figure 7.5, the differences between the fundamental frequencies in major and minor music parallel the differences in fundamental frequencies evident in excited and subdued speech (cf. Figure 7.4).

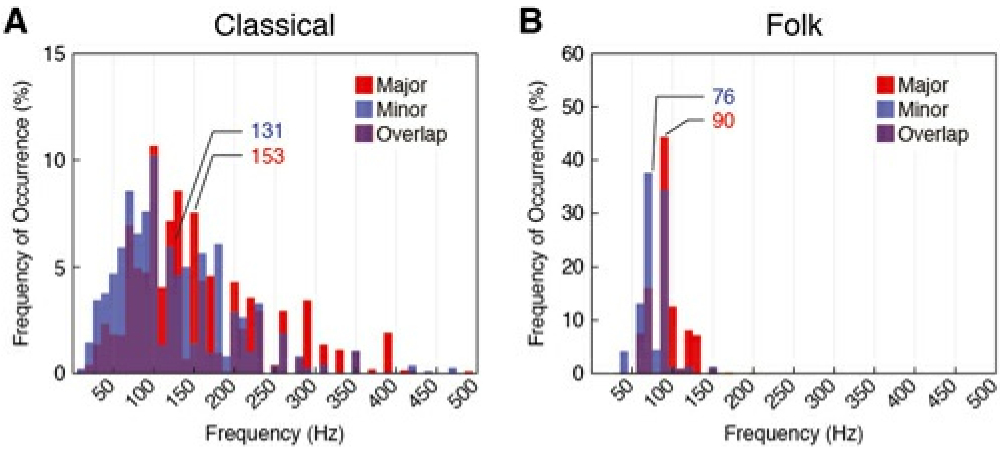


FIGURE 7.5 The fundamental frequencies of two-tone combinations in Western classical music (A) and folk music (B). The mean fundamental frequency values (i.e., the common denominator of the two frequencies) for each distribution are indicated. The average fundamentals in major music of both genres are greater than those in minor music. The relative narrowness of the data in (B) is a consequence of less variation in the key signatures of folk music compared to classical music. (After Bowling et al., 2010.)

The main reason for the tonal distinctions between excited and subdued speech is the fundamental frequency of the speaker's voice. When a speaker is excited, increased muscular tension acting on the vocal folds, the vocal tract, and the chest cavity raises the fundamental frequency of the signal and its prosodic variation; conversely, when a speaker is subdued, decreased tension lowers the fundamental frequency and decreases prosodic variation (see Figure 7.3). In music, the tones in minor music have less variation in pitch height because of the prevalence of smaller tonic and melodic intervals compared to the intervals characteristic of major compositions.

Comparison of Major and Minor Music and Speech Formant Ratios

Another type of analysis uses the frequency ratios that distinguish major and minor music compared with the ratios of speech formants. As described in Chapter 3, formants are amplitude peaks in the harmonic series that characterize the spectrum of any voiced speech signal, and are generated by the resonances of the vocal tract above the larynx (see Figure 3.2). The first two formants (called F1 and F2) represent the most salient resonances of the vocal tract and are necessary (and sufficient) for the production and discrimination of vowel sounds in any language. F2 / F1 ratios thus allow particular vowel sounds to be recognized as such across speakers with anatomically different vocal tracts (e.g., men, women and children).

In excited speech, more formant ratios correspond to those of major seconds, thirds, sixths, and sevenths, whereas ratios corresponding to minor seconds, thirds, sixths, and sevenths are absent (Figure 7.6). In contrast, in subdued speech fewer formant ratios correspond to major seconds, thirds, sixths, and sevenths, whereas more ratios correspond to those of minor seconds, thirds, sixths, and sevenths.

These parallel differences between the frequency of occurrence of formant ratios in excited and subdued speech and the ratios of the musical intervals that distinguish major and minor scales add further to the similarity between speech in different emotional states and the intervals that distinguish major and minor music, providing yet another basis for associations between emotions expressed in music and emotions expressed in speech.

Emotional Effects of Other Diatonic Scales

Why, then, is there little evidence about affective qualities conveyed by the five other diatonic modes in Figure 7.1?

One problem that militates against pursuing this issue is distinguishing other emotions in terms that would allow a comparison of their expression in speech and music. Figure 7.7 shows a standard way of parsing human emotional states. Notice, however, that a number of different emotional states are subsumed under the descriptors “excited” and “subdued” used in the studies described earlier in the chapter. Thus, in addition to “happy” and “sad,” the classifier “excited” includes emotional states such as “alarmed,” “afraid,” “angry,” and “astonished.” Similarly, the classifier “subdued” would include the emotional states called “bored,” “tired,” and “gloomy.” Although the terms *excited* and *subdued* apply well enough when assessing the expression of these states in music, attempts to more specifically align other emotional states with the remaining modes in Figure 7.1 or other scales would be a difficult task, even if one were to focus on genres such as opera where emotions are highlighted.

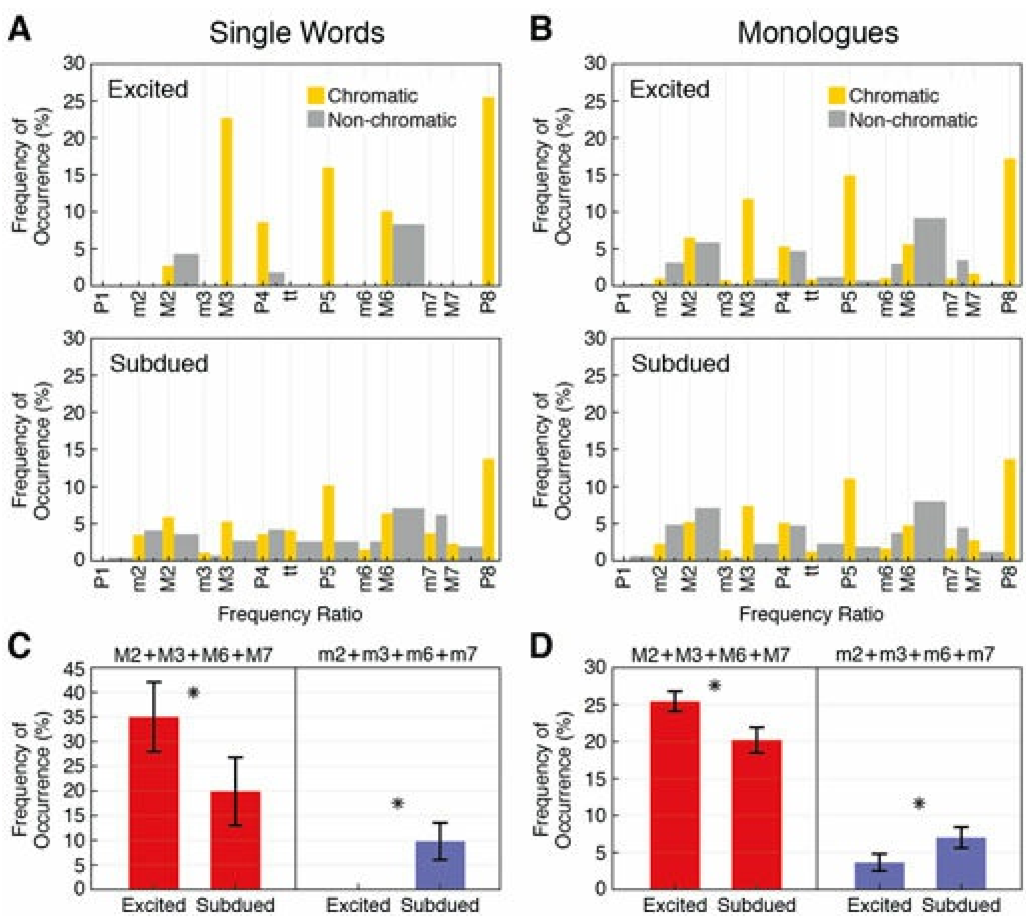


FIGURE 7.6 Comparison of speech formant ratios in different emotional states with the ratios that define musical intervals. (A, B) The distribution of formant ratios in excited and subdued speech in recordings of single words and monologues. Yellow bars indicate ratios that fall within 1 percent of chromatic interval ratios (see Figure 4.2). Gray bars indicate ratios that do not correspond to chromatic intervals. (C, D) Percentages of chromatic intervals in (A) and (B) that correspond to the intervals that distinguish major and minor music. Asterisks indicate significant differences. See text for explanation. (After Bowling et al., 2010.)

FIGURE 7.7 A diagram of human emotional states. Diagrams like this—called “circumplex” models of emotion—are based on studies in which participants rate the words designating an emotion on a graph

where the horizontal axis specifies degree of arousal (i.e., activity versus torpor) while the vertical axis specifies the degree of pleasantness or unpleasantness (the "valence") of an emotional state. (From Purves et al., 2013.)

Another problem is the subtlety of the emotional impact conveyed by the remaining modes in Figure 7.1, or any other scales, and the difficulty describing exactly what they are. Whereas major and minor scales have been tied to excited and subdued emotional states for at least several centuries (and no doubt much longer in practice), there is little agreement about other associations between an emphasis on different tonal intervals in music and their emotional consequences.

These caveats, however, should not obscure the evidence here that tones play the same sort of role in generating emotions by imitation and association as do the dynamic, tempo, rhythm and timbre of a musical piece.

Conclusion

For aspects of music that do not entail tonality such as intensity, tempo, rhythm and timbre, the emotional quality of a composition is conveyed by imitating the way a given emotion is expressed in human behavior, motor behavior in particular. The same imitative principle and the associations it elicits in listeners can likewise explain the affective impact of the different tone collections. Thus, the acoustical characteristics of speech in an excited state more closely reflect the intervals that define music in a major scale, whereas the characteristics of subdued speech more closely reflect the intervals emphasized in minor music. The resulting associations between speech uttered in different emotional states and the different tone collections used in music appear to provide a biological basis for the emotions elicited.

Additional Reading

Blood, A. J., and R. J. Zatorre (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proc Natl Acad Sci USA 98: 11818–11823.

A paper that relates musical emotion with brain activity.

Bowling, D. L, K. Gill, J. D. Choi, J. Prinz, and D. Purves (2010). Major and minor music compared to excited and subdued speech. J Acoust Soc Am 127: 491–503.

A comparison of speech and music with respect to major and minor scales, providing a more detailed account of some of the work summarized in this chapter.

Darwin, C. (1872 / 2000). *The Expression of Emotion in Man and Animals.* Oxford, UK: Oxford University Press.

A classic on the origins of emotion that is still well worth reading.

Juslin, P. N., and P. Laukka (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? Psychol Bull 129: 770–814.

A meta-analysis of virtually all the studies that linked music, speech and emotion up to about 2003.

Koelsch, S. (2014). Brain correlates of music-evoked emotions. Nat Rev Neurosci 15: 170–180.

A good review of what is known of brain activity in response to music.

Spencer, H. (1857). The origin and function of music. Fraser Mag 56: 396–408.

Spencer's historically important speculations about the emotions elicited by music and their impact on human physiology.

1. Tetrachords were often combined to span an octave. The term comes from early instruments designed to play tone combinations, such as the ancient Greek lyre.

2. Diatonic in Greek simply means "across tones."

8

Music and Speech across Cultures

As already emphasized, the use of preferred subsets of chromatic scale intervals is characteristic of many musical traditions. However, as anyone who has traveled widely—or simply dined in a Chinese restaurant—will have noticed that the way the same or similar tone collections are used differs markedly across cultures. Most attempts to account for these differences between Western, Eastern, and other traditions refer to scale preferences, but this line of thinking raises the question of why particular sets of intervals should be preferred in the first place. Given the links between music and vocalization documented in earlier chapters, a possible explanation is that cultural differences in the way the same tonal sets are used arise because the tonal characteristics of a culture's speech influence the tonal characteristics of its traditional music. This chapter summarizes evidence that the characteristics of a culture's music are indeed biased by the character of its language.

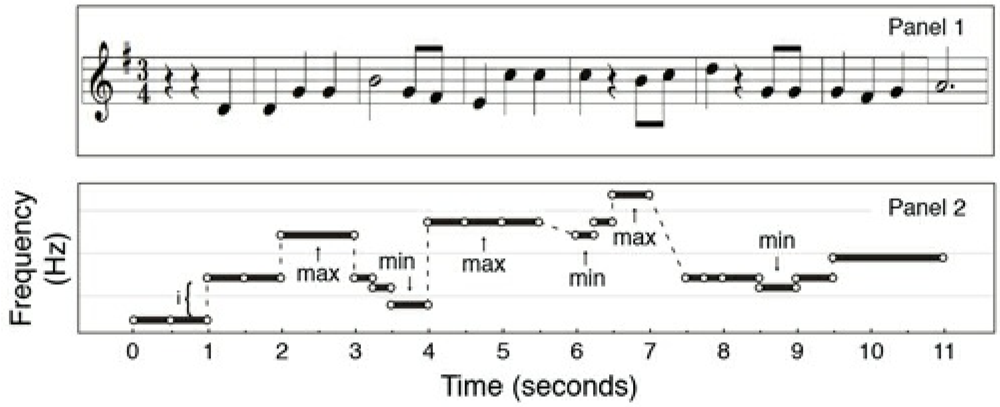
Tonal Differences among Languages

The use of tonality in speech varies greatly among languages. The most obvious example is the division of languages into *tonal* and *non-tonal* types. In *tone languages*, pitch contours and levels convey the lexical meaning of each syllable. For example, Standard Mandarin (the lingua franca of China) uses five tones, referred to as "high," "rising," "falling then rising," "falling," and "light tone," the latter being considered a "neutral" tone. The syllable "ma" uttered as a high tone means "mother." The same syllable uttered with a rising tone means "hemp," with a falling then rising tone "horse," and with a falling tone "scold." Other tone languages, such as Vietnamese and Thai are similar, but differ in the number of tones used to convey the lexical meaning of a syllable. Many other languages use tonal variations in a less comprehensive way (e.g., Japanese), or for different purposes (e.g., to indicate past and future, as in some African languages). In contrast, in *non-tone languages* (e.g., English, French, Spanish, German, Tamil) tone levels and / or syllabic contours do not affect lexical meaning, although syllabic stress is sometimes used to differentiate words with the same spelling, for example the English words CONtent and conTENT.

The use of tonality in speech and language is thus a continuum, with languages like Standard Mandarin and American English defining extremes in which tonality is used to convey the meaning of every syllable (Mandarin) versus languages in which variations convey only broader, generally non-lexical information (English). These differences in the uses of tonality in language presumably explain the fact that a higher percentage of Mandarin speakers (estimates vary widely but are on the order of 30 percent) have absolute or "perfect" pitch (defined as the ability to identify a musical tone without comparison to an explicit reference tone). In contrast, in non-tone language-speaking cultures this ability is considered relatively rare, and is restricted to highly trained musicians. The implication is that whether or not a person has absolute pitch is largely a matter of tonal practice, whether as music or language, particularly at an early age when neuronal connections are far more malleable than they are in adulthood.[1]

Music and Speech in Tone Language and Non-tone Language Cultures

In considering these observations, a pertinent question is whether differences in the traditional music of tone and non-tone language cultures are related to the different uses of tones in speech. To explore this issue, traditional melodies can be compiled for tone language and non-tone language cultures, and the tonal variations in speech compared to tonal variations in music.



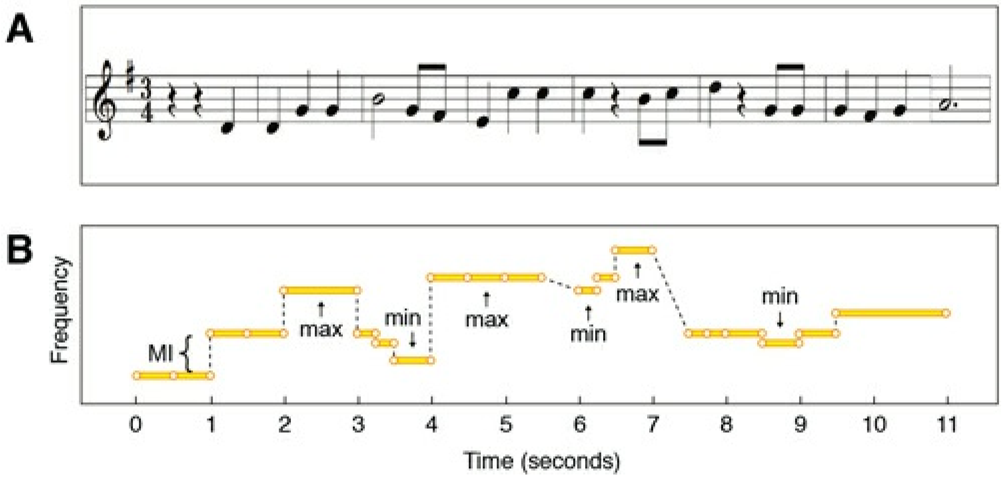


FIGURE 8.1 Analysis of tonal sequences in music. (A) The musical notation of a few measures of a traditional American melody ("Home on the Range"). (B) The same melody reformatted for analysis of slope reversals and melodic interval sizes. Yellow bars represent pitch heights in the melody, with open circles marking the beginning and ending of a new level. Local maxima (max) and minima (min) indicate levels where the slope of the melodic contour reverses; melodic interval size (MI) is the vertical distance between successive pitch heights in the melody, indicated by the dashed lines. (After Han et al., 2011.)

Figure 8.1 shows how tonal differences in music can be analyzed using simple parameters such as the frequency of melodic "slope" reversals and the size of melodic intervals. Slope reversals refer to changes in the direction of the contour of a melody as it unfolds, and interval "size" refers to the difference in pitch height between successive notes in a melody.

Insofar as possible, speech is analyzed in the same way. For instance, Figure 8.2A shows examples of emotionally neutral monologues recorded from readings by native speakers of six different languages. Figure 8.2B indicates how the analogous features of speech—slope reversals in the prosodic contour and the size of the pitch height differences between syllables—are measured.

Figure 8.3 compares the number of melodic slope reversals in the music of tone language and non-tone language cultures with the number of prosodic slope reversals in speech. The median number of melodic slope reversals per one hundred notes is greater in the music of tone language cultures compared with non-tone language-speaking cultures. Similarly, the median number of prosodic slope reversals per one hundred syllables is greater in the speech of tone language cultures than in non-tone language cultures.

Figure 8.4A compares the size of melodic intervals in the music of tone and nontone language cultures with the size of prosodic intervals in speech. Melodic intervals in all cultures are relatively small (0 to 500 cents), presumably because intervals much greater than a perfect fourth (500 cents) are harder to sing or play.[2] This overall tendency notwithstanding, the distribution of melodic intervals in tone and nontone language cultures differs. The music of tone language cultures uses fewer melodic intervals that are less than or equal to a semitone (100 cents; 16 percent versus 37 percent) compared with non-tone language cultures, and more larger intervals (200 to 500 cents; 84 percent versus 64 percent). The only inconsistency in the comparison is the major third (400 cents), which is used somewhat more in the music of the non-tone language cultures (in traditional Western music, for example).

Finally, the distribution of prosodic interval sizes in speech is shown in Figure 8.4B. As in music, there is a lower incidence of smaller pitch changes (<200 cents) and a higher incidence of larger changes (>200 cents) between adjacent syllables in tone-language speech compared with non-tone language speech.

A

English Monologue

“I applied for a job that would give me a good work experience. I was placed in a department that I like and my boss seems like a reasonable guy. My co-workers are also nice, although I don’t really get together with them very often outside the office. Overall, I am very satisfied with my new position”

French Monologue

“Je viens de’être engage poste qui va me donner une bonne experience professionelle. Je suis dans un département que j’aime et mon chef semble être une personne raisonnable. Mes collègues sont aussi très sympas bien que je ne sorte pas souvent avec eux. Globalement je suis très satisfait de mon nouveau travail.”

German Monologue

“Ich habe mich für einen Job beworben, in dem ich viel Erfahrung sammeln kann. Die Abteilung, in der ich eingesetzt werde, gefällt mir gut, und mein Chef scheint auch vernünftig zu sein. Meine Kollegen sind auch sehr nett, aber außerhalb der Arbeit treffe ich sie nicht sehr häufig. Insgesamt bin ich mit meiner neuen Stelle sehr häufig. Insgesamt bin ich mit meiner neuen Stelle sehr zufrieden.”

Mandarin Monologue

“一直以来，我期望第一份工作能够是一个好的开始。很幸运，我找到了一份自己喜欢的工作，而且还有一个还算不错的老板。同事们人都很好。不过下班后我很少参加他们的活动。总之，我对这份工作很满意。”

Vietnamese Monologue

“Tôi kiếm được một việc mà tôi nghĩ sẽ học hỏi được nhiều kinh nghiệm. Tôi được xếp vào một phòng tôi thích và sếp tôi có vẻ là một ông sếp tốt. Đồng nghiệp của tôi cũng rất dễ thương, mặc dù tôi không đi chơi với họ sau giờ làm. Nhìn chung, tôi rất hài lòng với công việc mới của mình.”

Thai Monologue

“ตอมนี้ฉันได้งานใหม่แล้ว ฉันว่าจะไปได้ดี ทั้งเป็นงานที่ฉันชอบ เจ้านายก็ท่าทางเป็นคนมีเหตุผล และเพื่อนร่วมงานก็น่ารัก แต่ว่ฉัน ไม่ค่อย ได้ออก ไปไหนกับพวกเขามากเท่า ไหร่ โดยรวมๆ ตอนนี้ฉันพอใจกับชีวิตฉันมากเลยทีเดียว”

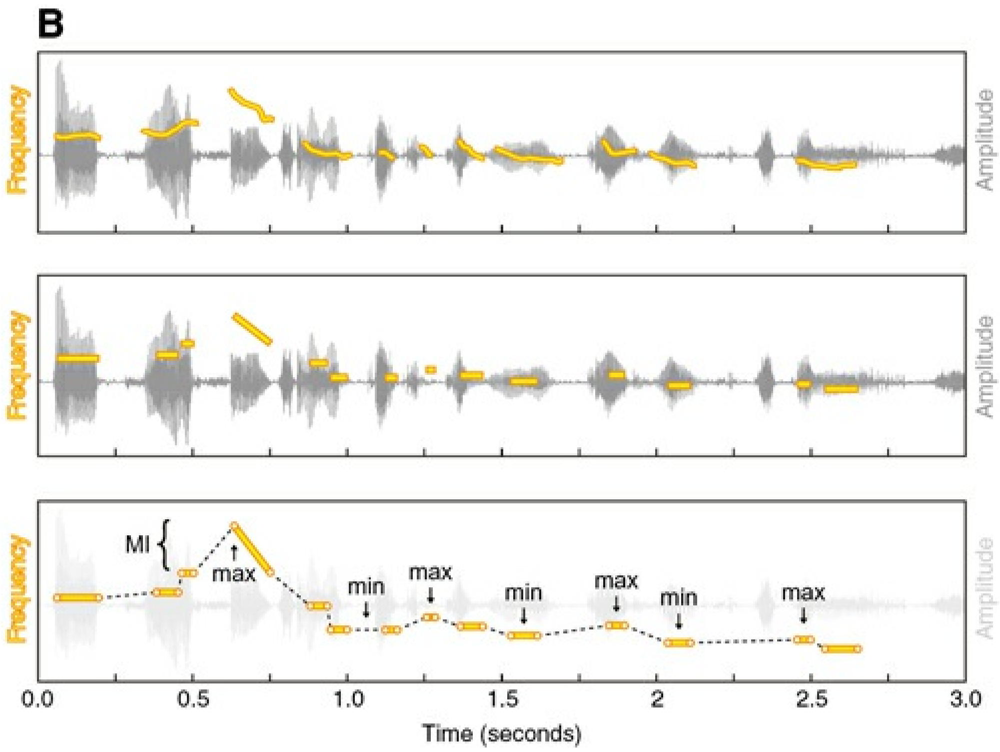


FIGURE 8.2 Analysis of tonal sequences (prosody) in speech. (A) Examples of monologues read by native speakers of the languages examined. (B) The sentence "I applied for a job that would give me a good work experience" spoken in American English. The upper panel shows the sound signal generated by the spoken sentence, with the fundamental frequencies of the voiced speech segments shown in yellow. The middle panel shows the fundamentals of the voiced portions of the syllables simplified as bars, and the lower panel shows the data with open circles marking the beginnings and endings of the bars. As in Figure 8.1, local maxima (max) and minima (min) indicate slope reversals in the prosodic contour; the analog in speech of melodic interval size (MI) is indicated by the vertical distance between the pitch heights of successive tonal components in the utterance. (After Han et al., 2011.)

In sum, the tonal patterns in music and speech again track each other, in this case across the different cultures examined. The music of tone language cultures changes pitch direction more frequently and uses larger melodic intervals on average compared to the music of non-tone language cultures. Similarly, the frequencies of occurrence of changes in pitch direction and the size of the pitch intervals between subsequent syllables are greater in tone language compared to non-tone language speech. The implication is that the tonal character of a culture's speech influences the way tones are used in its music, and perhaps vice versa.

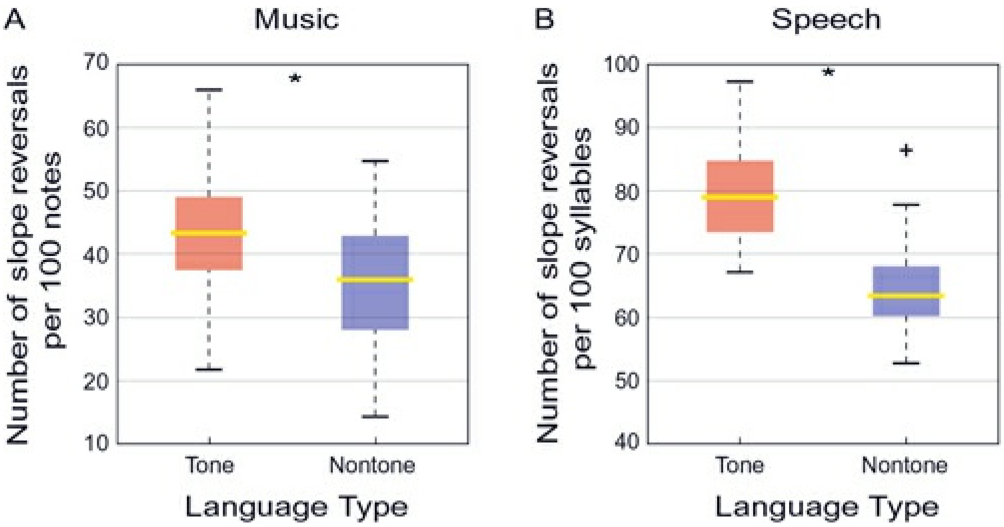


FIGURE 8.3 Slope reversals in the music and speech of a tone and non-tone language speaking culture. (A) The number of melodic slope reversals per 100 notes in the traditional music of the tone (red) and non-tone (blue) language culture. Yellow lines indicate the median number of reversals; boxes indicate the interquartile range and dashed lines the overall range. (B) The number of prosodic slope reversals per 100 syllables in tone and non-tone language speech illustrated in the same way. (From Han et al., 2011. CC BY 3.0.)

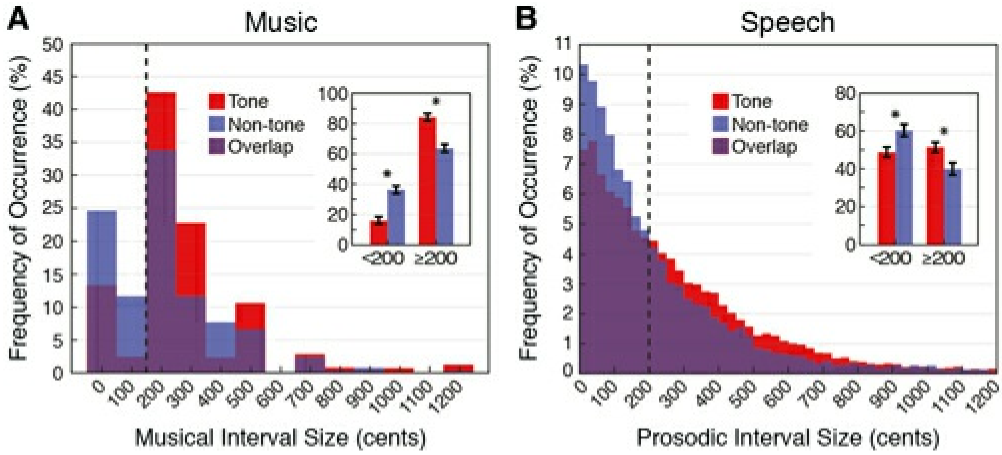


FIGURE 8.4 Comparison of interval sizes in the music and speech of tone and non-tone language speaking cultures. (A) The distribution of melodic interval size per melody in the music of tone (red) and non-tone (blue) language cultures (the overlap is shown in purple); the inset summarizes these data in terms of intervals greater or smaller than 200 cents (a major second; dashed line in the main graph). (B) The distribution of prosodic interval size in the speech of tone (red) and non-tone (blue) languages. The inset again shows the percentage of prosodic interval sizes greater or smaller than 200 cents. Asterisks indicate that the differences are statistically significant. (From Han et al., 2011. CC BY 3.0.)

The Expression of Emotion in Eastern and Western Music

Another issue is whether the relationship between the way emotion is expressed in speech and in music holds for non-Western cultures. As explained in Chapter 7, in Western music the major scale (the Ionian mode) is typically used to convey excited, happy, bright, or martial feelings while the minor scale (the Aeolian mode) is used to convey subdued, sad, or darker emotional states. The preeminent empirical difference between these tone collections is that larger intervals tend to be used in major music and smaller ones in minor music. What then about the expression of emotion in classical Eastern music? Is the same connection between music and speech tonalities apparent?

An especially useful corpus in this regard is traditional South Indian Carnatic music, in which the emotions conveyed by traditional *ragas* are specified (the Hindustani music of Northern India is much the same).[3] *Ragas,* as briefly mentioned in Chapter 7, are collections of tones analogous to scales or modes in Western music, but more elaborate (Table 8.1). In addition to specifying a group of tonal relationships that tend to convey a different emotional state, *ragas* also specify patterns of tonal movement and particular ways that notes are to be played. Thus, while using what is essentially the chromatic scale, some additional variants are specified and used in different contexts. Other ornaments called *gamakas* indicate notes that are to be "shaken," or where the musician is meant to "slide" to the next note. These elaborations are similar to the stylized ornamentation in seventeenth and eighteenth century Baroque music and other classical genres in which vibrato, glissandos, trills, and other figures adorn melodies in both sung and played music. Less formal ornamentation is abundant today in popular Western rock, R&B, and blues music, in which melismas, pitch bends, and glissandos are typical. These less formal nuances in musical pitch also seem to imitate human vocalization in different emotional states, although this possibility has not been studied.

Table 8.1 Tonal intervals in classical Western music compared to classical Indian music. (After Han et al., 2011.)

A. The 13 notes and 12 principal intervals used in South Indian Carnatic music over an octave are indicated by the names in the first column (here and in Table 8.2, a dot above or below the abbreviated name indicates the octave above or below the primary octave).
B. The 13 notes and 12 intervals of the Western chromatic scale for comparison. The names, abbreviations, fundamental frequency ratios in just intonation (JI), and interval sizes in both JI and equal temperament (ET) are indicated. The 12 principal intervals in the two systems are similar.

			Interval size (cents)	
Interval name (s)	Abbreviation	Frequency ratio	JI	12-TET
A. Twelve-interval, thirteen-note system used in Carnatic music				
Sa	S	1 : 1	0	0
Shuddha Ri	R1	16 : 15	112	100
Chathusruthi Ri, Shuddha Ga	R2,G1	9 : 8	204	200
Shatsruthi Ri, Sadharana Ga	R3,G2	6 : 5	316	300
Anthara Ga	G3	5 : 4	386	400
Shuddha Ma	M1	4 : 3	498	500
Prati Ma	M2	45 : 32	590	600
Pa	Pa	3 : 2	702	700
Shuddha Dha	D1	8 : 5	814	800

Chathusruthi Dha, Shuddha Ni	D2,N1	27 : 16	906	900
Shatsruthi Dha, Kaisiki Ni	D3,N2	9 : 5	1,018	1,000
Kakali Ni	N3	15 : 8	1,088	1,100
Sa	Ṡ	2 : 1	1,200	1,200
B. Chromatic scale used in Western music				
Perfect unison	P1	1 : 1	0	0
Minor second	m2	16 : 15	112	100
Major second	M2	9 : 8	204	200
Minor third	m3	6 : 5	316	300
Major third	M3	5 : 4	386	400
Perfect fourth	P4	4 : 3	498	500
Tritone	tt	7 : 5	583	600
Perfect fifth	P5	3 : 2	702	700
Minor sixth	m6	8 : 5	814	800
Major sixth	M6	5 : 3	884	900
Minor seventh	m7	9 : 5	1,018	1,000
Major seventh	M7	15 : 8	1,088	1,100
Perfect octave	P8	2 : 1	1,200	1,200

Table 8.2 Comparison of Eastern *ragas* and Western major and minor modes.
A. Carnatic *ragas* commonly associated with positive/excited and negative/subdued emotion. The positive/excited *ragas* are taken to fall under the emotional themes (*rasas*) of happiness, courage, pride and confidence. The negative/subdued *ragas* are mostly associated with the *rasas* of disgust, depression and self-pity.
B. The Western modes commonly associated with positive/excited and negative/subdued emotion from Chapter 7 for comparison. Abbreviations are those defined in Table 8.1.

Emotional theme	Name of *raga* (Indian) or mode (Western)	Interval names (abbr.) (↑ = ascending) (↓ = descending)	Number of melodies examined
A. Carnatic Indian *ragas*			
positive/excited	*Bilahari*	↑ S, R2, G3, P, D2, Ṡ ↓ Ṡ, N3, D2, P, M1, G3, R2, S	51
	Mohanam	↑ S, R2, G3, P, D2, Ṡ ↓ Ṡ, D2, P, G3, R2, S	42
negative/subdued	*Naadanaamakriya*	↑ Ṇ3, S, R1, G3, M1, P, D1, N3 ↓ N3, D1, P, M1, G3, R1, S, Ṇ3	5
	Punnaagavaraali	↑ Ṇ2, S, R1, G2, M1, P, D1, N2 ↓ N2, D1, P, M1, G2, R1, S, Ṇ2	30
	Varaali	↑ S, G1, R1, G1, M2, P, D1, N3, Ṡ ↓ Ṡ, N3, D1, P, M2, G1, R1, S	57
B. Western modes			
positive/excited	Major	↑ P1, M2, M3, P4, P5, M6, M7, P8 ↓ P8, M7, MS, P5, P4, M3, M2, P1	566
negative/subdued	Minor	↑ P1, M2, m3, P4, P5, m6, m7, P8 ↓ P8, m7, m6, P5, P4, M3, M2, P1	376

As with major and minor Western modes, particular *ragas* are used to convey different emotions. In Hindu tradition, nine emotional themes (called *rasas*) are described, and different *ragas* fall into one or another of these categories. Thus, the *rasa* of joy, happiness, and mirth (called *Hasya*), and the *rasa* of sadness, grief, and pity (called *Karuna*), together with other *rasas* that express positive and excited or negative and subdued emotional states, are similar to the emotional states that are conveyed by major and minor Western musical modes, respectively (Table 8.2).

In comparison with the negative and subdued *raga*s, positive and excited *raga*s emphasize major tonic intervals, with approximately 20 percent more major sixths, 15 percent more major thirds, 11 percent more major seconds, and 4 percent more unison or octave intervals (Table 8.3A). With respect to melodic intervals, the main difference is in melodic major seconds: positive and excited *raga*s exhibit 31 percent more major seconds than negative and subdued *ragas*.

Table 8.3 An empirical comparison of the interval differences used to convey emotion in Eastern and Western music. (From Bowling et al., 2012. CC BY 3.0.)
A. Intervals used in Carnatic *ragas* that express positive/excited emotions (red) compared to negative/subdued emotion (blue). Western interval abbreviations are used for convenience; asterisks indicate statistically significant differences.

A. Interval prevalence in Carnatic Indian melodies						
Bin center (cents)	Approx. interval	positive/excited *raga* melodies	negative/subdued *raga* melodies	Degrees of freedom	t-statistic	*P*-values
Tonic intervals positive/excited *raga* > negative/subdued *raga*						
900*	M6	**19.9%**	0%	192	39.66	2×10^{-94}
400*	M3	**16.8%**	0.9%	192	27.36	3.5×10^{-68}
200*	M2	**17.5%**	7.3%	192	11.87	0
0 / 1,200*	P1 / P8	**19%**	14.9%	192	4.75	3.9×10^{-6}
700	P5	**17.8%**	16.7%	192	1.34	0.18
positive/excited *raga* < negative/subdued *raga*						
100*	m2	0%	**15.9%**	192	−23.56	1.5×10^{-58}
800*	m6	0%	**12.3%**	192	−19.39	4.1×10^{-47}
300*	m3	0%	**8%**	192	−7.3	6.7×10^{-12}
600*	tt	0%	**6.6%**	192	−9.64	0
500*	P4	3.7%	**7.4%**	192	−3.58	4.4×10^{-4}
1,100*	M7	4.5%	**7.3%**	192	−3.29	1.2×10^{-3}
1,000*	m7	0.8%	**2.6%**	192	−4.23	3.6×10^{-5}
Melodic intervals positive/excited *raga* > negative/subdued *raga*						

200*	M2	55%	23.2%	192	12.47	0
300*	m3	**19.3%**	87.%	192	8.37	1.2×10^{-14}
500*	P4	**3.5%**	1.5%	192	6.31	1.8×10^{-9}
700*	P5	**0.7%**	0.4%	192	2.15	0.032
0	P1	**11.5%**	11.1%	192	0.38	0.71
900	M6	**0.1%**	0%	192	1.78	0.076
> 1,200	> P8	**0%**	0%	192	1.33	0.18
		positive/excited *raga* < negative/subdued *raga*				
100*	m2	8.3%	**48.7%**	192	−17.87	1.1×10^{-42}
400*	M3	1.3%	**5.5%**	192	−9.03	2.2×10^{-16}
600*	tt	0%	**0.4%**	192	−3.68	3×10^{-4}
800*	m6	0%	**0.2%**	192	−1.9	0.058
1,000	m7	0.1%	**0.1%**	192	−0.08	0.94
1,100	M7	0%	**0.1%**	192	−0.72	0.41
1,200	P8	0%	**0.1%**	192	−0.82	0.49

B. Western interval distributions for comparison. Recall that "tonic interval" refers to an interval in the context of harmony where the metric is based on the reference note of the dyads in a scale or the root tone of more complex chords. "Melodic interval" refers to the frequency distance between successive notes in a melody line.

B. Interval prevalence in classical Western melodies

Bin center (cents)	Approx. interval	Major melodies	Minor melodies	Degrees of freedom	t-statistic	*P*-values
			Tonic intervals Major > Minor			
400*	M3	**18.1%**	0.8%	202	32.42	0
900*	M6	**8.4%**	1.4%	202	16.12	0
1,100*	M7	**7.7%**	5.3%	202	5.06	5.1×10^{-7}
0 / 1,200	P1 / P8	**20.2%**	19.7%	202	0.68	0.5
			Minor < Major			
300*	m3	0.9%	**15.6%**	202	−36.98	0
800*	m6	0.6%	**7.8%**	202	−20.97	0
1,000*	m7	0.7%	**3.4%**	202	−9.56	0
700	P5	19.2%	**20.6%**	202	−2.01	0.05
600	tt	1.2%	**1.6%**	202	−1.53	0.13
500	P4	10.1%	**10.4%**	202	−0.46	0.65
200	M2	12.6%	**12.8%**	202	−0.43	0.67
100	m2	0.4%	**0.6%**	202	−1.29	0.2
			Melodic intervals Major > Minor			
200*	M2	33.8%	26.9%	940	5.94	4.0×10^{-9}

400*	M3	**7.4%**	5.5%	940	3.66	2.7×10^{-4}
0	P1	**11.1%**	10.7%	940	0.4	0.69
500	P4	**8%**	7.6%	940	0.69	0.49
900	M6	**1.4%**	1.2%	940	1.06	0.29
> 1,200	> P8	**1%**	0.9%	940	0.64	0.52
1,200	P8	**1.3%**	1.2%	940	0.37	0.71
1,000	m7	**0.4%**	0.4%	940	0.62	0.53
			Minor < Major			
100*	m2	21%	**28.2%**	940	−7.12	2.2×10^{-12}
300*	m3	9.8%	**11.1%**	940	−2.01	0.044
800*	m6	1.3%	**2%**	940	−2.76	5.9×10^{-3}
600*	tt	0.5%	**1.1%**	940	−3.01	2.7×10^{-3}
700	P5	3%	**3.3%**	940	−0.92	0.36
1,100	M7	0.1%	**0.1%**	940	−0.12	0.9

In contrast, negative and subdued *raga* melodies are characterized by an overall increase in minor intervals compared with positive and excited *raga* melodies. Negative and subdued melodies on average comprise 16 percent more minor seconds, 13 percent more minor sixths, 7 percent more minor thirds, and 4 percent more tonic minor sevenths. Other intervals that are more prevalent in negative and subdued *raga*s are the tritone (6 percent more) and the perfect fourth (5 percent more). In terms of melodic intervals, negative and subdued *ragas* exhibit 40 percent more melodic minor seconds than positive and excited *ragas*.

The upshot is that whether in terms of tonic or melodic intervals, positive and excited *ragas* use larger intervals, whereas negative and subdued *ragas* use smaller ones. This pattern of interval sizes used to express emotion in Carnatic music is thus much the same as the pattern apparent in Western major and minor modes (see Table 8.3).

Expression of Emotion in Eastern and Western Speech

The question that follows is whether these differences in music used to convey different emotions also track differences in emotional speech across cultures, or at least this particular comparison of Eastern and Western cultures.

Figure 8.5 shows the distributions of prosodic interval sizes in positive and excited compared with negative and subdued speech in Tamil (the non-tone language spoken in South India, where Carnatic music arose) and American English. As might be expected from the evidence described in Chapter 7, prosodic intervals in positive and excited speech are larger on average than prosodic intervals in negative and subdued speech in both languages.

In short, the tonal characteristics of speech in different emotional states and the characteristic intervals used to convey these feelings in music are much the same in this example of Eastern and Western cultures.

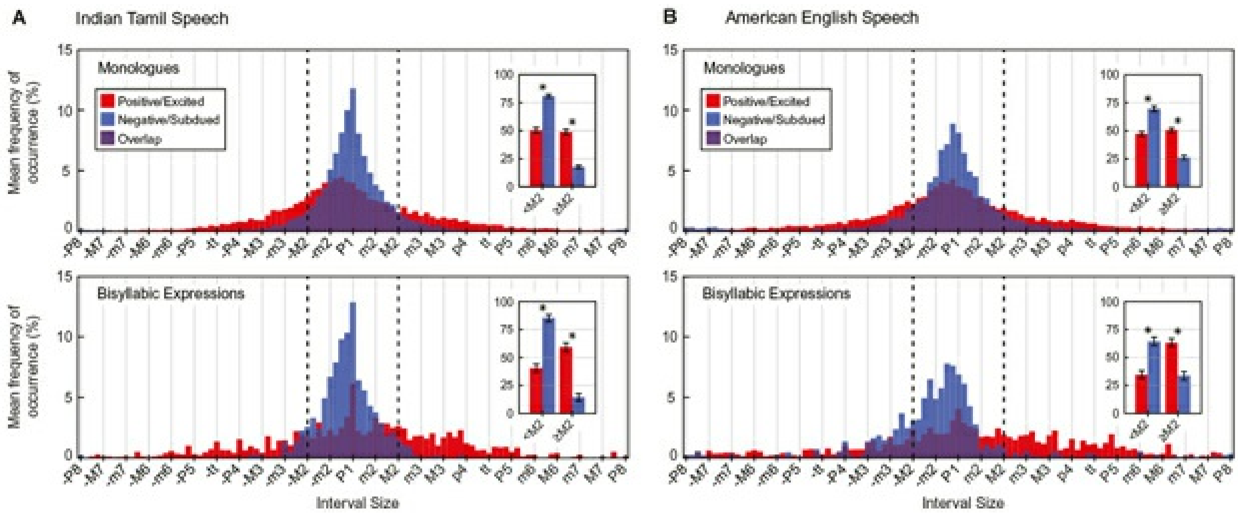


FIGURE 8.5 Prosodic intervals in Tamil (A) and American English (B), based on analyses of monologues (above) and two-syllable utterances (below). As before, the distribution of positive/excited speech is shown in red and of negative/subdued speech in blue; purple shows overlap. The dashed lines indicate the boundary between intervals smaller than a major second (between the dashed lines) and intervals that are larger. The upper panel in (B) is the same data shown in Figure 7.3. It is worth emphasizing again that musical intervals as such are not apparent in the prosodic variations of speech in either language.

Conclusion

The co-variation of tonal characteristics in music and speech across cultures accords with the intimate relationship between these two forms of social communication. With respect to the possible reasons for different of use of notes within the same scales, the characteristics of music and speech in tone-language and non-tone-language culture track each other pretty well. Further supporting the connection between the tones used in speech and those in music, the use of larger intervals, which are perceived as emotionally positive and excited, versus smaller intervals, which tend to be perceived as emotionally negative or subdued, is also much the same in the two nontone language cultures examined.

Additional Reading

Bowling, D. L., J. Sundararajan, S. Han, and D. Purves (2012). Expression of emotion in Eastern and Western music mirrors vocalization. PLOS ONE 7(3): e31942. doi:10.1371 / journal.pone.0031942

A more detailed account of much of the data presented in this chapter.

Han, S., J. Sundararajan, D. L. Bowling, J. Lake, and D. Purves (2011). Co-variation of tonality in the music and speech of different cultures. PLOS ONE 6(5): e20160. doi:10.1371 / journal.pone.0020160.

A full account of the tone-language culture versus non-language study.

Ho, L. T., and K. H. Han (1982). On Chinese scales and national modes. Asian Music 14: 132–154.

How scales are used in traditional Chinese music.

Patel, A. D. (2008). *Music, Language, and the Brain.* New York: Oxford University Press.

An extensive review of the literature on some of the issues discussed in this chapter.

Satyanarayana, N. C. (2005). *The Science of Indian Music.* Hyderabad, India: Sri Dattasai Graphics.

A full account of Carnatic music.

Vos, P. G., and J. Troost (1989). Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. Music Percept 6: 383–396.

A classic paper on melodic intervals.

1. P. Loui, A. Zamm, and G. Schlaug (2012) provides evidence that variations in the ability to identify absolute pitch is a function practice, with references to other literature in this somewhat controversial field.

2. See Figure 4.2 for a review of interval measurements in cents.

3. The language of South India is Tamil, a non-tone language similar in this respect to English and other Western languages.

9

Implications

The goal of neuroscience—whether in audition or any other domain—is to understand the operation of animal nervous systems, the human brain in particular. And in most respects, this effort has been a great success. The cellular and molecular basis of neural signaling is well understood; the connectivity of the brains of many species is known in detail; the functional properties of individual neurons and circuits in numerous nervous systems have been thoroughly described; and observing neural activity in the brains of humans carrying out various tasks is now routine. In one respect, however, understanding the brain has been deeply elusive: despite a wealth of information and technology, the link between the physical world and the odd way we perceive it is not understood. By the same token, almost nothing is known about biological basis of why we like some sensory inputs (stimuli) more than others (i.e., aesthetic preferences). This last chapter focuses on what tonal music as biology implies about these and other broader issues.

Implications for Musical Phenomenology and Aesthetics

A theme in previous chapters is that a biological framework for understanding tonal music can rationalize a range of musical phenomenology that conventional music theory cannot. Most accounts of musical tonality can be traced back to the mathematical ratios first promoted by Pythagoras, who according to legend found that subjectively pleasing tone combinations are generated by plucking strings whose lengths or tensions are related by small integer ratios. These ratios define musical unison, octaves, fifths, and fourths. Other frequency ratios were formally included in Western theory (and other traditions) over the centuries, leading eventually to the thirteen-note / twelve-interval chromatic scale that defines the set from which much music is drawn today.

Beginning with the scientific insights in the Renaissance and culminating with the work of Helmholtz in the nineteenth century, subjectively pleasing (consonant) tone combinations were recognized as ratios of the fundamental frequencies of vibrating strings or air columns, putting musical tonality and instrument tuning on a physical footing. Music on the basis of subjective preferences for particular ratios based on math and physics does not, however, explain any of the puzzles in music taken up in previous chapters. Moreover, this way of looking at music flirts with a tautology: the musical tone combinations we like are used to explain why we like musical tone combinations.

The biological perspective laid out here offers a way to resolve these deficiencies by taking tonal music to be, at bottom, a consequence of our inherent attraction to tones, driven by the evolutionary advantages of recognizing and responding to conspecific vocalizations.[1] Given this framework, understanding the phenomenology of tonal music becomes a lot easier. The salient phenomena that make more sense when looked at in this way are consonance and dissonance, musical scales, the special role of octaves, the small number of scales humans use, the small number of tones in scales, why scales have "home base" reference tone, musical tension and resolution, why musical tones convey emotion, and why tonal palettes vary across cultures.

A biologically based understanding of tonal preferences also provides insight into the evolution of a sense of tonality in the first place, its extension over the eons to tonal music, the limited musicality of other species, and even human aesthetics. Whereas data about preferences in other art forms are hard to generate and discuss in scientific terms, music provides an abundance of carefully documented information that indicates what we humans have been and are attracted to. These data in turn allow exploration of why these attractions exist and what science might have to say about aesthetics, which have generally been relegated to the humanities.

Chapter 5 on consonance and dissonance, Chapter 6 on musical scale preferences and their characteristics, Chapter 7 on the emotions inspired by different scales, and Chapter 8 on the similarities and differences in tonal preferences across cultures all point in the same direction. We like tonal stimuli that help us detect and interpret human and other animal vocalizations, and dislike (or better said, like less) tone combinations that indicate less clearly a voiced sound source and its significance. The basis of tonal preferences in the art form we call music is, in this conceptual framework, the degree to which the sound signals accord with the tonal characteristics of human vocalization, the harmonic series of voiced phones in particular.

Implications for Auditory Physiology and Anatomy

The biological importance of recognizing conspecific sound signals based on vocal similarity accords with the physiology and anatomy that is being pursued in auditory systems of nonhuman primates and other experimental animals. For example, many neurons in the auditory cortex of monkeys are driven not only by fundamental frequencies of a tone, but by integer multiples of that frequency (i.e., by its harmonics). Moreover, when tested with two-tone combinations (dyads), many neurons show stronger excitation or inhibition when the tones are related as integer multiples (i.e. by octave intervals). These findings led physiologist Xiaoqin Wang working at Johns Hopkins to propose that responding to harmonics is an organizing principle of the primate auditory cortex.[2] Adding to this evidence, other researchers have described a region in the anterior temporal lobe of macaque monkeys that responds specifically to the variety of calls that these animals use to communicate socially.[3]

Readers will have noticed, however, that little has been said in the book about regions of the human brain that support responses to music, despite the fact that many studies have addressed this issue over the past couple of decades using functional magnetic resonance imaging, electroencephalography, transcranial magnetic stimulation, as well as clinical evidence from neurological patients. While these studies are certainly of interest, the results have shown that many poorly understood brain regions that also serve other neural functions are activated by music.[4]

Thus, interpreting the role of the various cortical and subcortical structures activated by listening to or performing music (or any other stimuli for that matter) remains a difficult challenge. This work has been well summarized in a number of reviews; those by Robert Zatorre and colleagues at McGill University listed at the end of the chapter are especially recommended.

Implications for Musicality in Other Species

Social communication using vocal sound signals has been amply documented in many species, including species of amphibians, birds, whales, dolphins, other mammals, and non-human primates. Why then is tonal music absent or at least very limited in other animals?

If the human sense of tonal consonance derives from the biological importance of the information conveyed by conspecific vocalization, other animals that generate sound signals that entail harmonics series—including the species just mentioned—should also be attracted to a uniform harmonic series for the same reasons we are: an indication of animacy, conspecific animacy and its biological significance in particular. In principle then, lots of species have the wherewithal to make tonal music.

What seems to be missing is the social learning that drives human culture, which has evolved in humans to a far greater degree than in other animals. The term *culture* of course carries a good deal of baggage. Although its definition is debated, learning from parents and peers (social learning) lies at the core of arguments about the similarities and differences between human culture and the simpler versions of culture in other species.[5] Whereas human culture depends critically on what we are taught by parents, teachers and peers, social learning in other species, although it clearly exists, is obviously limited.

Absent the vastly increased social learning that has led to speech and language in human culture and all that has followed therefrom, tonal expression as we know it in music should be limited in other animals, as it appears to be. Although many animals can discriminate pitches, monkeys can recognize transposed melodies by octave similarity,[6] and some birds produce a few musical intervals,[7] that seems to be about the extent of non-human musicality.

Implications for the Operation of Sensory Systems in General

Since the mid-twentieth century, the focus of sensory neuroscience has been on the properties of neurons in the relevant input systems and the neural circuits they form in experimental animals. The seemingly sensible assumptions underlying this approach are that perception arises from neural mechanisms that encode stimulus features, filter out redundant or otherwise less important information, and combine what remains to represent a "sparse" version of external reality mediated by the population activity of functionally specific neurons in the brain's sensory cortices. It seems self-evident that the goal of neural processing in any sensory modality, audition included, should be to accurately reveal the physical properties of the local environment so that successful behavior can follow.

As pointed out in Chapter 1, however, the problem with this assumption is that neither the real world nor its significance for behavior is specified by the energy that impinges on biological sensors. Figure 9.1 illustrates this problem in vision in much the way Figure 1.4 illustrated the problem in audition. The reason for bringing up vision here is that this confounding issue has been far more extensively studied in the visual system, where it is easier to understand both the problem and its apparent resolution. Moreover, the similar quandary in another modality implies that evolution has had to resolve the same challenge in all sensory systems. The quandary is that stimuli derived from electromagnetic radiation, local atmospheric pressure changes, or any other form of biologically usable energy in the environment could have arisen from many combinations of physical sources. This fact precludes mapping stimulus features back onto physical reality. The information needed to do this is simply not available to biological sensing systems.[8] Information about physical sources, however, is what animals need to succeed in the world.

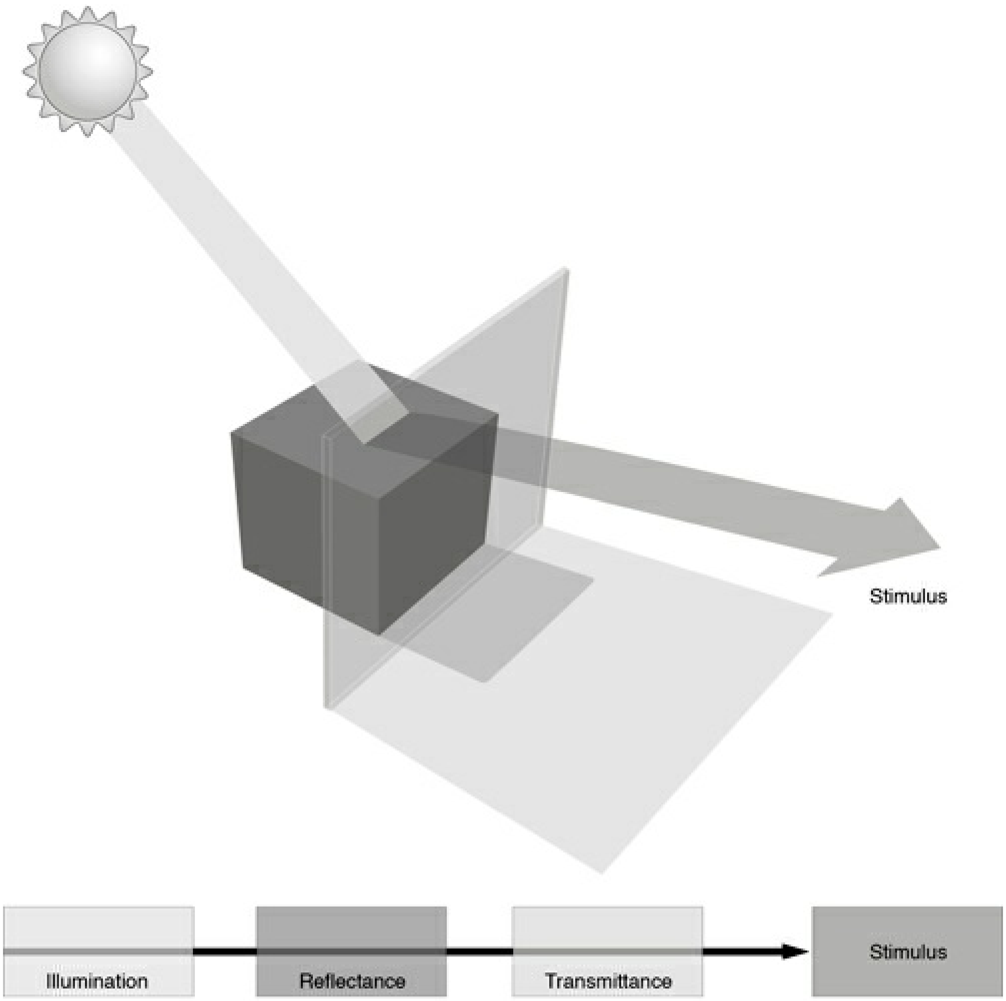
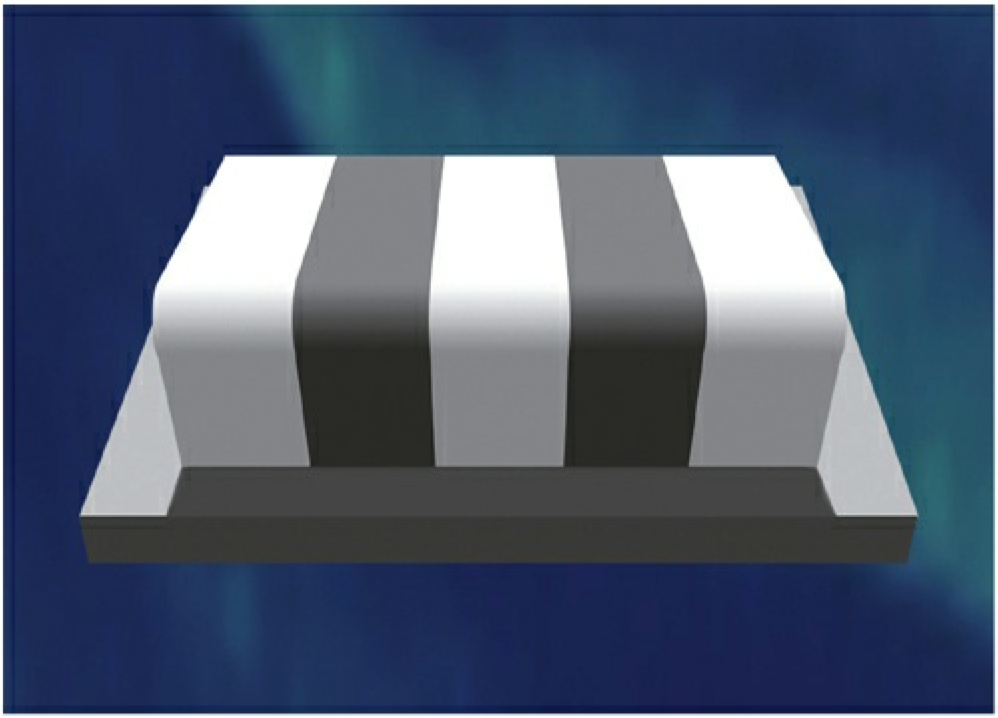


FIGURE 9.1 An example of why the physical sources of visual stimuli cannot be apprehended in vision. Luminance values in retinal stimuli are determined by illumination and reflectance, as well as a host of other physical factors (e.g., atmospheric transmittance, spectral content, occlusion, distance, object orientation, and many more). These physical parameters, however, are conflated in light stimuli, precluding biological measurements of the objective world in which perceptions and other behaviors must play out. (After Purves and Lotto, 2011; cf. Figure 1.4.)

Since what we see, hear, and otherwise perceive in response to the energy affecting peripheral sensors works quite well as a guide to behavior, the idea that objects and conditions in the environment cannot provide information about the physical world may be difficult to credit. But despite our deeply held belief that the world we experience via the senses represents reality, it does not. The universal discrepancies between the objective world and our subjective impression of it are not hard to appreciate, especially in vision, where they have been much better documented than in audition.[9] As an standard example, Figure 9.2 demonstrates the strange relationship between luminance (a physical measure of light intensity) and lightness (the perceived lightness or darkness of object surfaces). What we see is clearly not what physical instruments like photometers measure.

These puzzling observations in vision and audition imply that we contend with the inherently uncertain meaning of stimuli by generating perceptions that have been determined empirically by trial and error. Perceptual responses will have been rewarded or not according to their ultimate contribution to behavior and reproductive success. As diagrammed in Figure 9.3, the result of evolutionary and lifetime learning is that perceptions end up tracking the frequency of occurrence sensory input patterns, not physical reality. In simplistic terms,[10] perceptual and other behavioral responses to stimuli would have been made randomly at first. Over time, however, useful responses would have been selected for, while those that contributed less to the success of the species would have been weeded out. By gradually ranking the qualities seen or heard over a subjective range that had proven useful in responding to recurring stimulus patterns, we and other animals would have succeeded in the world despite being unable to measure reality (see Figure 9.1).[11]



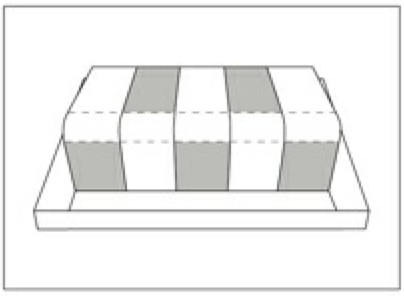


FIGURE 9.2 Differences between objective measurements of light and perception. Although the target patches in the inset have the same luminance and appear the same in a "neutral" setting, there is a striking perceptual difference in lightness when the patches are viewed in the scene with context. This phenomenon and others like it are not "illusions" but the inevitable outcome of the way humans evolved to perceive any and all patterns of luminance in retinal images, which cannot be mapped back to their physical sources. (After Purves and Lotto, 2011; cf. Figure 9.1.)

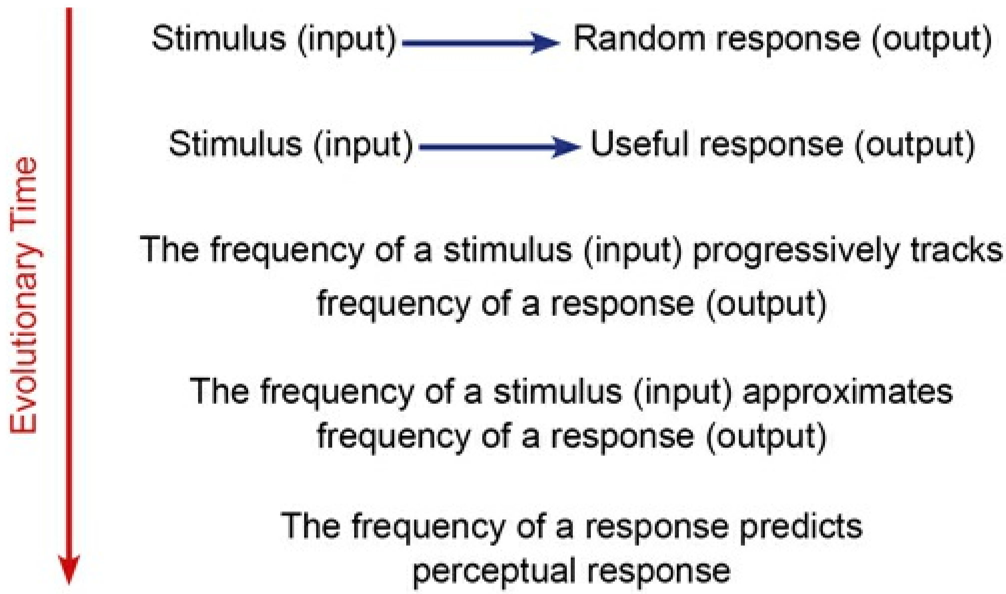


FIGURE 9.3 How and why the frequency of occurrence of stimulus patterns predicts perceptions. Early in evolution, different behaviorally significant input patterns are associated with responses made more or less randomly by trial and error. As evolution progresses, however, sensory inputs are gradually linked to perceptual (and other behavioral) outputs based on their contribution to survival and reproductive success. As a result, the frequency of occurrence of stimulus patterns and the evolving agent's responses to them eventually accord with (and can be used to predict) their perception (e.g., the lightness values we see in Figure 9.2, or the auditory qualities we hear as loudness, pitch and timbre). (After Purves et al., 2014.)

Although input–output equivalence in any trial-and-error process is never reached, after sufficient evolution and lifetime learning the function that describes the input should approximate the function that describes the output, which includes perception. Once evolution of a sensory system has brought the input–output functions into approximate alignment, the frequency of occurrence of stimulus input patterns should predict what we see, hear, or otherwise experience subjectively. And since the metrics of reproductive success differ fundamentally from metrics of the physical world, the observed discrepancies between objective measures and subjective percepts are expected.

This strategy implies that pre-neural apparatus in the sensory periphery and all related sensory system circuitry have evolved to serve the same overriding purpose: to *rank-order* the subjective qualities we perceive (and responses to stimuli generally) according to their promotion of survival and reproductive success. Although linking objective and subjective domains in this way does not bring perceived values closer to the relevant physical parameters (see Figure 9.2), the strategy endows us and other agents with the ability to perceive and act in ways that lead to biological success without having to access physical reality as such.

Implied Mechanisms

Generally speaking, the biological mechanisms underlying this empirical way of understanding sensory systems and perception are well known. The driving force that instantiates the links between sensory stimuli and reproductive success during the evolution of a species is natural selection: random changes in the structure and function of the sensory systems in ancestral forms have persisted or not in descendants according to how well they serve the reproductive success of the agent who harbors the variants. Any configuration of pre-neural apparatus and neural circuitry that mediated more successful responses to stimuli will increase among members of a population, whereas less useful circuit configurations and operations will not. The significance of this conventional statement about the phylogeny of any biological system is simply the existence of a well-established mechanism for instantiating and updating the empirical associations that underlie perceptions and other behavioral responses.

Neural circuitry is of course modified over the lifetimes of individuals according to particular experience by the mechanisms of neural plasticity, taking advantage of information that allows individuals to benefit from their circumstances in innumerable ways. What we learn over the course of a lifetime is not, however, heritable.[12] Content, including the music we experience, is passed on to future generations by means of culture. Thus it is evolutionary experience that does the heavy lifting in this way of explaining how sensory systems circumvent the inverse problem illustrated in Figures 1.4 and 9.1.

Implication That Audition and Other Sensory Systems Are Reflexive

A further implication of this strategy is that perceptions and other behaviors are *reflexive*. Although the concept of "reflex" has always been imprecise, it alludes to behaviors such as the "knee-jerk" response that depend on the automatic transfer of information from sensory input to motor (or other) output via circuitry established by feedback from behavioral success. The advantages of reflex responses are clear enough: once natural selection and the mechanisms of neural plasticity have done their work over evolutionary and individual time, the nervous system can respond to sensory or other inputs with greater speed and accuracy.

It does not follow, however, that reflex responses must be "simple," that they are limited to motor acts, or that they entail only "lower order" neural circuitry. Charles Sherrington, who pioneered the study of reflex circuits in the early twentieth century, was well aware that the concept of a "simple" reflex is, in his words, a "convenient … fiction." As he pointed out, no part of the nervous system acts without affecting other parts, however remote. Moreover as alluded to earlier, cortical neurons are not limited to performing a single task. It follows that an enormous number of interacting neural influences must be taken into account for the optimal execution of even the basic reflex.

There is no evidence that responses to sensory input based on feedback from behavioral outcomes, no matter how complex, differ in any important way from a spinal reflex or the reflexive regulation of homeostatic processes by the autonomic components of mammalian nervous systems. Although the concatenation of neural circuitry and brain regions involved may be far more extensive in cognitive functions like perception, understanding them as responses based on connections previously established by empirical success makes good biological sense.

Conclusion

Musical tonality understood in terms of our attraction to sound signals that are similar to human vocalization (i.e., that conform to a harmonic series) can rationalize many issues that conventional music theory cannot. The most important of these issues are consonance and dissonance, musical scales and their organization, the number of scales used, the special role of the octave, the number of tones in scales, why a reference tone acts as a musical "home base," musical tension and resolution, how musical tones convey emotion, and the tonal palettes of different cultures. Like other sensory qualities that humans find attractive, music is evidently based on a subjective preference for auditory stimuli that have promoted reproductive success, in this instance by recognizing the biologically critical information conveyed by conspecific vocalizations. Although this reinvention of tonal music theory as biology is unlikely to have gotten everything right, it seems a step toward the goal of understanding music in a more rational way.

Additional Reading

Belin, P., R. J. Zatorre, P. Lafaille, P. Ahad, and B. Pike (2000). Voice-selective areas in human auditory cortex. Nature 403: 309–312.

Evidence that a region of the human brain is dedicated to vocal recognition.

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. Cognition

100: 173–215.

A review of musical ability in a variety of other species.

Hoeschele, M., H. Merchant, Y. Kikuchi, Y. Hattori, and C. ten Cate (2015). Searching for the origins of musicality across species. Phil Trans R Soc B 370: 20140094.

A different perspective on the origins of music.

McDermott, J., and M. Hauser (2005). The origins of music: innateness, uniqueness and evolution. Music Percept 23: 29–59.

Another good review of music across species.

Nettl, B. (1956). *Music in Primitive Culture.* Cambridge, MA: Harvard University Press.

Ideas about origins of music based on the music of primitive societies.

Petkov, C. I., C. Kayser, T. Steudel, K. Whittingstall, M. Augath, and N. K. Logothetis (2007). A voice region in the monkey brain. Nat Neurosci 11: 367–374.

Evidence that a nonhuman primate devotes a region of the brain to conspecific vocal processing.

Purves, D., B. Monson, J. Sundarajan, and W. T. Wojtach (2014). How biological vision succeeds in the physical world. Proc Natl Acad Sci USA 111: 4750–4755.

A more detailed overview of the relationship between the objective and subjective domains in another sensory modality.

Sherrington, S. C. (1906/1947). *The Integrative Action of the Nervous System,* 2nd ed. New Haven, CT: Yale University Press.

A classic description of reflexes and the organization of the nervous system in these terms.

Zatorre, R. J., and J. M. Zarate (2012). Cortical processing of music. In: *The Human Auditory Cortex* (D. Poeppel et al., eds.). New York: Springer.

Zatorre, R. J., J. L. Chen, and V. B. Penhune (2007). When the brain plays music. Nat Rev Neurosci 8: 547–558. doi :10.1038 / nrn2152.

Both these papers review the brain regions activated by music.

1. There are of course many ethnomusicologists and some psychologists who argue that harmony and melody derive from cultural experience rather than the demands of biology. A recent example is J. H. McDermott, A. F. Schultz, E. A. Undurraga, and R. A. Godoy (2016).

2. X. Wang (2013).

3. C. Perrodin, C. Kayser, N. K. Logothetis, and C. I. Petkov (2011).

4. The reason for "multiplexing" is presumably that if the same brain regions and neurons have several different functions, brains of fixed size can generate far more neural processing and behavioral output than otherwise.

5. A set of excellent essays on this issue can be found in K. N. Laland and B.G. Galef (2009).

6. A. A. Wright, J. J. Rivera, S. H. Hulse, M. Shyan, and J. J. Neiworth (2000).

7. E. L. Doolittle, B. Gingras, D.M. Endres, and W. T. Fitch (2014).

8. Of course physical instruments can carry of this mapping using non-biological mechanisms.

9. D. Purves and R. B. Lotto (2011).

10. “Simplistic” because these changes over evolutionary time would have happened concurrently, not sequentially. Moreover, the term *percepts* suggests a level of awareness that must have arisen relatively late in animal evolution with the advent of complex brains.

11. A fuller account of these difficult ideas can be found in D. Purves, Y. Morgenstern, and W. T. Wojtach (2015).

12. Evidence from the relatively new field of epigenetics makes this statement less absolute than it used to be, but the argument stands.

APPENDIX
An Overview of the Human Auditory System

This primer is much abbreviated and amended information from Chapter 13 of Purves et al., *Neuroscience,* 5th edition (2012), and Chapter 4 of Purves et al., *Principles of Cognitive Neuroscience,* 2nd edition (2013). It is presented here to give readers a general idea of the sensory system that, among other things, is critical for an appreciation of music. It should be understood, however, that other brain systems such as those that generate emotion and speech are also involved. Thus the relationship between the structure and function of the human auditory system and music is only part of a complex anatomical and physiological story that is not well understood.

The Ear

In humans and other mammals, the auditory system transforms mechanical energy carried by the movement of air molecules into neural signals that ultimately give rise to the perceptual qualities we hear, as described in Chapter 1. The first stage of this transformation entails pre-neural effects produced by the external ear and the middle ear (see Figure 1.1). By virtue of their anatomy and resonance properties, these structures collect, filter, and amplify sound energy so that stimuli of particular ecological importance (e.g., vocalizations) are transmitted with greater emphasis and efficiency. Thus, the odd-looking cartilaginous structures of the external ear, called the *concha* and *pinna*, function much like an old-fashioned "ear trumpet" to collect and focus sound energy, while the resonance of the ear canal helps filter out the less relevant aspects of sound signals. The three bones of the middle ear link the resulting deflections of the tympanic membrane (eardrum) to the inner ear, further enhancing the energy transmitted to the inner ear at the oval window. This bony mechanism between the eardrum and oval window enhances pressure in much the same way that the pressure on the plunger of a syringe is increased at the bore of the needle.

The Cochlea

To reiterate some of the anatomical points made in Chapter 1, the oval window marks the entry to the cochlea, which houses the neural receptor apparatus of the inner ear (see Figure 1.1). The major features of the cochlea, so named because the overall shape of its bony shell is similar to that of a snail (*cochlea* means “spiral shell” in Latin), are the basilar membrane and its embedded receptor cells, the hair cells. The movement of the oval window is transmitted to the fluid in the cochlea, which in turn moves protrusions on the tips of the hair cells called stereocilia. The movement of the stereocilia depolarizes the membrane of the hair cells, leading to the release of transmitter molecules from their basal ends, which in turn elicits synaptic potentials and, if these are sufficient, action potentials in the endings of the axons that form the auditory nerve. The relevant neuronal cell bodies are in the nearby spiral ganglion (see again Figure 1.1).

Action potentials in auditory nerve fibers convey information about the frequency, amplitude, and phase of sound stimuli to the auditory processing regions of the brain, leading eventually to the primary auditory cortex and higher-order auditory cortices (see Figure A.1). The frequency and intensity of a given sound signal are transduced, respectively, by the region of the basilar membrane that is most affected by a stimulus and the amplitude of the deflection. These spatially specific responses of the basilar membrane rely on its mechanical properties: the stiffer portion near the oval window moves in response to relatively high frequencies, while the more compliant portion at the cochlear apex reacts to low frequencies (see Figure 2.1).

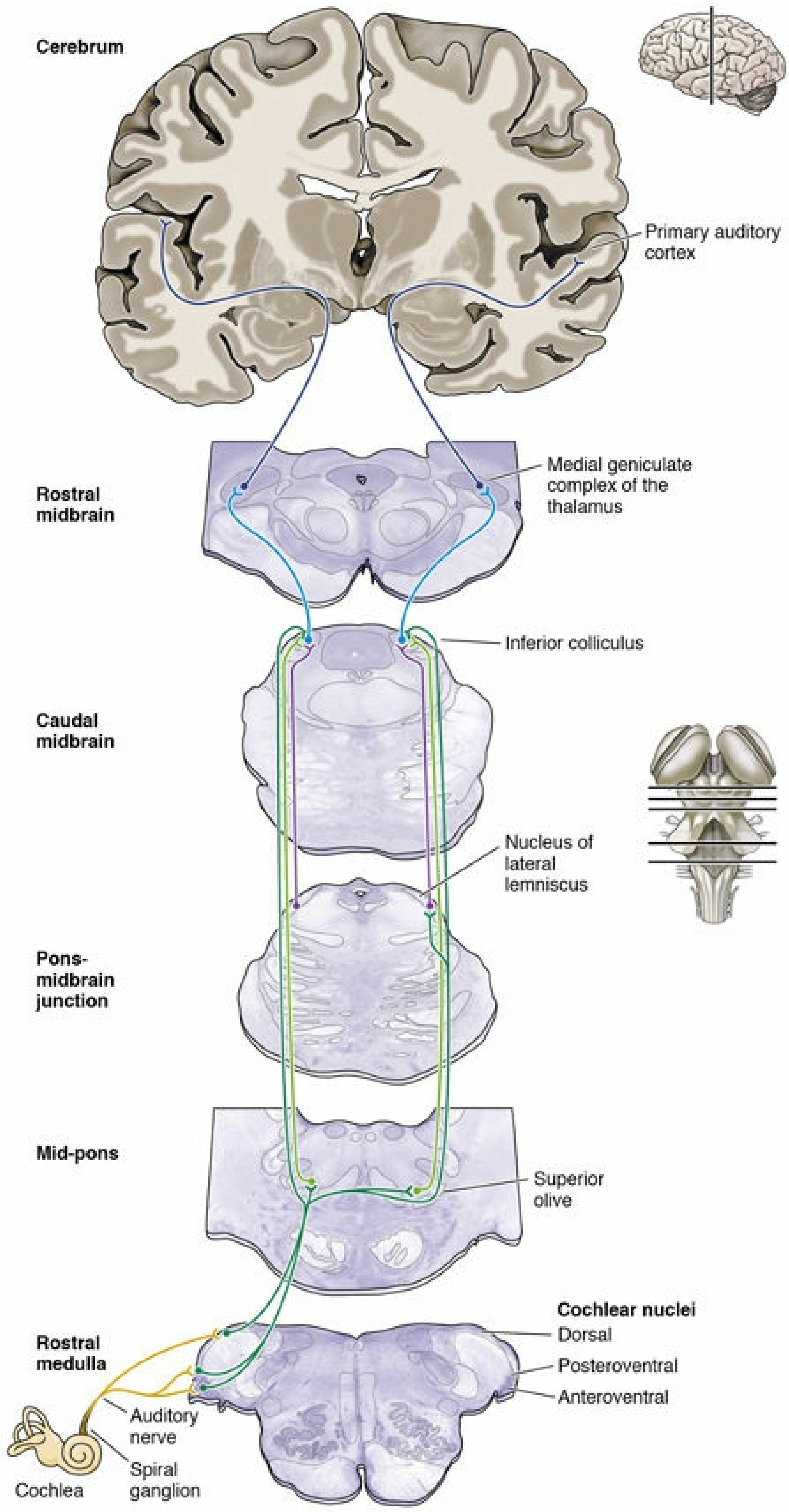


FIGURE A.1 The primary auditory pathway, showing the major stations of the human auditory system. The insets indicate the levels at which the cross sections in the diagram are taken. Only the ascending pathways are shown; there are equally complex descending pathways at every processing level.

The Primary Auditory Pathway

The primary auditory pathway shown in Figure A.1 begins with the hair cells in the cochlea and entails further processing in the cochlear nuclei in the brainstem, the inferior colliculi in the midbrain, and the medial geniculate nuclei of the thalamus before arriving at the primary auditory cortex (see below).

The first stage of central auditory processing occurs in the *cochlear nucleus* in the rostral medulla of the brainstem, the initial target of the auditory nerve axons that convey information from the basilar membrane. From there, peripheral auditory information diverges into a number of parallel pathways that project to one or more of several targets:

- The *superior olivary complex*, the first place that information from the two ears interacts, is the site of the initial processing of cues that allow listeners to localize sound signal sources in space.
- Axons from the neurons of the cochlear nucleus also project to the *inferior colliculus* in the midbrain, a major integrative center and the first place where auditory information interacts with the motor system to initiate auditory-guided behavior (e.g., turning the head toward a sound in order to see what caused it).
- Neurons in the cochlear nuclei also send projections to the nucleus of the *lateral lemniscus* in the midbrain, whose neurons process temporal aspects of sound stimuli that are also involved in locating sound signal sources in space.

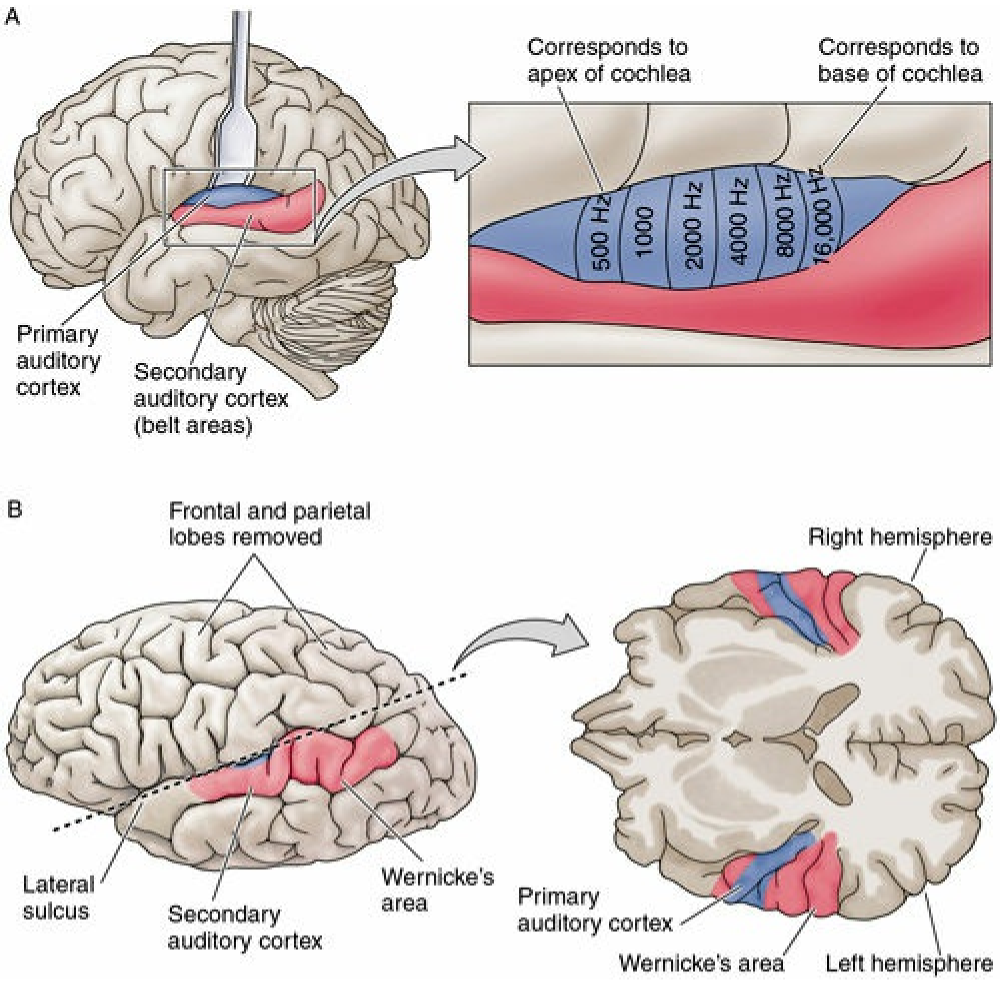
As in the case of the other the major sensory modalities (with exception of olfaction), information from these stations in the brainstem and midbrain is sent to the *thalamus*, where it is further processed and relayed to the *primary auditory cortex*. The relevant thalamic target in this case is the medial geniculate nucleus, a station homologous to the lateral geniculate nucleus in the primary visual pathway. How these thalamic nuclei alter the incoming information is unclear

The Auditory Cortices

The auditory cortices are located in the superior temporal lobe and adjacent regions of the parietal lobe. Like the other sensory cortices, the auditory cortex is divided into primary and secondary regions (Figure A.2). The primary auditory cortex (also called A1) lies on the superior aspect of the temporal lobe along the superior temporal gyrus, and is defined by being the major cortical recipient of the thalamic projections. The adjoining areas of the temporal and parietal lobes comprise the secondary auditory cortex (called A2; the auditory cortical areas surrounding the primary auditory cortex are also referred to as “belt” areas). These secondary regions are where higher-order auditory processing occurs, including the processing germane to understanding speech sounds and the recognition and comprehension of words (see below). These auditory areas are thus similar sensory association areas where more complex integrative processing of stimuli occurs. These secondary areas combine information from other sensory modalities as well as from additional brain regions.

An important feature of the primary auditory cortex is its *tonotopic* arrangement, which accords with the tonotopy of the basilar membrane: in both instances neurons are selective for particular stimulus frequencies and arranged from low to high in an orderly “map.” The higher-order processing of sound stimuli that gives rise to auditory percepts occurs in the *secondary auditory cortex* adjacent to A1, which is roughly analogous to the cortical areas adjacent to V1 in the visual system of S1 in the somatic sensory system. As shown in Figure A.3, the tonotopic arrangement in A1 is reiterated at least twice more in these higher-order areas.

Information from both ears is processed in both hemispheres, although there is a slight tendency toward greater hemispheric processing of signals originating in the opposite ear. This bilateral processing is in contrast to the visual or somatic sensory systems in which information arising from one side of the visual field or the body is processed in the opposite hemisphere.



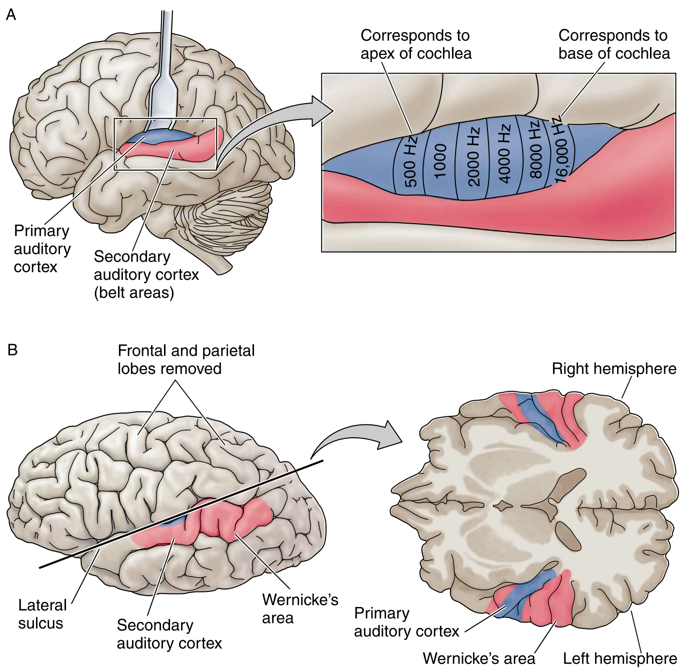


FIGURE A.2 Location and tonotopic organization of the primary auditory cortex.

Higher-Order Cortical Processing

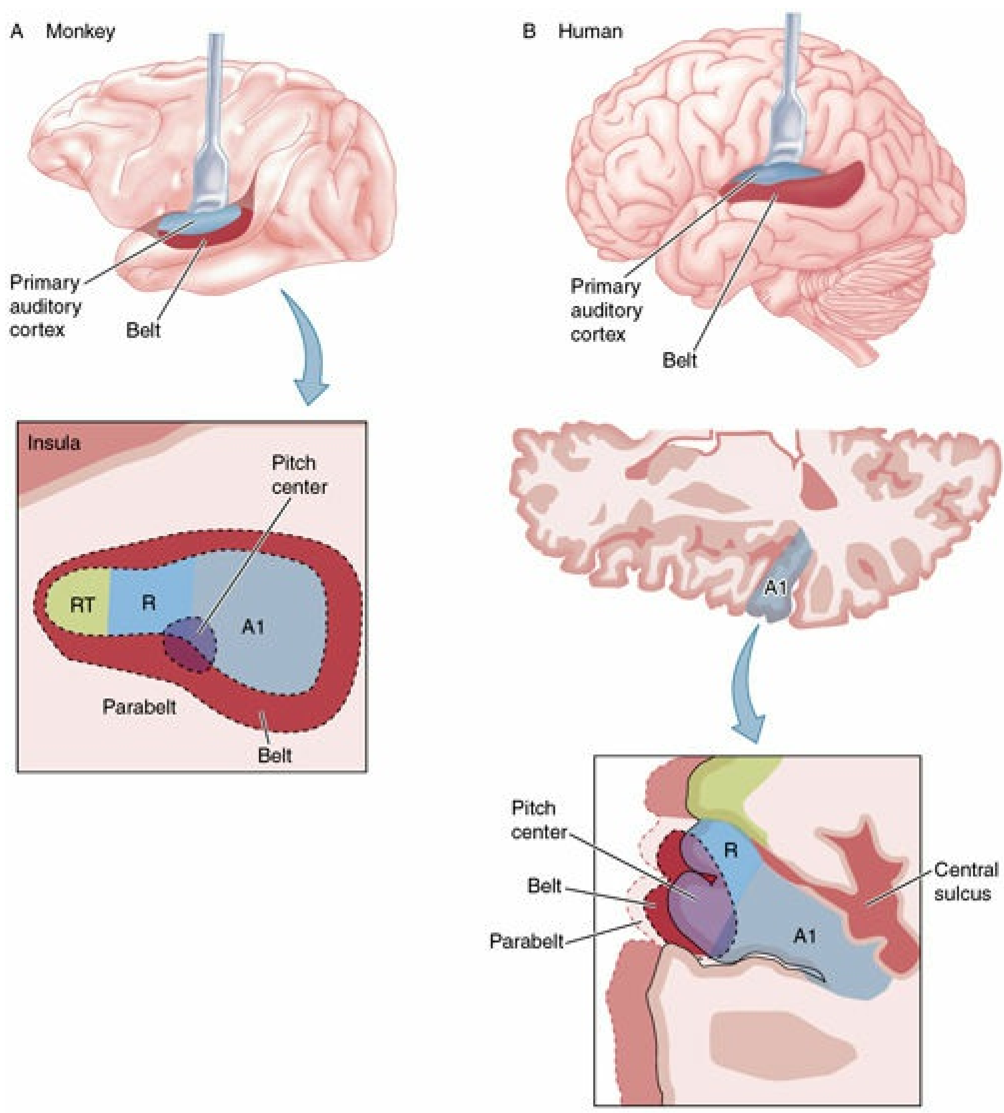
As in other sensory systems, higher-order auditory cortical areas tend to be specialized for processing particular categories of information based on biological importance (Figure A.3; see also Chapters 2 and 3).

The clearest evidence about the organization of these specialized secondary areas involves areas devoted to speech and tonal processing in both humans and non-human primates. Some key discoveries about these higher-order auditory specialization are:

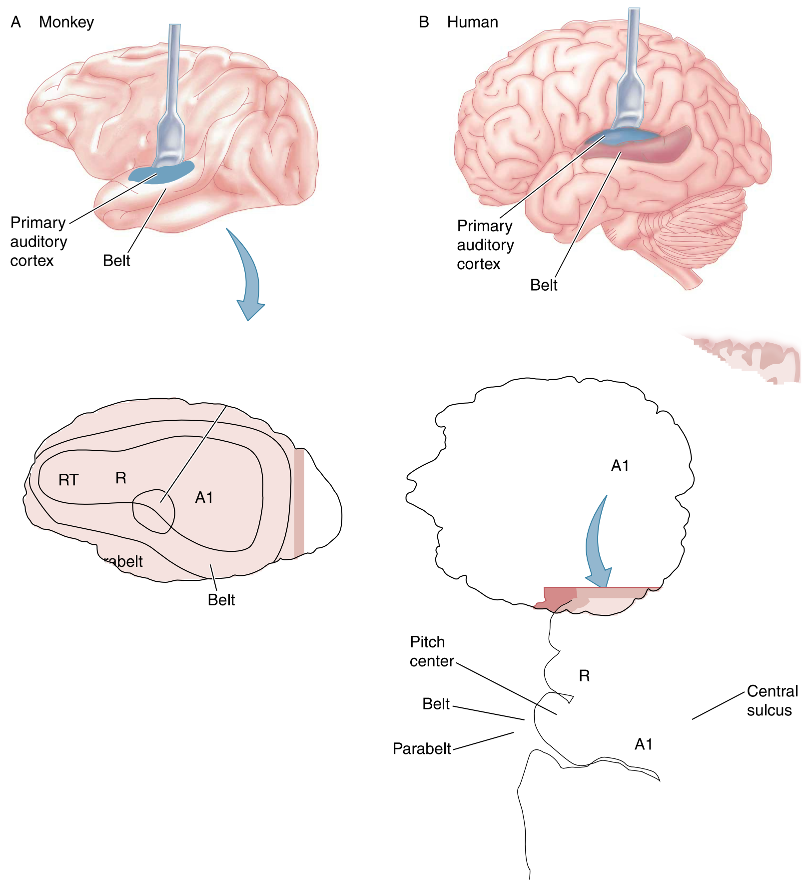
- *An area adjacent to A1 in the superior and posterior region of the temporal lobe called Wernicke's area.* This region links speech sounds to their meanings. Thus patients with lesions in this general region of the left hemisphere tend to suffer a particular language deficiency called comprehension aphasia, in which the affected individuals are able to produce fluent speech, but unable to use words in the correct way (i.e., the meanings are garbled). This disorder contrast with the production aphasia due to damage of the motor areas in the left frontal lobe (meanings are clear but fluent production is slow and halting)

- *Functional magnetic resonance imaging (fMRI) evidence for the existence of an area of auditory cortex that is particularly concerned with the motion of auditory signals.* This specialization for processing information about sound sequences in space is arguably similar to the specialization of some regions of secondary (extrastriate) visual cortex for processing image sequences that give rise to the visual motion.

- *Animal studies showing that important "natural" sound stimuli are further processed in especially well-developed cortical areas in many species.* In humans, it has been known for more than a century that the regions used to process speech sounds are not only overrepresented but also lateralized (see above). Thus, speech sound processing is predominantly carried out in the left hemisphere in most people, whereas the processing of other environmental sound stimuli occurs in both hemispheres.
- *The processing of tonal harmonics evident in the belt areas of nonhuman primates.* As mentioned in Chapter 9, electrophysiological recordings in monkeys show that many neurons are driven not only by the fundamental frequencies of tones, but also by integer multiples of their fundamental frequencies. Many belt area neurons also respond to stimuli that are missing the fundamental frequency of a harmonic series, much like this psychophysical phenomenon in humans (see Chapter 4).
- *Neurons in the auditory cortex of non-human primates that are especially responsive to conspecific vocalizations.* These observations show that the general importance of vocal sounds that lack language. In some studies, animals are specifically responsive to different types of calls, suggesting the roots of human language in other extant primates.
- *Auditory nerve axons tuned to broadband stimuli characteristic of the sounds animals hear in nature.* Such responses are also consistent with the idea that natural sound signals are more fully and efficiently processed than sound stimuli created in the laboratory.

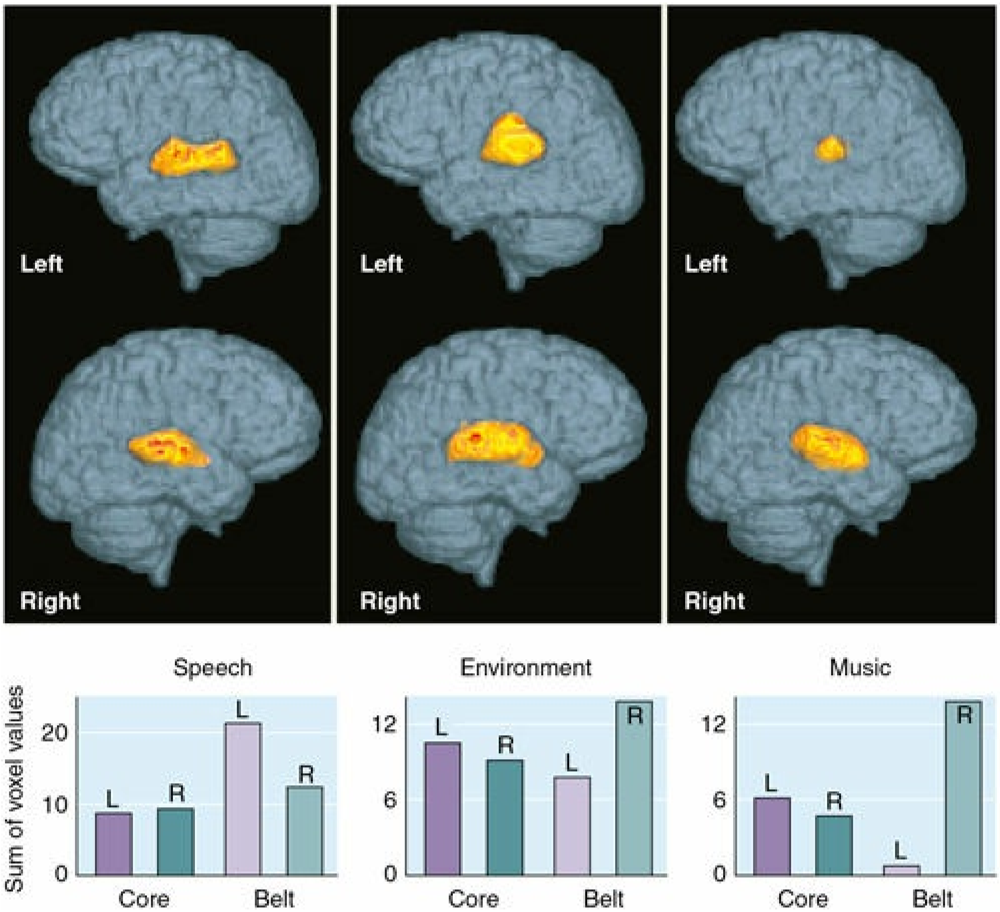


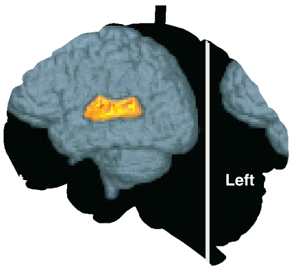
A Monkey
B Human
Primary auditory cortex
Belt
Primary auditory cortex
Belt
Insula
Pitch center
RT
R
A1
Parabelt
Belt
A1
Pitch center
Belt
Parabelt
R
A1
Central sulcus



A Monkey
B Human
Primary auditory cortex
Belt
Primary auditory cortex
Belt
RT
R
A1
Belt
A1
Pitch center
Belt
Parabelt
R
A1
Central sulcus

FIGURE A.3 Specialized processing in the auditory cortex of a nonhuman primate determined by electrophysiological recording compared to human cortical anatomy. (A) The location of neurons in the monkey auditory cortex that respond specifically to the missing fundamental of a sound stimulus is the region labeled "Pitch center." (B) The approximate location of this region in the human auditory cortex is shown for comparison. R and RT indicate more rostral auditory regions in which the primary auditory (A1) map of tonotopy is reiterated. (From Purves et al., 2013; after Bendor and Wang, 2006a, b.)





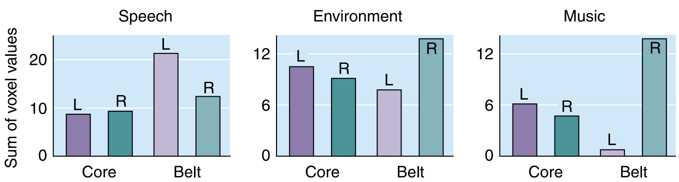


FIGURE A.4 Functional MRI evidence for lateralization of musical sound processing. In most individuals, musical sounds signals tend to be processed primarily in the auditory cortex (belt areas) of the right temporal lobe. The graph shows the relative activation in the primary and secondary areas of auditory cortex in each hemisphere in the subject studied, which contrasts with the predominantly left-hemisphere processing of speech sound stimuli. (From Purves et al., 2012, courtesy of Jagmeet Kanwal.)

This range of observations indicates that processing in the human auditory cortex is strongly biased in favor of stimuli whose perception is particularly relevant to successful behavior, social sound signals in particular.[1]

Brain Areas Activated by Music

In contrast to vocal signals, there is not much consensus about the representation of the music in the brain, a large number of recent studies using fMRI notwithstanding (see Chapter 9). Some studies suggest that in contrast to the predominantly left hemisphere activity elicited by speech sounds, the auditory cortex is more balanced, with only a modest bias toward the right hemisphere processing for musical sound signals (Figure A.4). This result is consistent with other evidence showing that speech prosody, which as described in Chapter 7 helps convey emotion in speech, is also more strongly represented in the auditory areas of the right hemisphere.

Many other potentially important observations about human brain activity in response to music have been made, but there is little agreement about what such activity implies beyond the fact that music stimulates a wide range of brain areas, including those associated with emotion.

1. The organization of the human versus non-human primate auditory cortex remains debated. A good review of the controversy is S. Baumann, C. I. Petkov, and T. D. Griffiths (2013).

Glossary

action potential The electrical signal conducted along neuronal axons by which information is conveyed from over long distances in the nervous system.

adaptation Resetting receptors or other elements in a sensory system to different levels of sensitivity; allows sensory systems to operate over a wide range of input values despite the limited number of action potentials per unit time that neurons can produce.

aesthetics In biological terms, the principles that govern which stimuli in a given sensory modality attract us and which do not.

aphasia A language deficit that arises from damage to one of the cortical language areas, typically in the left hemisphere.

atonal music A form of modern music that intentionally avoids the creation of a musical anchor, or “home base.” See *12-tone music.*

auditory scene The overall perception of the auditory environment at a point in time. Analogous to the perception of a visual scene.

awareness A cognitive / perceptual state in which an individual (or non-human animal) can report subjective experience.

axon The extension of a neuron that carries the action potential from the nerve cell body to a target cell.

bandwidth A range of frequencies used to transmit sound or other signals.

Baroque music The corpus of music composed between ~1600 and 1750; signifies the beginning of the prevalent use of major and minor scales, often with melodies elaborately ornamented.

basilar membrane The membrane in the inner ear (cochlea) that contains the receptor neurons (hair cells) that initiate audition.

beating Fluctuating sound signal amplitudes perceived when two or more signals constructively and destructively interfere with each other to create a sense of auditory "roughness."

brainstem The portion of the brain that lies between the diencephalon and the spinal cord; comprises the midbrain, pons, and medulla.

cadence The end of a musical phrase or piece of music; typically characterized by the use of the tonic, or notes / harmonies closely related to the tonic.

Carnatic music A form of classical South Indian music derived from Hindu tradition that emphasizes vocal music.

central nervous system The brain and spinal cord of vertebrates (and by analogy, the central nerve cord and ganglia of invertebrates).

cerebral hemispheres The two halves of the forebrain in mammals.

chord Two or more notes played more or less simultaneously to create harmonies. Often thought of as triad that comprises the tonic or reference note together with the third and fifth chromatic intervals.

chromatic scale The 13 notes and 12 intervals over an octave that comprise the superset from which the scales in most Western music, as well as some other traditions, are taken.

circumplex model A graphical representation of the relationships among emotions by ordering them around a circle, with axes indicating valence and arousal.

cochlea The portion of the inner ear specialized for transducing sound energy into neural signals (the other component is the vestibular system, which monitors head position and acceleration).

consonance The subjectively pleasing sense elicited by specific combinations of notes in harmonies or melodies.

consonant Typically an unvoiced (atonal) element of speech that begins and / or ends syllables.

cortical association areas The regions of cerebral cortex that are not involved in primary sensory or motor processing.

critical bands ~1.0-millimeter lengths along the basilar membrane; frequencies that fall within such bands interfere with each other, creating a perception of beating.

dendrite The branches of a neuron that receive synaptic input.

diatonic scale A heptatonic scale defined by the seven Greek modes.

dissonance When melodic or harmonic combinations of notes are perceived as relatively inharmonious or less pleasing than consonant combinations.

dynamic A term referring to the composer's intended loudness for a section of a musical piece (designated by the notations *forte, mezzo forte, piano,* etc.).

ear canal The tube that extends from the external ear to the eardrum.

emotion Physiological responses and subjective feelings that motivate humans and other animals to react appropriately to events of biological and / or individual significance.

equal temperament A tuning system nearly universal today that allows playing in different keys without retuning. Each semitone is tuned to be exactly the same normalized frequency distance from the preceding semitone in the chromatic scale (a 5.9463 percent increment).

external ear The cartilaginous elements of the visible parts of the ear (the pinna and concha).

forebrain The anterior portion of the mammalian brain that includes the cerebral hemispheres (the telencephalon and diencephalon).

formant One of several frequency peaks in the harmonic spectrum of voiced vocalizations; determined by the resonances of the vocal tract above the vocal folds.

functional magnetic resonance imaging (fMRI) A noninvasive method for imaging the locus of human brain activity in a given condition or during a task; the measured signal is caused by changes in blood flow and oxygenation induced by local neural activity.

fundamental frequency The first vibratory mode in the harmonic series evident in a sound spectrum generated by a vibrating string or column of air.

gamaka An ornament in classical Indian music where a note is "shaken," or where a musician "slides" to the next note.

Gamelan music Music performed in parts of Java by a traditional instrumental ensemble that uses of bells, gongs, cymbals, and other metallophone instruments.

gyrus A "peak" in the corrugation of the cerebral cortex.

hair cell The receptor cell in the inner ear that transduces the mechanical energy in sound stimuli into neural signals.

harmonic minor scale A natural minor scale with the seventh scale degree raised to create a stronger musical pull back toward the tonic note.

harmonic series The series of vibratory modes evident in the spectra of strings, air columns, or other objects that produce amplitude peaks at integer multiples of their fundamental resonance frequency.

harmony Combinations of musical tones played more or less simultaneously to create chords.

heptatonic scale A scale that uses eight notes that divides octaves into seven intervals. See *diatonic scale; mode*.

hexatonic scale A scale that uses seven notes to divide the octave into six intervals; not as formally prominent Western music theory as the pentatonic and heptatonic scales, but provides the foundation of blues music.

inference The proposition that our brains create percepts using unconscious guesses about the significance of ambiguous stimuli based on past experience.

inferior colliculus Midbrain station in the primary auditory pathway where sound is integrated with movement. Also receives input from other sensory systems.

interval The frequency distance between two notes, often expressed as a ratio in music theory.

inverse problem The impossibility of knowing the physical parameters of the world by means of sensory stimuli.

just intonation A tuning system that parses the octave into intervals according to small whole number ratios derived from subjective consonance.

lamellophone Family of musical instruments that entail metal plates fixed at one end, such a Jew's harp, a music box, or the African *mbira.*

larynx The portion of the respiratory tract that lies between the trachea and the pharynx. Contains the vocal folds (also called vocal cords).

lateral lemniscus Station in the primary auditory pathway concerned with sound localization. Projects primarily to the inferior colliculus.

learning A change in performance (behavior) caused by experience; presumed to arise by cell biological mechanisms that alter the strength of neural connections (synapses).

loudness The perception of sound signal intensity; measured in newtons / m^2.

low note The note in a musical scale with the lowest frequency. Also called the reference note of a scale.

major scale A scale that divides the octave into the specific intervals that emphasize whole tone (major) intervals.

mbira A Zimbabwean instrument that consists of 22 to 28 metal keys fixed at one end to a wooden soundboard.

mechanoreceptors Neurons that respond to mechanical energy; includes the hair cells in the inner ear.

medial geniculate nucleus Region of the thalamus that processes and relays auditory information to the primary auditory cortex.

melisma Singing the same syllable using a succession of notes.

melodic interval The difference in pitch height between two sequential notes in a melody line.

melodic minor scale A musical scale that has the same intervals as the natural minor scale when descending, but raised sixth and seventh scale degrees when ascending, allowing smoother melodic transitions. Creates a musical pull toward the fifth scale degree if descending, and toward the tonic if ascending.

melody A sequence of musical tones.

metallophone A family of musical instruments that entails tuned metal bars; characteristic of some traditional Indonesian music (see *Gamelan music*).

midbrain The most rostral of the three components of the brainstem.

middle ear The portion of the ear between the eardrum and the oval window; contains the three small bones that amplify sound signals mechanically.

modes A set of seven heptatonic scales, each with a different Greek name and a nominally different quality; modes were especially popular during the Renaissance, and include the major (Ionian) and minor (Aeolian) scales that are dominant today.

music Complex periodic sounds produced by a variety of physical instruments (including the human vocal tract) that are appreciated by humans as pleasing and affective, typically implemented by specific tone combinations in melodies and harmonies.

natural minor scale A musical scale that contains the same intervals as the major scale except that the third ("mi"), sixth ("la") and seventh ("ti") scale degrees are lowered by one semitone.

nervous system The full complement of neurons and their connections throughout the body of an animal.

neural circuit A collection of interconnected neurons mediating a specific function.

neuron A cell specialized for processing information by means of bio-electrical signals. Also called a nerve cell.

neurotransmitter A chemical agent released at synapses that mediates signaling between nerve cells.

noise A sound signal that is aperiodic and perceived as atonal. Compare *tone.*

nucleus An anatomically identified cluster of nerve cells with a shared function.

octave The frequency distance defined by doubling of the fundamental frequency of a periodic sound signal.

parietal lobe One of the four lobes of the cerebral cortex; along with the superior temporal lobe houses regions of the higher auditory cortex concerned with speech recognition and comprehension.

pentatonic scale A scale with six notes that divide octaves into five intervals. Prevalent in relatively simple music (e.g., folk music).

perfect Referring to a dyadic ratio that is non-repeating as a decimal (e.g., 1.5 or 2.0).

peripheral nervous system All the nerves and neurons that lie outside the brain and spinal cord. Compare central nervous system.

phone One of about 200 different sound signals the human vocal apparatus can produce. A subset of these is used in any given language.

phoneme The perceptual response generated by a phone.

pitch The perceived height of a sound signal that has a periodic frequency over the range of human hearing.

pitch shift of the residue The change in pitch that occurs when a set of successive harmonics are adjusted by a constant amount such that they no longer have a common divisor.

pitch strength Refers to the fact that sound signals with frequency repetition rates of ~200 to 500 Hz evoke a stronger sense of pitch than signals outside this range.

pons One of the three components of the brainstem, lying between the midbrain rostrally and the medulla caudally.

primary auditory cortex (A1) The cortical target of the neurons in the medial geniculate nucleus; the terminus of the primary auditory pathway. Compare *secondary auditory cortex.*

primary auditory pathway The pathway from the inner ear to the primary auditory cortex in the temporal lobe.

prosody The fluctuating pitch heights that endow speech with emotional and other information. In some accounts prosody includes the timbre and loudness.

Pythagorean tuning A tuning system based on increasing the fundamental frequency of each successive note in a scale by a fifth (i.e., by a ratio of 3:2 or 50 percent).

raga Any one of a number of scales used in classical Indian music.

rasa One of nine emotional themes associated with different *ragas*. Comparable to the affective intent of modes.

receptive field The region of the receptor surface (e.g., basilar membrane, retina, skin) that when stimulated elicits a response in the neuron being examined. Can also allude to the stimulus properties that neuron in question prefers.

reference note The low note in a scale, to which the other notes in the scale are referred.

register The height of an octave on the human scale of pitch.

resonance The inherent tendency of physical objects to vibrate maximally at a certain frequency.

rhythm Patterns of accented versus unaccented beats in music.

root note The tonic or reference note in a chord.

roughness The generally annoying quality of a sound signal comprising two or more tones whose component frequencies interfere with each other to create audible "bumps."

scale A collection of notes (pitch heights) and intervals used in music; bounded by a low starting note and the note octave above.

secondary auditory cortex (A2) Cortical region surrounding the primary auditory cortex concerned with higher-order processing of auditory information. Also called "auditory belt area."

semitone The smallest interval in the Western chromatic scale; approximately a 6 percent increment in frequency over that of the preceding note.

single-unit recording A method of studying the activity of individual neurons using a microelectrode.

slope reversals Changes in the direction of a contour of a melody; i.e., when the pitch goes down after previously going up, or vice versa.

sound pressure level The physical parameter used to measure sound signal intensity in newtons / m^2.

sound signal Pressure change in the local atmosphere within the range of human hearing that arises when an resonant object is disturbed by a force.

sound spectrum An analysis of a sound signal that reveals the distribution of its energy (signal amplitude) as a function of frequency.

sound stimulus Result of the transformation of a sound signal into neural activity by the receptor neurons in the basilar membrane of the inner ear.

sound wave The periodic compression and rarefaction of air molecules in a sound signal.

source-filter model A generally accepted model for the production of speech sound signals that entails the vocal fold vibrations as a source, and the rest of the vocal tract as a dynamic filter.

spectral dominance The fact that when the frequencies of only some of the harmonics of a series are increased or decreased, listeners perceive a fundamental that is biased toward the midrange of speech signal frequencies.

spiral ganglion Locus of the cell bodies of the neurons whose bipolar axons form the auditory nerve in one direction and in the other contact the base of the hair cells.

standing wave The wave generated by a taut string with both ends fixed; vibrates in a series of modes caused by the reflection of the wave that cancels oscillations that are not integer multiples of the fundamental frequency. Can also be created in a column of air.

string bending A technique used by guitar players to create a continuum of frequencies between two semitones by pushing a sting laterally

subcortical Pertaining to brain structures that anatomically underlie the cerebral cortex.

sulcus A "valley" in the corrugation of the cerebral cortex.

superior olivary complex A complex of brainstem nuclei in the primary auditory pathway.

synapse A specialized contact between the axon of a neuron (the presynaptic cell) and a target (postsynaptic) cell. Information is transferred between the presynaptic and postsynaptic cells by the release and receipt of neurotransmitter molecules.

temporal lobe One of the four lobes of the cerebral cortex; houses the primary auditory cortex.

tetrachord Four adjacent notes comprising three intervals; a formal structure in early Western music before pentatonic and heptatonic scales became prevalent.

tetrad A combination of four notes played together as a chord.

thalamus A collection of nuclei that forms the major component of the diencephalon. A primary role is to relay sensory information from the periphery to the cerebral cortex, and to integrate this input with cortical feedback.

timbre The perceptual quality that distinguishes tones that have the same pitch and loudness; arises from the distribution of energy in a sound signal over time.

tonal Pertaining to a sound stimulus that, by virtue of its periodic repetition, produces the perception of a tone.

tone The sound heard in response to a particular frequency of vibration or combination of vibrations that are strongly periodic. Compare *noise.*

tonic chord The triad of notes (the tonic or root note, together with a third and a fifth above it) that forms the anchor or home base chord in most musical compositions.

tonic interval The difference in pitch height between the higher note of a dyad and the tonic note.

tonic note Note that forms a reference for the other notes in a dyad or chord. Usually used in the context of harmony. Also called the root note in a chord.

tonotopy The anatomical organization of the auditory cortex that accords with the systematic layout of responses to different tones on the basilar membrane.

transduction The conversion of energy into a neural signal.

traveling wave A wave generated when only one end of a string or other object is fixed, as is the case for the basilar membrane.

triad A combination of three notes played together.

tritone The interval between the perfect fourth and the perfect fifth in the chromatic scale; also called a diminished fifth or an augmented fourth, depending on whether P5 is lowered or P4 raised.

tuning system A protocol that dictates how to adjust the tonal sounds of an instrument to generate specific musical intervals.

12-tone music Music composed so that all notes in the chromatic scale occur equally. See *atonal music.*

tympanic membrane The eardrum.

valence The subjective attractiveness or unattractiveness of a stimulus.

vibrato A pulsating effect accomplished by rapidly changing the pitch; often used to add emotional intensity to a note that is being held.

vocal similarity theory Referring to the general idea that the attraction of tonal music derives from the biological advantages of recognizing and responding to the harmonic series characteristic of the human voice.

voiced speech Pertaining to a speech sound signal characterized by laryngeal harmonics, typically a vowel sound.

vowel Typically a voiced (tonal) element of speech that forms the nucleus of syllables.

wave cycle All the positions a wave can inhabit between one amplitude maximum and the next.

waveform The shape of a wave over time.

whole tone An interval in Western and many other scales that comprises two semitones (e.g., the frequency distance between “do” and “re”).

Bibliography

Abramson, A. S. (1962). The vowels and tones of standard Thai: acoustical measurements and experiments. Int J Am Ling 28 (2).

Alain, C., and S. R. Arnott (2000). Selectively attending to auditory objects. Frontiers Biosci 5: 202–212.

Aldwell, E., and C. Schacter (2003). *Harmony and Voice Leading.* Florence, KY: Wadsworth / Thomson.

Aoyagi, T., and R. A. Kendall (1996). Perceptual relationships among nine septatonic western modes: Convergence across distinct methods. In: *Proceedings of the Fourth International Conference on Music Perception and Cognition* (Pennycook, B., and E. Costa-Giomi, eds). pp. 65–70. Montreal, PQ, Canada: Society for Music Perception and Cognition.

Atcherson, W. (1973). Key and mode in seventeenth-century music theory books. J Music Theory 17: 204–232.

Bagchee, S. (1998) *Nad: Understanding Raga Music.* Delhi, India: BPI (India).

Balzano, G. J. (1980) The group-theoretic description of 12-fold and microtonal pitch systems. Comp Mus J 4: 66–84.

Banse, R., and K. R. Scherer (1996). Acoustic profiles in vocal emotion expression. J Pers Soc Psychol 70(3): 614–636.

Barbour, J. M. (1951). *Tuning and Temperament: A Historical Survey.* Lansing: Michigan State College Press.

Barlow, H., and S. Morgenstern (1974). *A Dictionary of Musical Themes.* New York: Crown.

Baumann, S., C. L. Petkov, and T. D. Griffiths (2013). A unified framework for the organization of the primate auditory cortex. Front Syst Neurosci 7:11. doi: 10.3389/fnsys.2013.00011. eCollection 2013.

Beck, J., and W. A. Shaw (1961). The scaling of pitch by the method of magnitude-estimation. Am J Psychol 74(2): 242–251.

Belin, P., R. J. Zatorre, P. Lafaille, P. Ahad, and B. Pike (2000). Voice-selective areas in human auditory cortex. Nature 403: 309–312.

Bendor, D., and X. Wang (2006a). Cortical representations of pitch in monkeys and humans. Curr Opin Neurobiol 16: 391–399.

Bendor, D., and X. Wang (2006b). The neural representation of pitch in primate auditory cortex. Nature 436: 1161–1165.

Berckmoes, C., and G. Vingerhoets (2004). Neural foundations of emotional speech processing. Am Psychol Soc 13(5): 182–185.

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

Bernstein, L. (1976). *The Unanswered Question: Six Talks at Harvard.* Cambridge, MA: Harvard University Press.

Bilsen, F. A. (1995). What do dichotic pitch phenomena tell us about binaural hearing. In: *Advances in Hearing Research* (G. A. Manley, G.M. Klump, C. Köppl, and H. Fastl, eds.). Singapore: World Scientific, pp. 334–341.

Blood, A. J., and R. J. Zatorre (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proc Natl Acad Sci USA 98: 11818–11823.

Blood, A. J., R. J. Zatorre, P. Bermudez, and A. C. Evans (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. Nat Neurosci 2: 382–387.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proc Inst Phonetic Sci 17: 97–110.

Boersma, P., and D. Weenink (2001). PRAAT 4.0.7: Doing phonetics by computer. Department of Phonetic Sciences, University of Amsterdam (available at http://fonsg3.let.uva.nl/praat).

Bowling, D. L., and D. Purves (2015). A biological rationale for musical consonance. Proc Natl Acad Sci USA 112: 11155–11160. doi: 10.1073 / pnas.1505768112

Bowling, D. L., K. Gill, J. D. Choi, J. Prinz, and D. Purves (2010). Major and minor music compared to excited and subdued speech. J Acoust Soc Am 127(1): 491–503.

Bowling, D. L., J. Sundararajan, S. Han, and D. Purves (2012). Expression of emotion in Eastern and Western music mirrors vocalization. PLOS ONE 7(3): e31942. doi:10.1371 / journal.pone.0031942.

Braun, M. (1999). Auditory midbrain laminar structure appears adapted to f0 extraction: Further evidence and implications of the double critical bandwidth. Hear Res 129: 71–82.

Brazil, D. (1997). *The Communicative Value of Intonation in English.* New York: Cambridge University Press.

Bregman, A. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge, MA: MIT Press.

Bright, W. (1963). Language and music: areas for cooperation. Ethnomusicology 7(1): 26–32.

Broughton, S., M. Ellingham, and R. Trillo (2000). *World Music: The Rough Guide.* London: Rough Guides.

Brown, K., and S. Ogilvie (2009). *Concise Encyclopedia of Languages of the World.* New York: Elsevier.

Brown, S. (2000). The "music language" model of music evolution. In: *The Origins of Music* (N. L. Wallin, B. Merker, and S. Brown, eds.), pp. 271–300. Cambridge, MA: MIT Press.

Brown, S., and J. Jordania (2011). Universals in the world's music. Psych Music 41: 229–248.

Budge, H. (1943). *A Study of Chord Frequencies.* New York: Bureau of Publications, Teachers College, Columbia University.

Burkholder, J. P., D. J. Grout, and C. V. Palisca (2014). *A History of Western Music,* 9th ed. New York: Norton.

Burns, E. M. (1982). Pure-tone pitch anomalies. I. Pitch-intensity effects and diplacusis in normal ears. J Acoust Soc Am 72(5): 1394–1402.

Burns, E. M. (1999). Intervals, scales, and tuning. In: *The Psychology of Music* (D. Deutsch, ed.), pp. 215–264. New York: Academic.

Butler, J. W., and P. G. Daston (1968). Musical consonance as musical preference: A cross-cultural study. J Gen Psych 79: 129–142.
Buus, S., H. Musch, and M. Florentine (1998). On loudness at threshold. J Acoust Soc Am 104(1): 399–410.
Campbell, S., and C. Shaweevongs (1957). *The Fundamentals of the Thai Language.* Chicago: Paragon Book Gallery.
Cansino, S., A. Ducorps, and R. Ragot (2003). Tonotopic cortical representation of periodic complex sounds. Hum Brain Mapp 20: 71–81.
Cariani, P. A., and B. Delgutte (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J Neurophys 76: 1698–1716.
Cariani, P. A., and B. Delgutte (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. J Neurophys 76: 1717–1734.
Carterette, E. C., and R. A. Kendall (1999). Comparative music perception and cognition. In: *The Psychology of Music* (D. Deutsch, ed.), pp. 725–791. New York: Academic.
Chen, Z., G. Hu, B. R. Glasberg, and B. C. J. Moore (2011). A new method of calculating auditory excitation patterns and loudness for steady sounds. Hear Res 282: 204–215.
Christensen, T. (1993). *Rameau and Musical Thought in the Enlightenment.* Cambridge, UK: Cambridge University Press.
Cohen, A. (1961). Further investigation of the effects of intensity upon the pitch of pure tones. J Acoust Soc Am 33: 1363–1376.
Cohen, D. (1971). Palestrina counterpoint: A musical expression of unexcited speech. J Music Theory 15: 85–111.
Cook, P. (1999). Pitch, periodicity, and noise in the voice. In: *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics* (P. Cook, ed.), pp. 195–208. Cambridge, MA: MIT Press.
Cook, P. R., ed. (2001). *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics.* Cambridge, MA: MIT Press.
Cooke, D. (1959). *The Language of Music,* pp. 89–90. Oxford, UK: Oxford University Press.
Cousineau, M., J. H. McDermott, and I. Peretz (2012). The basis of musical consonance as revealed by congenital amusia. Proc Natl Acad Sci USA 109(48): 19858–19863.
Craig, A. D. (2002) How do you feel? Interoception: The sense of the physiological condition of the body. Nat Rev Neurosci 3: 655–666.
Creel, W., P. C. Boomsliter, and S. R. Powers (1970). Sensations of tone as perceptual forms. Psych Rev 77(6): 534–545.
Crocker, R. L. (1963) Pythagorean mathematics and music source. J Aesth Art Culture 22: 189–198.
Crowder, R. G. (1984). Perception of the major / minor distinction: hedonic, musical, and affective discriminations. Bull Psychonom Soc 23: 314–316.
Crystal, D. (1976). *Prosodic Systems and Intonation in English.* New York: Cambridge University Press.
Crystal, D. (1997). *The Cambridge Encyclopedia of Language,* 2nd ed. New York: Cambridge University Press.
Curtis, M. E., and J. J. Bharucha (2010). The minor third communicates sadness in speech,

mirroring its use in music. Emotion 10(3): 335–348.
d'Alessandro, C., and P. d' Mertens (1995). Automatic pitch contour stylization using a model of tonal perception. Comp Speech Lang 9(3): 257–288.
d'Alessandro, C., S. Rosset, and J. Rossi (1998). The pitch of short-duration fundamental frequency glissandos. J Acoust Soc Am 104(4): 2339–2348.
Dai, H. (2000). On the relative influence of individual harmonics on pitch judgment. J Acoust Soc Am 107: 953–959.
Danielou, A. (1980). *The Ragas of Northern Indian Music,* pp. 40–42. New Delhi: Munshiram Manoharlal.
Darwin, C. (1872 / 2000). *The Expression of Emotion in Man and Animals.* Oxford, UK: Oxford University Press.
Day-O'Connell, J. (2007). *Pentatonicism from the Eighteenth Century to Debussy.* New York: University of Rochester Press.
de Beer, A. (1969). The development of 31 tone music. Sonorum Speculum 38: 26–38.
Delattre, P., A. M. Liberman, F. S. Cooper, and L. J. Gerstman (1952). An experimental study of the acoustic determinants of vowel color: observation of one- and two-formant vowels synthesized from spectrographic patterns. Word 8: 195–210.
Denes, P. (1963). On the statistics of spoken English. J Acoust Soc Am 35(6): 892.
Deutsch, D. (1973). Octave generalization of specific interference effects in memory for tonal pitch. Percept Psychophys 13: 271–275.
Deutsch, D. (ed.) (2013). *The Psychology of Music,* 3rd ed. New York: Academic Press.
Doolittle, E. L., B. Gingras, D. M. Endres, and W. T. Fitch. (2014). Overtone based pitch selection in hermit thrush song: Unexpected convergence with scale construction in human music. PNAS 111: 16616–16621, doi: 10.1073/pnas.1406023111, doi: 10.1073/pnas.1406023111
Dowling, W. J., and D. L. Harwood (1986). *Music Cognition.* New York: Academic.
Driscoll, T. A. (1997). Eigenmodes of isospectral drums. SIAM Rev 39: 1–17.
Duffin, R. W. (2007). *How Equal Temperament Ruined Harmony (and Why You Should Care).* New York: Norton.
Eerola, T., and P. Tovianien (2004). Suomen Kasan eSävelmät (Finnish Folk Song Database). Available from http://www.jyu.fi/musica/sks (accessed November 2008).
Escabi, M. A., and C. E. Schreiner (2002). Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. J Neurosci 22(10): 4114–4131.
Fastl, H., and G. Stoll (1979). Scaling of pitch strength. Hear Res 1: 293–301.
Fechner, G. T. (1860 / 1966). *Elements der Psychophysik.* Leipzig: Brietkopf und Hartel. (Vol. 1 translated as *Elements of Psychophysics* by H. E. Adler.) New York: Holt, Rinehart and Winston.
Ferguson, J. (2000). *All Blues Scales for Jazz Guitar.* Pacific, MO: Mel Bay.
Field, D. J. (1994). What is the goal of sensory coding? Neural Comp 6: 559–601.
Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. Cognition 100: 173–215.
Fitch, W. T. (2015). Four principles of biomusicology. Phil Roy Soc B. doi: 10.1098/rstb.2014.0091
Flanagan, J. L., and N. Gutman (1960). On the pitch of periodic pulses. J Acoust Soc Am 32:

1308–1319.
Fletcher, H. (1924). The physical criterion for determining the pitch of a tone. Phys Rev 23: 427–437.
Fletcher, H. (1934). Loudness, pitch and the timbre of musical tones and their relation to the intensity, the frequency and the overtone structure. J Acoust Soc Am 6(2): 59–69.
Fletcher, H., and W. A. Munson (1933). Loudness: Its definition, measurement, and calculation. J Acoust Soc Am 5: 82–108.
Fletcher, H., and R. H. Galt (1950). The perception of speech and its relation to telephony. J Acoust Soc Am 22(2): 89–151.
Fletcher, N. H. (1992). *Acoustic Systems in Biology.* New York: Oxford University Press.
Florentine, M., S. Buus, and T. Poulsen (1996). Temporal integration of loudness as a function of level. J Acoust Soc Am 99(3): 1633–1644.
Forster, C. M. L. (2010). *Musical Mathematics.* San Francisco: Chronicle Books.
Fritz, C., J. Curtin, J. Poitevineau, H. Borsarello, I. Wollman, F.-H. Tao, and T. Ghasarossian (2014). Soloist evaluations of six old Italian and six new violins. Proc Natl Acad Sci USA 111: 7224–7229.
Fritz, T., S. Jentschke et al. (2009). Universal recognition of three basic emotions in music. Current Biol 19: 573–576.
Fujioka, T., B. Ross, H. Okamoto, Y. Takeshima, R. Kakigi, and C. Pantev (2003). Tonotopic representation of missing fundamental complex sounds in the human auditory cortex. Eur J Neurosci 18: 432–440.
Gabrielsson, A., and E. Lindström (2001). The influence of musical structure on emotional expression. In: *Music and Emotion: Theory and Research* (P. N. Juslin and J. A. Sloboda, eds.). New York: Oxford University Press.
Garofolo, J. S., L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren (1990). DARPA-TIMIT acoustic-phonetic continuous speech corpus [CD-ROM]. Gaithersburg, MD: Department of Commerce.
Gerson, A., and J. L. Goldstein (1978). Evidence for a general template in central optimal processing for pitch of complex tones. J Acoust Soc Am 63: 498–510.
Gill, K. Z., and D. Purves (2009). A biological rationale for musical scales. PLOS ONE 4: e8144. doi:10.1371 / journal.pone.0008144.
Gingras B., H. Honing, I.Peretz, L.Trainor, and S. Fisher (2015). Defining the biological bases of individual differences in musicality. Phil Roy Soc B 370 (1664), 1–15. doi: 10.1098/rstb.2014.0092
Glasberg, B. R., and B. C. J. Moore (2006). Prediction of absolute threshold and equal-loudness contours using a modified loudness model. J Acoust Soc Am 120(2): 585–588.
Glave, R. D., and A. C. M. Rietveld (1975). Is the effort dependence of speech loudness explicable on the basis of acoustical cues? J Acoust Soc Am 58: 875–879.
Gold, L. (2007). *Music in Bali: Experiencing Music, Expressing Culture.* New York: Oxford University Press.
Goldstein, J. L. (2000). Pitch perception. In: *Encyclopedia of Psychology* (A. Kazdin, ed.). Oxford, UK: Oxford University Press.
Gordon, C., D. Webb, and S. Wolpert (1992). One cannot hear the shape of a drum. Bull Am Math Soc 27: 134–138.

Gorman, P. (1979). *Pythagoras, a Life.* London: Routledge and Kegan Paul.
Grassi, M., and R. Burro (2003) Impact sounds. In: *The Sounding Object* (D. Rocchesso and F. Fontana, eds.). Florence, Italy: PHASAR Srl.
Gregory, A. H., and N. Varney (1996). Cross-cultural comparisons in the affective response to music. Psychol Music 24: 47–52.
Griffiths, D. G., and J. D. Warren (2004). What is an auditory object? Nat Rev Neurosci 5: 887–892.
Hall, J. W., and R. W. Peters (1981). Pitch for nonsimultaneous successive harmonics in quiet and noise. J Acoust Soc Am 69: 509–513.
Hammerschmidt, K., and U. Jurgens (2007). Acoustical correlates of affective prosody. J Voice 21: 531–540.
Han, S., J. Sundararajan, D. L. Bowling, J. Lake, and D. Purves (2011). Co-variation of tonality in the music and speech of different cultures. PLOS One 6: e20160. doi:10.1371 / journal.pone.0020160.
Harrison, M. (2003). *Blues Piano: Hal Leonard Keyboard Style Series (Keyboard Instruction).* Milwaukee: Hal Leonard.
Heinlein, C. P. (1928). The affective characters of the major and minor modes in music. J Comp Psychol 8: 101–142.
Heller, E. J. (2013). *Why You Hear What You Hear: An Experimental Approach to Sound, Music, and Psychoacoustics.* Princeton, NJ: Princeton University Press.
Hellman, R. P., and J. Zwislocki (1963). Monaural loudness function at 1000 cps and interaural summation. J Acoust Soc Am 35(6): 856–865.
Helmholtz, H. L. F. (1866 / 1924–1925). *Helmholtz's Treatise on Physiological Optics,* 3rd German ed., Vols. I–III, 1909 (J. P. C. Southall, trans.). New York: Optical Society of America.
Helmholtz, H. L. F. (1877 / 1954). *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (A. J. Ellis, trans.). New York: Dover.
Hevner, K. (1935). The affective character of the major and minor modes in music. Am J Psychol 47: 103–118.
Hillenbrand, J., L. A. Getty, M. J. Clark, and K. Wheeler (1995). Acoustic characteristics of American English vowels. J Acoust Soc Am 97: 3099–3111.
Hirst, D., and A. Di Cristo (1999). *Intonation Systems: A Survey of Twenty Languages.* New York: Cambridge University Press.
Ho, L. T., and K. H. Han (1982). On Chinese scales and national modes. Asian Music 14(1): 132–154.
Hoeschele, M., H. Merchant, Y. Kikuchi, Y. Hattori, and C. ten Cate (2015). Searching for the origins of musicality across species. Phil Trans R Soc B 370: 20140094.
Hogden, J., A. Loqvist, V. Gracco, I. Zlokarnik, P. Rubin, and E. Saltzman (1996). Accurate recovery of articulator positions from acoustics: New conclusions based upon human data. J Acoust Soc Am 100: 1819–1834.
Hollien, H. (1960). Some laryngeal correlates of vocal pitch. J Speech Hear Res 3: 52–58.
Holt, L. L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. J Acoust Soc Am 120: 2801–2817.
Houtsma, A. J. M. (1995). Pitch perception. In: *Hearing,* pp. 267–295. New York: Academic.

Houtsma, A. J. M., and J. L. Goldstein (1972). The central origin of the pitch of complex tones: Evidence from musical interval recognition. J Acoust Soc Am 51: 520–529.

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

Huron, D. (1994). Interval-class content in equally-tempered pitch-class sets: Common scales exhibit optimum tonal consonance. Music Percept 11: 289–305.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice leading from perceptual principles. Music Percept 19: 1–64.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation.* Cambridge, MA: MIT Press.

Hutchinson, W., and L. Knopoff (1978). The acoustic component of Western consonance. Interface 7: 1–29.

Huth, A.G. et al. (2016). Natural speech reveals semantic maps that tile the human cerebral cortex. Nature 532: 453–458.

Iamblichus (c.300 / 1989). *On the Pythagorean Life* (G. Clark, trans.). Liverpool, UK: Liverpool University Press.

Isacoff, S. (2001). *Temperament: The Idea That Solved Music's Greatest Riddle.* New York: Knopf.

Janata, P., J. L. Birk, J. D. Van Dorn, M. Leman, B. Tillman, and J. J. Bharucha (2002). The cortical topography of tonal structures underlying Western music. Science 298: 2167–2170.

Jaramillo, F., V. S. Markin, and A. J. Hudspeth (1993). Auditory illusions and the single hair cell. Nature 364: 527–529.

Järveläinen, H., T. Verma, and V. Välimäki (2002). Perception and adjustment of pitch in inharmonic string instrument tones. J New Music Res 31: 311–319.

Jesteadt, W., and L. J. Leibold (2011). Loudness in the laboratory, part I: Steady-state sounds. In: *Loudness* (M. Florentine, A. N. Popper, and R. R. Fay, eds.), pp. 109–144. New York: Springer.

Johnston, I. (2002). *Measured Tones: The Interplay of Physics and Music.* New York: Taylor and Francis.

Johnstone, T., and K. R. Scherer (2000). Vocal communication of emotion. In: *Handbook of Emotions,* 2nd ed. (M. Lewis and J. M. Haviland-Jones, eds.). New York: Guilford.

Juslin, P. N., and P. Laukka (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? Psychol Bull 129: 770–814.

Kaas, J. H., and T. A. Hackett (2000). Subdivisions of auditory cortex and processing streams in primates. Proce Natl Acad Sci USA 97: 11793–11799.

Kallman, H. J., and D. W. Massaro (1979). Tone chroma is functional in melody recognition. Percept Psychophys 26: 32–36.

Kameoka, A., and M. Kuriyagawa (1969a). Consonance theory, part I: Consonance of dyads. J Acoust Soc Am 45: 1451–1459.

Kameoka, A., and M. Kuriyagawa (1969b). Consonance theory, part II: Consonance of complex tones and its computation method. J Acoust Soc Am 45: 1460–1469.

Kanwal, J. S., J. Kim, and K. Kamada (2000). Separate distributed processing of environmental, speech, and musical sounds in the cerebral hemispheres. J Cogn Neurosci (Suppl): 32.

Keefe, D. H., E. M. Burns, and P. Nguyen (1991). Vietnamese modal scales of the Dan Tranh.

Music Percept 8(4): 449–468.
Koelsch, S. (2014). Brain correlates of music-evoked emotions. Nat Rev Neurosci 15: 170–180.
Kreuger, F. (1913). Consonance and dissonance. J Phil Psychol Scient Meth 10: 158.
Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch.* New York: Oxford University Press.
Krumhansl, C. L., and R. N. Shepard (1979). Quantification of the hierarchy of tonal functions within a diatonic context. J Exp Psychol 5: 579–594.
Ladefoged, P. (1962). *Elements of Acoustic Phonetics.* Chicago: University of Chicago Press.
Ladefoged, P. (2000). *Vowels and Consonants: An Introduction to the Sounds of Languages,* 2nd ed. Oxford, UK: Wiley-Blackwell.
Ladefoged, P., and D. E. Broadbent (1957). Information conveyed by vowels. J Acoust Soc Am 29: 98–104.
Ladefoged, P., R. Harshman, L. Goldstein, and L. Rice (1978). Generating vocal tract shapes from formant frequencies. J Acoust Soc Am 64: 1027–1035.
Ladefoged, P., and N. P. McKinney (1963). Loudness, sound pressure, and subglottal pressure in speech. J Acoust Soc Am 35: 454–460.
Laland, K. N. and B. G. Galef (2009). *The Question of Animal Culture.* Cambridge, MA: Harvard University Press.
Langner, G., M. Sams, P. Heil, and H. Schulze (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: Evidence from magnetoencephalography. J Comp Physiol 181: 665–676.
Lee, S., A. Potamianos, and S. Narayanan (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. J Acoust Soc Am 105(3): 1455–1468.
Lehrdahl, F., and R. Jackendorff (1996). *A General Theory of Tonal Music.* Cambridge, MA: MIT Press.
Lentz, D. A. (1961). *Tone and Intervals of Hindu Classical Music.* Lincoln: University of Nebraska.
Levitan, D. J. (2006). *This Is Your Brain on Music.* New York: Dutton.
Lewicki, M. (2002). Efficient coding of natural sounds. Nat Neurosci 5: 356–363.
Liberman, A. M., and I. G. Mattingly (1985). The motor theory of speech perception revised. Cognition 21: 1–36.
Licklider, J. C. R. (1954). "Periodicity" pitch and "place" pitch. J Acoust Soc Am 26: 945.
Lieberman, P., and S. E. Blumstein (1988). *Speech Physiology, Speech Perception, and Acoustic Phonetics*. New York: Cambridge University Press.
Lotto, A. J., and L. Holt (2006). Putting phonetic context effects into context: A commentary on Fowler. Percept Psychophys 68(2): 178–183.
Lotto, A. J., and K. R. Kluender (1998). General contrast effects of speech perception: Effect of preceding liquid on stop consonant identification. Percept Psychophys 60: 602–619.
Lotto, A. J., K. R. Kluender, and L. L. Holt (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). J Acoust Soc Am 102: 1134–1140.
Lotto, A. J., S. Sullivan, and L. L. Holt (2003). Central locus for non-speech context effects on phonetic identification. J Acoust Soc Am 113: 53–56.
Loui, P., H. C. Li, A. Hohmann, and G. Schlaug (2010). Enhanced cortical connectivity in

absolute pitch musicians: A model for local hyperconnectivity. J Cogn Neurosci 23(4): 1015–1026.

Loui, P., A. Zamm, and G. Schlaug (2012). Enhanced functional networks in absolute pitch. Neuroimage 63: 632–640.

Mach, E. (1886 / 1959). *The Analysis of Sensations and the Relation of the Physical to the Psychical,* 1st German ed. (C. M. Williams, trans.). New York: Dover.

Malm, W. P. (1996). *Music Cultures of the Pacific, the Near East, and Asia,* 3rd ed. Upper Saddle River, NJ: Prentice Hall.

Malmberg, C. F. (1918). The perception of consonance and dissonance. Psychol Monogr 25: 93–133.

McDermott, J., and M. Hauser (2005). The origins of music: innateness, uniqueness and evolution. Music Percept 23: 29–59.

McDermott, J. H., A. J. Lehr, and A. J. Oxenham (2010). Individual differences reveal the basis of consonance. Curr Biol 20: 1035–1041.

McDermott, J.H., A.F. Schultz, E.A. Undurraga. and R.A. Godoy (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. Nature 535: 547–550.

McDermott, J. H., and E. P. Simoncelli (2011). Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. Neuron 71: 926–940.

McGurk, H., and J. MacDonald (1976). Hearing lips and seeing voices. Nature 264: 746–748.

McKinney, M. F., M. J. Tramo, and B. Delgutte (2001). Neural correlates of musical dissonance in the inferior colliculus. In: *Physiological and Psychophysical Bases of Auditory Function* (D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven, eds.), pp. 83–89. Maastricht, the Netherlands: Shaker.

McPhee, C. (1966), *Music in Bali.* New Haven, CT: Yale University Press.

Meddis, R., and M. Hewitt (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. J Acoust Soc Am 89: 2866–2882.

Mertens, P. (2004). The prosogram: Semi-automatic transcription of prosody based on tonal perception model. Presented at Speech Prosody 2004, Nara, Japan, March 23–26, 2004.

Miller, G. A. (1991). *The Science of Words.* New York: Freeman.

Miskiewicz, A., and A. Rawkowski (2012). A psychophysical pitch function determined by absolute magnitude estimation and its relation to the musical pitch scale. J Acoust Soc Am 131(1): 987–992.

Monson, B. B., S. Han, and D. Purves (2013). Are auditory percepts determined by experience? PLOS ONE 8(5): e63728. doi:10.1371 / journal.pone.0063728

Moore, B. C. J. (1973). Some experiments relating to the perception of complex tones. Q J Exp Psychol 25: 451–475.

Moore, B. C. J. (1982). *An Introduction to the Psychology of Hearing,* 2nd ed. London: Academic.

Moore, B. C. J. (1993). Frequency analysis and pitch perception. In: *Human Psychophysics* (W. A. Yost, A. N. Popper, and R. Foy, eds.), pp. 56–115. New York: Springer.

Moore, B. C. J. (2003a). *An Introduction to the Psychology of Hearing,* 5th ed. London: Academic.

Moore, B. C. J. (2003b). Coding of sounds in the auditory system and its relevance to signal

processing and coding in cochlear implants. Otol Neurotol 24:243–254.
Moore, B. C. J., J. I. Alcantara, and B. R. Glasberg (2002). Behavioural measurement of level-dependent shifts in the vibration pattern on the basilar membrane. Hear Res 163: 101–110.
Moore, B. C. J., B. R. Glasberg, and T. Baer (1997). A model for the prediction of thresholds, loudness, and partial loudness. J Aud Eng Soc 45(4): 224–240.
Moore, B. C. J., B. R. Glasberg, and Peters, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. J Acoust Soc Am 77: 1853–1860.
Morgan, C. T., W. R. Garner, and R. Galambos (1951). Pitch and intensity. J Acoust Soc Am 23(6): 658–663.
Morton, D. (1976). *The Traditional Music of Thailand.* Berkeley: University of California Press.
Mukherjee, P. (2004), *The Scales of Indian Music.* New Delhi: Aryan Books International.
Muthusamy, Y. K., R. A. Cole, and B. T. Oshika (1992). The OGI multi-language telephone speech corpus. Proceedings of the 1992 International Conference on Spoken Language Processing, ICSLP 92, Banff, Alberta, Canada, October.
Nelken, I., A. Fishbach, L. Las, N. Ulanovsky, and D. Farkas (2003). Primary auditory cortex of cats: Feature detection or something else? Biol Cybernet 89: 397–406.
Nettl, B. (1956). *Music in Primitive Culture.* Cambridge, MA: Harvard University Press.
Neuhoff, J. G. (2004). *Ecological Psychoacoustics.* San Diego: Elsevier.
Norman-Haignere, S., N.G. Kanwisher, and J. H. McDermott (2015). Distinct cortical pathways for music and speech revealed by hypothesis free voxel decomposition. Neuron 88: 1281–1296.
Ohgushi, K., and T. Hatoh (1992). The musical pitch of high frequency tones. In: *Auditory Physiology and Perception* (Y. Cazals, L. Demany, and K. Horner, eds.). Oxford, UK: Pergamon.
Palisca, C. V. (1961). Scientific empiricism in musical thought. In: *Seventeenth-Century Science and the Arts* (H. H. Rhys, ed.). Princeton, NJ: Princeton University Press, pp. 91–137.
Pantev, C., M. Hoke, B. Lutkenhoner, and K. Lehnertz (1989). Tonotopic organization of the auditory cortex: Pitch versus frequency representation. Science 246: 486–488.
Parncutt, R. (1989). *Harmony: A Psychoacoustical Approach.* Berlin: Springer.
Patel, A. D. (2008). *Music, Language, and the Brain.* New York: Oxford University Press.
Pear, T. H. (1911). Differences between major and minor chords. Br J Psychol 4: 56–94.
Peretz, I., A. J. Blood, V. Penhune, and R. Zatorre (2001). Cortical deafness to dissonance. Brain 124: 928–940.
Peretz, I., L. Gagnon, and B. Bouchard (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. Cognition 68: 111–141.
Perrodin, C., C. Kayser, N. K. Logothetis, and C. I. Petkov (2011). Voice cells in the primate temporal lobe. Curr Biol 21: 1408–1415.
Petersen, G. E., and H. L. Barney (1962). Control methods used in a study of the vowels. J Acoust Soc Am 24: 175–184.
Petkov, C. I., C. Kayser, T. Steudel, K. Whittingstall, M. Augath, and N. K. Logothetis (2007). A voice region in the monkey brain. Nat Neurosci 11: 367–374.
Pickett, J. M. (1957). Perception of vowels heard in noises of various spectra. J Acoust Soc Am 29: 613–620.

Pierce, J. R. (1966). Attaining consonance in arbitrary scales. J Acoust Soc Am 40: 249.
Pierce, J. R. (1991). Periodicity and pitch perception. J Acoust Soc Amer 90: 1889–1893.
Pierce, J. R. (2000). *The Science of Musical Sound.* New York: Scientific American Books.
Pierce, J. R. (2001). Introduction to pitch perception. In: *Music, Cognition and Computerized Sound* (P. Cook, ed.). Cambridge, MA: MIT Press.
Plack, C. J., and R. P. Carlyon (1995). Loudness perception and intensity coding. In: *Hearing* (B. C. J. Moore, ed.), pp. 123–160. New York: Academic.
Plack, C. J., and A. J. Oxenham (2005). The psychophysics of pitch. In: *Pitch: Neural Coding and Perception* (C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper, eds.), pp. 7–55. New York: Springer.
Plomp, R. (1964). The ear as a frequency analyzer. J Acoust Soc Am 36: 1628–1636.
Plomp, R. (1967). Pitch of complex tones. J Acoust Soc Am 41: 1526–1533.
Plomp, R. (1976). *Aspects of Tone Sensation.* London: Academic.
Plomp, R. (2002). *The Intelligent Ear: On the Nature of Sound Perception.* Mahwah, NJ: Erlbaum.
Plomp, R., and W. J. Levelt (1965). Tonal consonance and critical bandwidth. J Acoust Soc Am 28: 548–560.
Pollack, I. (1952). On the measurement of the loudness of speech. J Acoust Soc Am 24(3): 323–324.
Pressnitzer, D., R. D. Patterson, and K. Krumbholz (2001). The lower limit of melodic pitch. J Acoust Soc Am 109: 2074–2084.
Protopapas, A., and P. Lieberman (1996). Fundamental frequency of phonation and perceived emotional stress. J Acoust Soc Am 101: 2267–2277.
Purves D., G. A. Augustine, D. Fitzpatrick, W. Hall, A.-S. LaMantia, J. O. McNamara, and S. M. Williams (2008). *Neuroscience*, 4th ed. Sunderland, MA: Sinauer.
Purves, D., G. A. Augustine, D. Fitzpatrick, W. Hall, A.-S. LaMantia, and L. E. White (2012). *Neuroscience,* 5th ed. Sunderland, MA: Sinauer.
Purves, D., E. M. Brannon, R. Cabeza, S. A. Huettel, K. S. LaBar, M. L. Platt, and M. Woldorff (2013). *Principles of Cognitive Neuroscience,* 2nd ed. Sunderland, MA: Sinauer.
Purves D., and R. B. Lotto (2011). *Why We See What We Do Redux.* Sunderland, MA: Sinauer.
Purves, D., B. Monson, J. Sundarajan, and W. T. Wojtach (2014). How biological vision succeeds in the physical world. Proc Natl Acad Sci USA 111: 4750–4755.
Purves, D., Y. Morgenstern, and W. T. Wojtach (2015). Will understanding vision require a wholly empirical paradigm? Front Psychol http://dx.doi.org/10.3389/fpsyg.2015.01072.
Ramaeu, J.-P. (1722). *Traité de l'harmonie réduite à ses principes naturels.* Paris: Jean-Baptiste-Christophe Ballard.
Randel, D. M. (ed.) (1986). *The New Harvard Dictionary of Music,* revised 2nd ed. Cambridge, MA: Belknap.
Rasch, R., and R. Plomp (1999). The perception of musical tones. In: *The Psychology of Music* (D. Deutsch, ed.), pp. 89–112. New York: Academic.
Rauschecker, J. P., and B. Tian (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci USA 97: 11800–11806.
Read, H. L., J. A. Winer, and C. E. Scheiner (2002). Functional architecture of auditory cortex. Curr Opin Neurobiol 12: 433–440.

Relkin, E. M., and J. R. Doucet (1997). Is loudness simply proportional to the auditory nerve spike count? J Acoust Soc Am 101(5): 2735–2740.

Ritsma, R. J. (1967). Frequencies dominant in the perception of the pitch of complex sounds. J Acoust Soc Am 42: 191–198.

Ritsma, R. J. (1970). Periodicity detection. In: *Frequency Analysis and Periodicity Detection in Hearing* (R. Plomp, and G. F. Smoorenburg, eds.), pp. 250–263. Leiden: Sijthoff.

Robinson, D. W., and R. S. Dadson (1956). A re-determination of the equal-loudness relations for pure tones. Br J Appl Phys 7: 166–181.

Robles, L., M. A. Ruggero, and N. C. Rich (1991). Two-tone distortion in the basilar membrane of the cochlea. Nature 349: 413–414.

Rosner, B. S., and J. B. Pickering (1994). *Vowel Perception and Production.* New York: Oxford University Press.

Ross, D., J. Choi, and D. Purves (2007). Musical intervals in speech. Proc Natl Acad Sci USA 104(23).

Rossing, T. D., R. F. Moore, and P. A. Wheeler (2002). *The Science of Sound,* 3rd ed. San Francisco: Addison-Wesley.

Rugerro, M. A. (1992). Responses to sound of the basilar membrane of the mammalian cochlea. Curr Opin Neurobiol 2: 449–456.

Rugerro, M. A., N. C. Rich, A. Recio, S. S. Narayan, and L. Robles (1997). Basilar-membrane responses to tones at the base of the chinchilla cochlea. J Acoust Soc Am 101(4): 2151–2163.

Russell, I. J., and K. E. Nilsen (1997). The location of the cochlear amplifier: Spatial representation of a single tone on the guinea pig basilar membrane. Proc Natl Acad Sci USA 94: 2660–2664.

Saffran, J. R., E. L. Newport, and R. N. Aslin (1996). Word segmentation: The role of distributional cues. J Mem Lang 35: 606–621.

Samson, J. (1977). *Music in Transition: A Study of Tonal Expansion and Early Atonality, 1900–1920.* New York: Norton.

Satyanarayana, N. C. (2005). *The Science of Indian Music.* Hyderabad, India: Sri Dattasai Graphics.

Savage, P. E., S. Brown, E. Sakai, and T. E. Currie (2015). Statistical universals reveal the structures and functions of human music. PNAS 112: 8987–8992.

Schellenberg, E. G., A. M. Krysciak, and R. J. Campbell (2000). Perceiving emotion in melody: Interactive effects of pitch and rhythm. Music Percept 18: 155–171.

Schellenberg, E. G., and S. E. Trehub (1996). Natural music intervals: Evidence from infant listeners. Psychol Sci 7: 272–277.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. Speech Commun 40: 227–256.

Scherer, K. R., R. Banse, and H. G. Wallbott (2001). Emotional inferences from vocal expression correlate across languages and cultures. J Cross Cult Psychol 32: 76–92.

Schnupp, J., I. Nelken, and A. King (2012). *Auditory Neuroscience: Making Sense of Sound.* Cambridge MA: MIT Press.

Schouten, J. F. (1938). The perception of subjective tones. Proc Kon Ned Akad Weten 34: 1086–1093.

Schouten, J. F. (1940). The residue, a new component in subjective sound analysis. Proc Kon

Ned Akad Weten 43: 356–365.
Schouten, J. F., R. J. Ritsma, and B. I. Cardozo (1962). Pitch of the residue. J Acoust Soc Am 34: 1418–1424.
Schreiner, C. E., and G. Langner (1997). Laminar fine structure of frequency organization in auditory midbrain. Nature 388: 383–386.
Schreiner, C. E., H. L. Read, M. L. Sutter (2000). Modular organization of frequency integration in primary auditory cortex. Annu Rev Neurosci 23: 501–529.
Schwartz, D. A., and D. Purves (2004). Pitch is determined by naturally occurring periodic sounds. Hear Res 194: 31–46.
Schwartz, D. A., C. Q. Howe, and D. Purves (2003). The statistical structure of human speech sounds predicts musical universals. J Neurosci 23: 7160–7168.
Schwartz, O., and E. P. Simoncelli (2001). Natural signal statistics and sensory gain control. Nat Neurosci 4(8): 819–825.
Seebeck, A. (1841). Beobachtungenuber über einige Bedingungen der Entschung von Tönen. Annal Physik Chem 53: 417–436.
Sethares, W. A. (1993). Local consonance and the relationship between timbre and scale. J Acoust Soc Am 94: 1218–1228.
Sethares, W. A. (1998). *Timbre, Tuning, Spectrum, Scale.* New York: Springer.
Shepard, R. N. (1999). Pitch perception and measurement. In: *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics* (P. R. Cook, ed.), pp. 149–166. Cambridge, MA: MIT Press.
Sherrington, S. C. (1906 / 1947). *The Integrative Action of the Nervous System,* 2nd ed. New Haven, CT: Yale University Press.
Siegel, R. J. (1965). A replication of the mel scale of pitch. Am J Psychol 78(4): 615–620.
Smith, E., and M. S. Lewicki (2006). Efficient auditory coding. Nature 439: 978–982.
Smoorenburg, G. F. (1970). Pitch perception of two-frequency stimuli. J Acoust Soc Am 48: 924–942.
Snowdon, C. T., and D. Teie (2010). Affective responses in tamarins elicited by species-specific music. Biol Lett 6: 30–32.
Song, X., M. S. Omanski, Y. Guo, and X. Wang (2016). Complex pitch perception mechanisms are shared by humans and a New World monkey. PNAS 113: 781–786.
Spencer, H. (1857). The origin and function of music. Fraser Mag 56: 396–408.
Sridhar, S., and A. Kudrolli (1994). Experiments on not “hearing the shape” of drums. Phys Rev Lett 72: 2175–2178.
Stevens, S. S. (1935). The relation of pitch to intensity. J Acoust Soc Am 6: 150–154.
Stevens, S. S. (1936). A scale for the measurement of a psychological magnitude: Loudness. Psych Review 43(5): 405–416.
Stevens, S. S. (1955). The measurement of loudness. J Acoust Soc Am 27(5): 815–829.
Stevens, S. S. (1956). The direct estimation of sensory magnitudes: Loudness. Am J Psychol 69(1): 1–25.
Stevens, S. S. (1970). Neural events and the psychophysical law. Science 170(3962): 1043–1050.
Stevens, S. S. (1986). *Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects* (G. Stevens, ed.). Oxford UK: Transaction Books.

Stevens, S. S., and J. Volkmann (1940). The relation of pitch to frequency: A revised scale. Am J Psychol 53(3): 329–353.
Stewart, L. (2006). Music and the brain: Disorders of musical listening. Brain 129: 2533–2553.
Stumpf, C. (1898). Konsonanz and Dissonanz. Beitrage Ak Musikviss 1: 91–107.
Suen, C. Y. (1982). Computational analysis of Mandarin sounds with reference to the English language. Proceedings of the 9th conference on computational linguistics, Vol. 1, pp. 371–376. Amsterdam: North Holland.
Suzuki, Y., and H. Takeshima (2004). Equal-loudness-level contours for pure tones. J Acoust Soc Am 116(2): 918–933.
Tartini, G. (1754). Trattato di musica secondo la vera scienza dell'armonia. Padua, Italy: Nella Stemperia del Seminario.
Terhardt, E. (1974). Pitch, consonance, and harmony. J Acoust Soc Am 55: 1061–1069.
Terhardt, E. (1984). The concept of musical consonance: A link between music and psychoacoustics. Music Percept 1: 276–295.
Terhardt, E., G. Stoll, R. Schermbach, and R. Parncutt (1986). Tonhöhenmehrdeutigkeit, Tonverwandtschaft und Identifikation von Sukzessivintervallen. Acustica 61: 57–66.
Terhardt, E., G. Stoll, and M. Seewann (1982a). Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions. J Acoust Soc Am 71: 671–678.
Terhardt, E., G. Stoll, and M. Seewann (1982b). Algorithm for extraction of pitch and pitch salience from complex tonal signals. J Acoust Soc Am 71: 679–688.
Thompson, W. F., P. Graham, and F. A Russo (2005). Seeing music performance: Visual influences on perception and experience. Semiotica 156(1 / 4): 177–201.
Thompson, W. F., and L. L. Balkwill (2006). Decoding speech prosody in five languages. Semiotica 158(1 / 4): 407–424.
Touma, H. H. (1996). *The Music of the Arabs.* Portland OR: Amadeus.
Tramo, M. J. (2001). Biology and music: Music of the hemispheres. Science 291: 54–56.
Tramo, M. J., P. A. Cariani, B. Delgutte, and L. D. Braida (2001). Neurobiological foundations for the theory of harmony in western tonal music. In: *The Biological Foundations of Music* (I. Peretz and R. J. Zatorre, eds.), pp. 92–116. New York: New York Academy of Sciences.
Turner, R. S. (1977). The Ohm–Seebeck dispute, Hermann von Helmholtz, and the origins of physiological acoustics. Br J History Sci 10: 1–24.
Umeda, N. (1974). Consonant duration in American English. J Acoust Soc Am 61(3): 846–858.
Verschuure, J., and A. A. van Meeteren (1975). The effect of intensity on pitch. Acta Acust United Acust 32(1): 33–44.
Vilares, I., J. D. Howard, H. L. Fernandes, J. A. Gottfried, and K. P. Kording (2012). Differential representations of prior and likelihood uncertainty in the human brain. Curr Biol 22: 1–8.
von Békésy, G. (1962). Three experiments concerned with pitch perception. J Acoust Soc Am 35: 602–666.
von Hoerner, S. (1974). Universal music? Psych Music 2: 18–28.
von Hoerner, S. (1976). The definition of major scales, for chromatic scales of 12, 19 and 31 divisions per octave. Psych Music 4: 12–23.
Vos, P. G., and J. M. Troost (1989). Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. Music Percept 6: 383–396.

Vouloumanos, A., and J. F. Werker (2004). Tuned to the signal: The privileged status of speech for young infants. Dev Sci 7: 270–276.

Vouloumanos, A., and J. F. Werker (2007). Listening to language at birth: Evidence for a bias for speech in neonates. Dev Sci 10: 159–164.

Wang, X. (2013). The harmonic organization of auditory cortex. Front Syst Neurosci 7: 114. doi: 10.3389/fnsys.2013-00114.

Warren, R. M. (1974). Auditory temporal discrimination by trained listeners. Cogn Psych 6: 237–256.

Warren, R. M. (1999). *Auditory Perception: A New Analysis and Synthesis.* Cambridge, UK: Cambridge University Press.

Weinberger, N. M. (2004). Specific long-term memory traces in primary auditory cortex. Nat Rev Neurosci 5: 279–290.

Wiese, R. (2000). *The Phonology of German.* Oxford, UK: Oxford University Press.

Wright, A. A., J. J. Rivera, S. H. Hulse, M. Shyan, and J. J. Neiworth (2000). Music perception and octave generalization in rhesus monkeys. J Exp Psychol Gen 129: 291–307.

Wright, O. (1978). *The Modal System of Arab and Persian Music, A.D. 1250–1300.* Oxford, UK: Oxford University Press.

Wundt, W. (1862 / 1961). Contributions to the theory of sensory perception. (English translation taken from Wundt, W., 1961). In: *Classics in Psychology* (T. Shipley, ed.), pp. 51–78. New York: Philosophical Library.

Yost, W. A. (2000). *Fundamentals of Hearing: An Introduction.* San Diego: Academic.

Youngblood, J. (1958) Style as information. J Mus Theory 2: 24–31.

Zahorik P., and F. L. Wrightman (2001) Loudness constancy varying with sound source distance. Nat Neurosci 4: 78–83.

Zarlino, G. (1558). Bk. 3 of *Le Institutioni hamoniche* (As *The Art of Counterpoint,* 1968, Marco, G., and C. Palisca, trans.). New Haven, CT: Yale University Press.

Zatorre, R. J., and J. M. Zarate (2012). Cortical processing of music. In: *The Human Auditory Cortex* (D. Poeppel et al., eds.). New York: Springer.

Zatorre, R. J., J. L. Chen, and V. B. Penhune (2007). When the brain plays music. Nat Rev Neurosci 8: 547–558. doi: 10.1038/nrn2152.

Zeng, F. G., and R. V. Shannon (1994). Loudness-coding mechanisms inferred from electric stimulation of the human auditory system. Science 264: 564–566.

Zentner, M. R., and J. Kagan (1998). Infants' perception of consonance and dissonance in music. Infant Behav Dev 21: 483–492.

Zwicker, E., and H. Fastl (1990). *Psychoacoustics: Facts and Models.* Berlin: Springer.

Zwicker, E., and B. Scharf (1965). A model of loudness summation. Psych Rev 72(1): 3–26.

Acknowledgments

I am grateful to the long list of people who have helped me explore music in biological terms over the past fifteen years. When this project started, I had never had a course in music theory or appreciation, and what little I now know has come from innumerable conversations, arguments, and instruction from Nigel Barella, Dan Bowling, Jonathan Choi, Ruby Froom, Kamraan Gill, Danny Gotham, Henry Greenside, Catherine Howe, Brian Monson, Joey Prinz, Randy Reed, Debbie Ross, David Schwartz, Han Shui'er, and Jananini Sundararajan. I am especially grateful to my collaborators Dan, Kamraan, Jananini, Joey, Shui'er, and David for letting me use the results of their hard work. Finally, Hebron Daniel, Jeanne Shi, and Jay Shin corrected many details in earlier versions of these chapters, while Ruby Froom and Rich Mooney provided feedback from people who know music (and in Rich's case the auditory system) far better than I do.

Index

Table of Contents